PREMIER REFERENCE SOURCE

MEDICAL INFORMATICS

Concepts, Methodologies, Tools, and Applications

VOLUME I

JOSEPH TAN

TAN

MEDICAL INFORMATICS
Concepts, Methodologies, Tools, and Applications

Volume I

Information Science Reference

TAN

MEDICAL INFORMATICS
Concepts, Methodologies, Tools, and Applications

Volume II

Information Science Reference

TAN

MEDICAL INFORMATICS
Concepts, Methodologies, Tools, and Applications

Volume III

Information Science Reference

TAN

MEDICAL INFORMATICS
Concepts, Methodologies, Tools, and Applications

Volume IV

Information Science Reference

Medical Informatics:
Concepts, Methodologies, Tools, and Applications

Joseph Tan
Wayne State University, USA

Medical Information Science
REFERENCE

MEDICAL INFORMATION SCIENCE REFERENCE
Hershey · New York

Acquisitions Editor: Kristin Klinger
Development Editor: Kristin Roth
Senior Managing Editor: Jennifer Neidig
Managing Editor: Jamie Snavely
Typesetter: Michael Brehm, Jeff Ash, Carole Coulson, Elizabeth Duke, Jennifer Henderson, Chris Hrobak, Sean Woznicki
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com/reference

and in the United Kingdom by
Information Science Reference (an imprint of IGI Global)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 0609
Web site: http://www.eurospanbookstore.com

Library of Congress Cataloging-in-Publication Data

Medical informatics : concepts, methodologies, tools, and applications / Joseph Tan, editor.
p. ; cm.
Includes bibliographical references and index.
Summary: "This set of book is a complete collection of cutting-edge medical IT research available in topics such as clinical knowledge management, medical informatics, and gene expression"--Provided by publisher.
ISBN 978-1-60566-050-9 (hc)
1. Medical informatics. I. Tan, Joseph K. H.
[DNLM: 1. Medical Informatics. W 26.5 M48966 2009]
R858.M3972 2009
651.5'04261--dc22
2008023343

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

List of Contributors

Contents
by Volume

Volume I

Section 1. Fundamental Concepts and Theories

This section serves as the foundation for this exhaustive reference tool by addressing crucial theories essential to the understanding of medical informatics. Chapters found within these pages provide an excellent framework in which to position medical informatics within the field of information science and technology. Individual contributions provide overviews of knowledge management in healthcare, health portals, and health information systems, while also exploring critical stumbling blocks of this field. Within this introductory section, the reader can learn and choose from a compendium of expert research on the elemental theories underscoring the research and application of medical informatics.

Section 2. Development and Design Methodologies

This section provides in-depth coverage of conceptual architectures, frameworks and methodologies related to the design and implementation of medical information systems and technologies. Throughout these contributions, research fundamentals in the discipline are presented and discussed. From broad examinations to specific discussions on particular frameworks and infrastructures, the research found within this section spans the discipline while also offering detailed, specific discussions. Basic designs, as well as abstract developments, are explained within these chapters, and frameworks for designing successful e-health applications, mobile healthcare systems, and clinical decision support systems are discussed.

Volume II

Section 3. Tools and Technologies

This section presents extensive coverage of the interaction between healthcare and the various technologies that both derive from and support the research of practitioners, students, and patients alike. These chapters provide an in-depth analysis of mobile technologies and their use in healthcare, while also providing insight into new and upcoming technologies, theories, and instruments that will soon be commonplace. Within these rigorously researched chapters, readers are presented with countless examples of the tools that facilitate and support medical innovations. In addition, the successful implementation and resulting impact of these various tools and technologies are discussed within this collection of chapters.

This section introduces and discusses the ways in which information technology has been used to shape the medical and biomedical sectors and proposes new ways in which IT-related innovations can be implemented within organizations and in society as a whole. These particular selections highlight, among other topics, the implementation of telehealth and the evolution of new, computerized systems in hospitals. Contributions included in this section provide excellent coverage of today's medical environment and insight into how medical informatics impacts the fabric of our present-day global village.

Volume III

Section 5. Organizational and Social Implications

This section includes a wide range of research pertaining to the social and organizational impact of medical informatics around the world. Chapters introducing this section analyze organizational factors in health informatics and patient centric health information systems, while later contributions offer an extensive analysis of the Internet's role in revolutionizing accessibility to medical-related information. The inquiries and methods presented in this section offer insight into the implications of medical informatics at both a personal and organizational level, while also emphasizing potential areas of study within the discipline.

Section 6. Managerial Impact

This section presents contemporary coverage of the managerial implications of medical informatics, more specifically related to successful IT governance in the healthcare sector. Particular contributions address the implementation of large IT projects in healthcare institutions and the effective measurement of cost efficiency in healthcare. The managerial research provided in this section allows executives, practitioners, and researchers to gain a better sense of the role of IT in both healthcare and medical research.

Volume IV

Section 7. Critical Issues

This section addresses conceptual and theoretical issues related to the field of medical informatics, which include maintaining high ethical standards and ensuring that sensitive medical information remains private. Within these chapters, the reader is presented with analysis of the most current and relevant conceptual inquires within this growing field of study. Particular chapters also address mining biomedical literature, standardization of health and medical informatics, and fuzzy logic in medicine. Overall, contributions within this section ask unique, often theoretical questions related to the study of medical informatics and, more often than not, conclude that solutions are both numerous and contradictory.

Section 8. Emerging Trends

This section highlights research potential within the field of medical informatics while exploring uncharted areas of study for the advancement of the discipline. Chapters within this section highlight evolutions in evidence-based medicine, the future of e-health, and web portals for medical data. These contributions, which conclude this exhaustive, multi-volume set, provide emerging trends and suggestions for future research within this rapidly expanding discipline.

Preface

The need to efficiently deliver and process information in the healthcare and biomedical sectors is a primary concern among practitioners, researchers, and patients alike. Medical informatics—a field that has emerged at the intersection of information technology and medicine—has transformed modern healthcare and created new, more pervasive methods for access to information, records, and even medical advice. As medical informatics continues to evolve and researchers continue to create and implement technologies for use in the study and practice of medicine, we must continue to understand, develop, and utilize the latest in medical research and exploration.

In recent years, the applications and technologies generated through the study of medical informatics have grown in both number and popularity. As a result, researchers, clinicians, practitioners, and educators have devised a variety of techniques and methodologies to develop, deliver, and, at the same time, evaluate the effectiveness of their use. The explosion of methodologies in the field has created an abundance of new, state-of-the-art literature related to all aspects of this expanding discipline. This body of work allows researchers to learn about the fundamental theories, latest discoveries, and forthcoming trends in the field of medical informatics.

Constant technological and theoretical innovation challenges researchers to remain informed of and continue to develop and deliver methodologies and techniques utilizing the discipline's latest advancements. In order to provide the most comprehensive, in-depth, and current coverage of all related topics and their applications, as well as to offer a single reference source on all conceptual, methodological, technical, and managerial issues in medical informatics, Information Science Reference is pleased to offer a four-volume reference collection on this rapidly growing discipline. This collection aims to empower researchers, practitioners, and students by facilitating their comprehensive understanding of the most critical areas within this field of study.

This collection, entitled **Medical Informatics: Concepts, Methodologies, Tools, and Applications**, is organized into eight distinct sections which are as follows: 1) Fundamental Concepts and Theories, 2) Development and Design Methodologies, 3) Tools and Technologies, 4) Utilization and Application, 5) Organizational and Social Implications, 6) Managerial Impact, 7) Critical Issues, and 8) Emerging Trends. The following paragraphs provide a summary of what is covered in each section of this multi-volume reference collection.

Section I, **Fundamental Concepts and Theories**, serves as a foundation for this exhaustive reference tool by addressing crucial theories essential to understanding medical informatics. Opening this elemental section is "Evaluation of Health Information Systems: Challenges and Approaches" by Elske Ammenwerth, Stefan Gräber, Thomas Bürkle, and Carola Iller. This selection addresses some of the primary issues and challenges in evaluating the use of IT in healthcare and suggests methods for improvement. Specific issues in medical informatics, such as the emergence of the Internet as a healthcare tool and knowledge management as it pertains to the healthcare industry, are discussed in selections such as “The Telehealth Divide” by Mary Schmeida and Ramona McNeal and “Knowledge Management in Healthcare” by Sushil K. Sharma, Nilmini Wickramasinghe, and Jatinder N.D. Gupta. Within the contribution “Information Technology (IT) and the Healthcare Industry: A SWOT Analysis,” authors Marilyn M. Helms, Rita Moore, and Mohammad Ahmadi utilize the SWOT analysis (strengths, weaknesses, opportunities and threats) to conceptualize and further evaluate the many issues facing IT implementation in healthcare, which include improved patient safety, cost, and user resistance. The selections within this comprehensive, foundational section allow readers to learn from expert research on the elemental theories underscoring medical informatics.

Section II, **Development and Design Methodologies**, contains in-depth coverage of conceptual architectures and frameworks, providing the reader with a comprehensive understanding of emerging theoretical and conceptual developments within the development and utilization of medical technologies. “The Development of a Health Data Quality Programme” by Karolyn Kerr and Tony Norris presents the case of the New Zealand Ministry of Health’s construction of a new data quality strategy aligned with the current health information program. Other selections, such as “Building Better E-Health Through a Personal Health Informatics Pedagogy” by E. Vance Wilson and “The PsyGrid Experience: Using Web Services in the Study of Schizophrenia” by John Ainsworth and Robert Harper, offer insight into the design and use of Web services to guide and inform medical decisions. The design and implementation of mobile-based health services is explored at length in selections such as “Enabling Conceptual Framework for Mobile-Based Application in Healthcare” by Matthew W. Guah; “Design of an Enhanced 3G-Based Mobile Healthcare System” by José Ruiz Mas, Eduardo Antonio Viruete Navarro, Carolina Hernández Ramos, Álvaro Alesanco Iglesias, Julián Fernández Navajas, Antonio Valdovinos Bardají, Robert S. H. Istepanian, and José García Moros; and “The M-Health Reference Model: An Organizing Framework for Conceptualizing Mobile Health Systems” by Phillip Olla and Joseph Tan. From basic designs to abstract development, chapters such as “A Cross-Cultural Framework for Evolution” by Pekka Turunenand and “A Distributed Patient Identification Protocol Based on Control Numbers with Semantic Annotation” by Marco Eichelberg, Thomas Aden, and Wilfried Thoben serve to expand the reaches of development and design methodologies within the field of medical informatics.

Section III, **Tools and Technologies**, presents extensive coverage of various tools and technologies and their use in creating and expanding the reaches of health and biomedicine. The emergence of wireless and mobile devices and the opportunities these devices present for revolutionizing traditional patient care is the subject of articles such as “PDA Usability for Telemedicine Support” by Shirley Ann Becker; “A Preliminary Study toward Wireless Integration of Patient Information System” by Abdul-Rahman Al-Ali, Tarik Ozkul, and Taha Landolsi; and “Choosing Technologies for Handheld and Ubiquitous Decision Support” by Darren Woollatt, Paul Koop, Sara Jones, and Jim Warren. Advancements in imaging for medical and biomedical applications are analyzed and assessed in several selections, which include “Imaging the Human Brain with Magnetoencephalography” by Dimitrios Pantazis and Richard M. Leahy and “Imaging Technologies and their Applications in Biomedicine and Bioengineering” by Nikolaos Giannakakis and Efstratios Poravas. The latter of these two chapters discusses biomedical imaging technologies such as MRI and x-ray, offering insight into the research opportunities these technologies have provided as well as the limitations associated with their use. The rigorously researched chapters contained within this section offer readers countless examples of modern tools and technologies that emerge from or can be applied to the medical and healthcare sectors.

Section IV, **Utilization and Application**, investigates the use and implementation of medical technologies and informatics in a variety of contexts. This collection of innovative research begins with “Successful Health Information System Implementation” by Kristiina Häyrinen and Kaija Saranto in which primary success factors for employing health systems, such as system qualities, information quality, usage, user satisfaction, and individual impact, are analyzed. The delivery of health information via telecommunications networks (also known as telehealth) is studied in selections such as "Telehealth Organizational Implementation Guideline Issues: A Canadian Perspective" by researchers Maryann Yeo and Penny A. Jennett and "Tele-Medicine: Building Knowledge-Base Tele-Health Capability in New Zealand" by Nabeel A. Y. Al-Qirim. Further contributions explore other key strategies and factors that relate to the successful use of electronic health records, mobile e-health, ICT, and knowledge management in a medical environment. From established applications to forthcoming innovations, contributions in this section provide excellent coverage of today's global community and demonstrate how medical informatics impacts the social, economic, and political fabric of our present-day global village.

Section V, **Organizational and Social Implications**, includes a wide range of research pertaining to the organizational and cultural implications of medical informatics. "Using Hospital Web Sites to Enhance Communication" by Sherrie D. Cannoy and Lakshmi Iyer investigates patient communication-enhancing features of hospital Web sites, ultimately contending that a hospital's Web presence must both address and cater to users' communication needs in order to be effective. Web portals and their use in fostering social interaction and knowledge enhancement are explored at length in chapters such as "Intelligent Portals for Supporting Medical Information Needs" by Jane Moon and Frada Burstein, "Health Portals and Menu-Driven Identities" by Lynette Kvasny and Jennifer Warren, and "Empowerment and Health Portals" by Mats Edenius. Other issues that are surveyed within this section include the implications of the digital divide in healthcare within Michele Masucci's "Digital Divide and E-Health Implications for E-Collaboration Research" and community-centered IT outreaches within Rosanna Tarsiero's "Community-Based Information Technology Interventions for Persons with Mental Illness." Overall, the discussions presented in this section offer insight into the integration of medical informatics in society and the benefit these innovations have provided.

SectionVI, **Managerial Impact**, presents contemporary coverage of the managerial applications and implications of medical informatics. Core concepts such as information security management, outsourcing, and healthcare technology management are discussed in this collection. "Information Assurance in E-Healthcare" by Sherrie D. Cannoy and A. F. Salam addresses the healthcare industry's limited adoption of IT advancements, which is now being remedied by new advancements in information assurance that guarantee the safety of patients' medical records. Similarly, within their article "Modelling Context-Aware Security for Electronic Health Records," Pravin Shetty and Seng Loke suggest context-based security policies for health organizations, which are able to adapt to new, incoming threats. Later contributions, such as "E-Health Dot-Coms' Critical Success Factors," further investigate the Internet's role in reshaping healthcare. Within this selection, authors Abrams A. O'Buyonge and Leida Chen evaluate the business models utilized by heath-information Web sites (such as WebMD) and present the lessons learned from a managerial perspective. The comprehensive research in this section offers an overview of the major issues that healthcare practitioners, governments, and even consumers must address in order to remain informed about the latest managerial changes in the field of medical informatics.

Section VII, **Critical Issues**, presents readers with an in-depth analysis of the more theoretical and conceptual issues within this growing field of study by addressing topics such as the quality and security of medical information. Specifically, these topics are discussed in selections such as "Medical Ethical and Policy Issues Arising from RIA" by Jimmie L. Joseph and David P. Cook and "E-Health Security and Privacy" by Yingge Wang, Qiang Cheng, and Jie Cheng. The latter of these two selections investigates relevant concepts, technologies, limitations, challenges, and trends in e-health security and privacy, along with standards such as the Health Insurance Portability and Accountability Act (HIPAA). Similarly, in contributions such as "Reliability and Evaluation of Health Information Online" by Elmer V. Bernstam and Funda Meric-Bernstam, the issue of how to effectively evaluate online health information is debated and a how-to guide for obtaining medical information online is suggested. Later selections, which include "Ontology-Based Spelling Correction for Searching Medical Information," review more novel issues, such as the difficulty in retrieving medical information online due to errors in spelling medical terms. In this chapter, researchers Jane Moon and Frada Burstein from Monash University propose an ontology-based architecture that would assist users with medical information retrieval. In all, the theoretical and abstract issues presented and analyzed within this collection form the backbone of revolutionary research in and evaluation of medical informatics.

The concluding section of this authoritative reference tool, **Emerging Trends**, highlights research potential within the field of medical informatics while exploring uncharted areas of study for the advancement of the discipline. The development and deployment of new forms of health information technologies (HITs) are proposed in Avnish Rastogi, Tugrul Daim, and Joseph Tan's "Charting Health Information Technology Futures for Healthcare Services Organizations," while the latest innovations in e-heath systems are analyzed in Pirkko Nykänen's "E-Health Systems: Their Use and Visions for the Future." In the contribution "Outsourcing of Medical Surgery and the Evolution of Medical Telesurgery," Shawna Sando asserts that, due to the rising cost of healthcare in the United States, the best alternative for some low- and middle-class citizens is to seek medical care overseas—to engage in "medical tourism." Other new trends, such as the emergence of evidence-based medicine, the creation of biotechnology portals in medicine, and revolutions in emergency medical services, are discussed in this collection. This final section demonstrates that medical informatics, with its propensity for constant change and evolution, will continue to both shape and define the modern face of healthcare and the ways in which we interact with health-related information.

Although the contents of this multi-volume book are organized within the preceding eight sections which offer a progression of coverage of important concepts, methodologies, technologies, applications, social issues, and emerging trends, the reader can also identify specific contents by utilizing the extensive indexing system listed at the end of each volume. Furthermore, to ensure that the scholar, researcher, and educator have access to the entire contents of this multi-volume set, as well as additional coverage that could not be included in the print version of this publication, the publisher will provide unlimited, multi-user electronic access to the online aggregated database of this collection for the life of the edition free of charge when a library purchases a print copy. In addition to providing content not included within the print version, this aggregated database is also continually updated to ensure that the most current research is available to those interested in medical informatics.

As medical technologies and the methods for evaluating medical data continue to evolve and new ways to store, process and access information are discovered, medical informatics will become an increasingly critical field of study. The nature of personal heath records, diagnosis, and even treatment have changed drastically due to the efforts of researchers, practitioners, and patients to make medical information more easily available, more secure, and of a higher quality than ever before. Innovations in the effective storage, retrieval, and understanding of medical information capitalize on the constant technological changes that seek to streamline and improve modern society.

The diverse and comprehensive coverage of medical informatics in this four-volume, authoritative publication will contribute to a better understanding of all topics, research, and discoveries in this developing, significant field of study. Furthermore, the contributions included in this multi-volume collection series will be instrumental in the expansion of the body of knowledge in this enormous field, resulting in a greater understanding of the fundamentals while also fueling the research initiatives in emerging fields. We at Information Science Reference, along with the editor of this collection, hope that this multi-volume collection will become instrumental in the expansion of the discipline and will promote the continued growth of medical informatics.

Medical Informatics (MI): Major Concepts, Methodologies, Tools & Applications

Joseph Tan
Wayne State University, USA

INTRODUCTION

Medical informatics (MI) is an inherently complex subject. The widening scope of its knowledge may be traced to the cross-disciplinary interactions of specialized knowledge domains – in particular, those of information sciences, computer sciences, and clinical sciences.

Over the last several decades, the field and sub-fields of MI – with its focus on clinical-based data and processes, and that of the even broader health informatics (HI) area (which embraces both clinical-based and health services administrative domain datasets and processes) – have now matured with the aim of achieving notable goals by transforming both complex organizational knowledge-based services as well as data-intensive, information-laden processes. These goals include improving physician-patient communications and relationships, promoting a higher quality of life, and achieving a more efficient and effective healthcare services delivery system.

With the first record of medical information came the germination of the MI concept. Its rapid diffusion is evidenced by the support it receives from an eclectic well of traditionally established medical practices, an increasing range of applied health data methodologies, and more efficient and effective clinical research tools and applications. Dissatisfied with the insufficient ways in which early forms of medical data were largely acquired and interpreted, academics and learned practitioners from diverse clinical and informatics disciplines have spurred the development of MI through their contributions.

Today, characteristic of its expanding scope, MI has branched into administrative health informatics, bioinformatics, clinical informatics, consumer health informatics, dental informatics, health management and services informatics, health sciences informatics, telematics, nursing informatics, pharmacy informatics, public health informatics, and veterinary informatics. A variety of names for MI have also arisen, including, but not limited to, computer-based medical information systems (MIS), methods of information in medicine, medical computer science, medical information processing, medical decisional models, medical computer technology, among other, closely related medical information systems/information technology (IS/IT) terminologies.

MI Definitions & Evolving Perspectives

Every MI student should be cognizant of MI's roots and the many attempts, over the years, to differentiate among the diverse perspectives of its developing sub-fields. Accordingly, a survey on early efforts to define the MI field and sub-fields can be extracted from the extant literature including archived, existing, and more recent MI-related journals such as *Angewandte Informatik*, *BMC Medical Informatics & Decision Making*, *Computer Programs in Biomedicine*, *Computers and Biomedical Research, International Journal of Bio-Medical Computing, International Journal of Electronic Healthcare, International Journal of Healthcare Information Systems & Informatics, International Journal of Medical Informatics*, *Canadian Medical Informatics, Journal of the American Medical Informatics Association, Journal of the American Society for Information Science, Journal of Clinical Computing, Journal of Healthcare Information Management, Journal of Medical Systems, Medical & Biological Engineering & Computing, Medical Decision Making, and Methods of Information in Medicine*. Owing to space limitation, the readers are asked to seek out further readings, as provided in the citations and other parts of this 4-volume work.

Primarily among countries influenced by Western Culture and Civilization, the conceptualization of MI began with a series of academic debates and educational efforts. These debates and efforts sought to provide substantive arguments for the preference of one school of thought over another, so that clinical students might be better educated, and to allow for the many aspects and identities of this germinating field to be differentiated. Indeed, the key definitions of MI found over the decades have provided us with a rich understanding of its different roots and perspectives. Among the earliest definitions, for instance, Collen (1977) conceived MI broadly as the computerization of medicine, including medical services, education, and research. Years later, van Bemmel (1984) defined MI on the basis of the knowledge and experience gained from processing and communicating medical and health care information via a paired theoretical-practical lens on the medical process. The Blois-Shortliffe (1990) description of MI largely evolved the *central hypothesis*, which is presented separately in this overview. Their biomedical perspective rendered MI operational by relating it to the storage, retrieval, and optimal use of clinical data, information, and knowledge for diagnostic, therapeutic, and prognostic problem solving. Stead, in 1998, referred to MI as a "science" on the nature, structure, collection, and use of health information. Most recently, Hersh (2002) defined the MI field as an integrative discipline that interweaves the applications of computational, cognitive, informational, organizational, and other sciences on the processes and use of clinical and biomedical information.

Put together, a cohesive definition of MI for the purpose of this work may be conceived as a *complex science that integrates relevant theories, design methodologies, and knowledge of best practices drawn from various cognitive, computational, informational, organizational, and other expert knowledge domains when applied synergistically to collecting, storing, organizing, manipulating, using, and disseminating clinical-based and health-related information.* Over the last two decades, we have seen the evolution of MI to a broadening diversification of sub-fields and major themes. Some of these sub-disciplines include, for example, health management information systems (HMIS) and/or health information management (HIM), health care information systems (HCIS) and/or health care information technology (HCIT), health decision support systems (HDSS) and/or clinical decision support systems (CDSS), telehealth and/or telemedicine, consumer health informatics, e-health, patient-centered e-health (PCEH), and m-health.

For the convenience of the readers, Table 1 provides a list of the acronyms mentioned in this review with its associated terminology listed for easy reference.

Table 1. Acronyms for MI Readers (©J. Tan, 2008. Used with permission)

Acronym	Associated Term
AHIMA	The American Health Information Management Association
AI	Artificial Intelligence
ATA	American Telemedicine Association
CCHIT	Certification Commission for Healthcare Information Technology
CDSS	Clinical Decision Support Systems
Cio/cto	Chief Information Officer/Chief Technology Officer
COSTAR	Computer Stored Ambulatory Record System
CPR	Computerized Patient Records
CSFs	Critical Success Factors
CTA	Cognitive Task Analysis
DEA	Data Envelopment Analysis
EBM	Evidence-Based Medicine
E-DSS	Expert Diagnostic Support System
EHRs	Electronic Health Records
EKG or ECG	Electrocardiography
ES	Expert Systems
GIS	Geographical Information Systems
GST	General Systems Theory
HCIS/HCIT	Health Care Information Systems/Health Care Information Technology
HCTM	Health Care Technology Management
HDSS	Health Decision Support Systems
HELP	Health Evaluation (through) Logical Processing
HHS	Health and Human Services
HI	Health Informatics
HIM	Health Information Management
HIMSS	Health Information Management & Systems Society
HIPAA	Health Insurance Portability and Accountability Act
HIS	Hospital Information Systems
HMIS	Health Management Information Systems
HMOs	Health Maintenance Organizations
ICT	Information & Communications Technology
IS/IT	Information Systems/Information Technology
JCAHO	Joint Commission on Accreditation of Healthcare Organizations
M-Health	Mobile Health
MI	Medical Informatics
MIS	Medical Information Systems
MS	Multiple Sclerosis
MUMPS	Massachusetts General Hospital Utility Multi-Programming System
NNs	Neural Networks

Table 1. continued

Acronym	Associated Term
ONCHIT	The Office of the National Coordinator for Health Information Technology
OOAD	Object-Oriented Analysis and Design
OSS	Open Source Software
PCEH	Patient-Centered E-Health
POMR/PROMIS	Problem-Oriented Medical Record/ Problem-Oriented Medical Information Systems
RMRS	Regenstrief Medical Record System
SA/SD	Systems Analysis/Systems Design
SDLC	Systems Development Life Cycle
SNOMED	Systematized Nomenclature of Medicine
SSM	Soft System Methodology
TMR	The Medical Record
TPR	True-Positive Rate
TNR	True-Negative Rate
u-health	Ubiquitous Healthcare
v-health	Virtual Healthcare
VR	Virtual Reality
VPR	Virtual Patient Records

MI History

Notwithstanding, "informatics" as a revolution in scientific methodologies may be traced to the 1946 invention of ENIAC, the first truly "electronic" computer used for high-speed mathematical tabulations during WWII. What followed evolved into the era of mainframes, where only the very large and well-funded hospitals of developed countries such as the United Kingdom, France, Sweden, Italy, Japan, Germany, Canada, and the US could afford to house/use these first-generation, large-scale computers that tended to fill entire floors. Processing computerized health data on a massive scale dictated that the dawn of MI thinking be characterized by the use of apparently sophisticated mechanisms, which could only be manned by highly-paid scientists and skilled professionals. During those early years, a key challenge was that of developing reliable and dependable health information and recording systems for hospitals supporting patient care and laboratory services. Eventually, specialized computer languages such as Medlars, Medline, SNOMED (Systematized Nomenclature of Medicine), and MUMPS (Massachusetts General Hospital Utility Multi-Programming System) were developed. These eventually led to the successful design and implementation of major computerized health record systems and databases, such as COSTAR (Computer Stored Ambulatory Record System), RMRS (the Regenstrief Medical Record System), TMR (The Medical Record), HELP (Health Evaluation through Logical Processing) system, and POMR/PROMIS (the Problem-Oriented Medical Record/Information Systems). Historical developments of these various systems play a paramount role in both the broader MI and the more specific hospital information systems (HIS) movements.

Indeed, functional hospital information systems soon became a reality when a growing group of hospitals across the US (including Akron Children's, Baptist, Charlotte Memorial, Desconess, El Camino, Henry Ford, Latter Day Saints, Mary's Help, Monmouth Medical Center, St. Francis, Washington Veteran's Administration, and others) as well as Europe (including Sweden's Danderyd Hospital and Karolinska Hospital, England's London Hospital and Kings Hospital, and Germany's Hanover Hospital) collaborated to advance the development of a prototype for efficient and effective management of health records. Despite the risk of major failures, these "pioneering" hospitals invested large sums of money, time, and effort to accomplish the mission of hospital computerization. With the surging interest expressed by these hospitals, and the potential market opportunities that arose from this movement, large computer vendors such as Burroughs, Control Data, Honeywell, IBM, and NCR began providing generous support for this development. Lockheed Information Systems Division, McDonnell-Douglas, General Electric, Technicon Corporation, and several other companies experienced in applying computers to the management of complex systems also joined in the effort. Ultimately, the Technicon system, which was initiated by Lockheed for the El Camino Hospital in Mountain View, California, and later bought over and improved by the Technicon Corporation under the leadership of Edwin Whitehead, became the "successful" prototype that laid the foundation for the majority of functional hospital information systems we know today – systems that gradually diffused throughout North America and Europe. The major lesson in the El Camino project was the significance of paying attention to users' information needs and of effecting a change in users' attitudes, especially if there is major resistance from key stakeholder user groups, such as physicians and nurses.

As evidence of continuing gains in productivity and efficiency were traced back to computerization, the early and mid-70's saw a growing diffusion of large-scale data-processing applications in medicine and health record systems among major teaching hospitals. As noted, COSTAR, RMRS, TMR, and POMR were among these early and successful prototypes. COSTAR, a patient record system developed at the Massachusetts General Hospital by Octo Barnett, was later extended to record patient data on different types of ailments, such as multiple sclerosis (MS-COSTAR). RMRS is a summary-type patient record system implemented in 1972, and TMR is an evolving medical record system developed in the mid 1970s at the Duke University Medical Center. POMR, developed by Lawrence Weed in the 1970s, is an important health record methodology offering a logical and group-based approach to medical problem solving. All participating members of the patient's healthcare team can easily follow through on the prescribed treatment protocol by linking the problem list (serving as the record's table of contents) in POMR to its database (comprising the history, physical examination, and initial laboratory findings), the initial plan (including tests, procedures, and other treatments), and the documented progress notes.

When minicomputers and microcomputers were introduced into medicine during the late 1970s and early 1980s, physicians and clinical practitioners quickly realized the speed and astounding harnessing power with which computers could process large and massive volumes of medical and other health-related data. This would improve not only clinical decisional efficiencies, but also its effectiveness. During this time, increasing interest in the application of artificial intelligence (AI) to medicine and health decision support systems (HDSS) soon created an explosive growth in the MI field. At the same time, major progress in MI was achieved by various attempts among clinicians to use medical computation for dentistry, radiology, pharmacy, nursing, as well as the incorporation of rule-based and frame-based algorithms into expert systems for the purpose of training resident physicians or less-than-expert clinicians. Notably, MYCIN and INTERNIST-1 are among the most popularly cited and well-documented clinical expert systems (Shortliffe et al. 2001). Altogether, the MI movement contributed heavily to the use of these automated record systems, as well as their applications of expert systems for training physician specialists and guiding decisions of less-than-experts. Eventually, it was only a matter of time before a considerable need for the diffusion of MI concepts into all the different sub-specialty medical areas would arise.

In summary, the MI discipline has continued to expand, with immense diversification and integration of concepts found in the medical sub-fields linking healthcare information management, medicine, computer technology, and information science. In the early years, established areas of MI concerned general medical informatics, clinical medicine decision making, biomedical computing, computing in biomedical engineering, nursing informatics, pharmacy informatics, dentistry informatics, hospital information systems, and education. Today, the extension of the MI concept is recognized by every discipline under the umbrella of health sciences that relates to linking rapidly advancing technologies to health and patient care services delivery.

At this point, it is important for the readers to glance briefly through the remaining sections of this overview chapter. Section II focuses on basic MI concepts, theories, and its central hypothesis. Section III highlights MI development and design methodologies. Section IV concentrates on MI tools and applications. Section V surveys MI major themes, utilization and applications, while Section VI shifts focus to MI organizational and social impacts. Section VII reviews key implementation issues and managerial impact of MI. Section VIII discusses critical issues, and Section IX highlights MI emerging trends. Section X concludes the entire exposition with a concentration on how the different underlying issues, themes, and emerging trends may be joined together to achieve effective medical informatics services, education, and research in the coming years.

BASIC CONCEPTS, THEORIES & CENTRAL HYPOTHESIS

Informatics basically refers to advances in the information sciences, a discipline which has accumulated an impressive array of methods for data, information, and knowledge processing. Informatics, when applied to medicine, combines human and computer elements to generate timely and relevant patient data, organized medical information, and accumulated knowledge, which, together, aids clinicians in retrieving, storing, analyzing, disseminating, and sharing diagnostic, therapeutic and prognostic decisions, treatment planning, and evaluation processes. Other commonly encountered MI terms include health informatics, clinical informatics, health information systems, health information technology, health information management, medical information science, as well as medical (health) informatics and telematics.

Health Data, Clinical Information & Medical Knowledge

In MI conceptualization, health data are subtly differentiated from clinical information and medical knowledge. Whereas the former emphasizes some form of source-gathered raw fact, such as recording a patient's medical history before generating any meaningful information processing, the latter terms entail converting data into organized clinical information-knowledge. Hence, the significance lies in the application of readily available methods or tools to produce accessible, appropriate, accurate, and aggregated information-knowledge from sourced data. At the level of converting information into knowledge, the same rules apply. With more complex methods or cognitive-based techniques, "wisdom," "rules of thumb," "probabilities," and other "likely associations" – all of which aid pattern recognition and future health-related decision situations, under increasingly complex or uncertain situations – are produced from seemingly unrelated information. Medical knowledge, in this sense, refers to the cumulative experience of applying useful clinical information management techniques to yield timely and significant decisions.

In essence, health data are specific facts and parameters. A good piece of datum is characterized by its accuracy, completeness, relevance, reliability, security, and timeliness. Accuracy is achieved when health data recorded are true, precise, and valid about the status of a patient's condition; for example, a temperature of 102°F recorded as 101°F is approximate but, nevertheless, inaccurate. Completeness dictates that all required health data should (and must) be recorded; for example, a unique identifier must exist in the patient master index (PMI) for each patient recorded in a database in order to differentiate among individual patients. Relevance ensures that appropriate and necessary data is made available to authorized clinical staff whenever and wherever necessary; for example, cardiologists should be able to view their patients' electrocardiography (EKG or ECG) reports in order to monitor the progress of their patients' heart conditions. Reliability requires that health data recorded are trustworthy and consistent; for example, if the allergy list of a patient exists in a hospital food services system, the same list should also be accessible from, and appear in, its pharmacy system. To ensure that patient data confidentiality is not compromised, and to safeguard against potential data misuse, security and privacy regulations stipulate that only designated persons with valid access rights can view or make authorized changes to recorded data. Last, but not least, timeliness ensures that the available health data are current and accessible for the decisions and tasks at hand, especially when they are of a crucial nature, such as in life and death situations.

Clinical data are unique and somewhat different from the financial and accounting data that are captured in routine business transactions. Clinical data are typically obtained in many different forms, so that the episodes and/or trajectory of care a patient encounters may be generated on a longitudinal basis. Even so, data about a patient may be entered by different clinicians, in different departments, and at different times to show the progress of the patient's wellness recovery: for example, a nurse may jot down the demographics, weight, temperature, and blood pressure measurements of a patient arriving at the clinic; then, through lengthy progress notes, the physician and/or other specialists would record their observations about and diagnosis of the patient's symptoms and complaints. While some of these data are textual in nature, quantitative analysis, in many instances, can also be made by comparing certain measurements taken of the patient to those from a relevant cohort of patients. Moreover, the physician and/or specialists may send the patient for further laboratory tests and other diagnostic scans, which will offer additional measures or markers. Subsequently, these data can be fitted with various biostatistical and/or complex computational models, tabulated and/or presented graphically to check for familiar patterns, and further digitized for convenient sharing among other clinical experts for the purpose of second opinions or referral interpretations. Indeed, it is the accumulated knowledge, embedded in the different biostatistical and mathematical models, that these expert medical consultants rely upon for accurate pattern matching when performing further diagnostic, therapeutic, and prognostic decisions about the patient. Thus, all the while, from admitting to discharging a patient, this cycle of data-information-knowledge processing is repeated continuously to enhance the art and science of MI.

In practice, every clinician is also a health informatician. MI starts with the collection of meaningful, accurate, relevant, complete, and usable datasets. While health databases must be properly organized and rendered accessible and available to health informaticians in a timely fashion, it is also critical to consider and understand, prior to the data gathering process, the requirements that these databases pose. Otherwise, managing and maintaining inappropriate or unnecessary data, especially in large medical databases, may wastefully drain away valuable and critical organizational resources.

Users of medical and health-related data range from patients to care providers, government agencies, health care planners, judicial agents, educators, researchers, and third-party payers. Different types of users may also require different scopes, formats, and presentations of data. To achieve best practices in MI, a full comprehension of the central hypothesis underlying the MI philosophy is critical. However, before discussing this core informatics ideology, we will first survey some of the key MI theories.

For years, MI researchers and academics from across the globe have attempted to etch an identity for the MI field. Yet, the evolution of its many definitions and perspectives, the explosive growth of the field and the continuing diversification of its sub-fields made it a very difficult task to cumulate a tradition for MI theories, methods and practices. Nonetheless, it may still be argued that MI conceptual foundations are rooted in several generally cited theories: the general systems theory, decision analytic theory, various human information processing, problem-solving and cognitive theories, as well as relevant social and organizational theories such as the theory of innovation diffusion, theory on management change, theory of interpersonal behavior and technology acceptance model. Due to space limitation, we will only cover a few of the more popularly cited theories.

GST

General Systems Theory (GST), or "Cybernetics," is a cornerstone of MI development. GST has played a key role in many other contemporary sciences during the emerging information era (Bertalanffy, 1968). The theory begins with the empirical observation that every "system," regardless of its disciplinary domain, shares some common characteristics in its underlying structure such as system synergy and system triad, and exhibits some similar behavioral patterns such as statistical constancy, growth, decay and/or other known (unknown) patterns.

All systems have objects and attributes. Objects constitute the components of a system, whereas attributes are the properties associated with these objects; essentially, an attribute is an abstract descriptor that characterizes and defines the component parts of the system. A person is an object and the attributes may be the person's demographics. A system combines all the objects and their attributes and defines the relationships among these objects, thereby enabling the different parts to add up to some greater whole than all of its individual parts. For example, the emergent property of a university system is more than just its faculty, students, classes and degree programs because of the many interfaces it coordinates among the objects related to the university, the community it serves, and the various donors and funding agencies that support the university's existence.

More generally, a system is a set of interrelated elements. An open system is one that interacts with its environment whereas a closed system does not. The structure of a system may involve a hierarchy of embedded subsystems, each having its own unified purpose that contributes jointly to the functioning of the larger system. The functioning of these subsystems can also vary in complexity. The simplest process involves a triad: an input, a process and an output. More complicated functioning may involve a series of conversion processes, positive and negative feedback mechanisms, and the channels through which the environment can exert its influence. Viewing the health service delivery industry from an open systems perspective will therefore provide valuable insights into the functioning and structure of the contextual system for MI technologies and applications. As such, understanding the application of GST on health services delivery systems can better define the role and function of the MI applications.

In designing an expert system for diagnosing problems of lower back pain (see Lin Lin et al. in Tan, 2005), for example, objects of the system may include the patients, the therapists attending to the patients, and the expert diagnostic system itself. The attributes describing the object "patient" may include the patient's demographics, a brief description of the pain patterns experienced by the patient, the cause of the injuries leading to the pain, the level of pain and an estimated length of time the pain has persisted. Moreover, the treatment plan for the pain to be administered daily may vary depending if the patient is highly motivated to get well, the potential for the patient to move and be tolerant of the pain, the patient-therapist relationship, whether or not there is a job action among the therapists at the time of the treatment administration, or if there has been a sudden snow-storm, making it difficult to transport the patient to the treatment facility. In general, it is expected that those therapists using the expert diagnostic support system will have gained a better judgment over time to guide the patient treatment protocol to reach some level of statistical stability or equilibrium than those not aided by the system, if the expert system is to be considered functioning efficiently and effectively.

A major body of MI knowledge and research is therefore the application of GST in the context of patient treatment planning and the development and use of computerized decision aids to obtain valuable insights into the behavior of complex, real-world systems. As computerized decision models are attempts to imitate systems from a particular viewpoint, and clinicians (or groups of clinicians) who run mostly through rational decision-making mechanisms are sub-systems in the larger context of MI user-computer systems, the close relationship between computerized models and clinicians cannot be overly emphasized. Such computerized aids are primarily supported by the application of decision analytic theory, which is discussed next.

Decision Analytic Theory

Decision analytic theory has also been widely applied in MI to provide logical rationalization on relatively complex judgmental tasks, for example, the application of well-tested decision rules such as Bayesian technique to ease computation and reduce uncertainties about clinical guidance on whether to proceed or not proceed with a certain form of therapy or surgical intervention for a certain patient. Its appeal in MI is, therefore, its ability to provide the clinical decision makers with a model, such as a decision analytic or computerized model, that would yield a more accurate and rational representation of the clinical case reality. In the MI context, the application of decision analytic theory on clinical decision problems has been largely defined by the use of programmable models to reduce complexities, and achieving a decision outcome that lessen its influence from that of mere human emotion and biases.

When incorporating programmable and semi-programmable models for MI applications, a taxonomy of models along a decision complexity continuum, including, but not limited to, decision analysis techniques, mathematical programming, computer simulation models, heuristic programming, and non-quantitative (qualitative data) modeling should be considered. Incidentally, these models are not intended to replace the decision makers, but to serve as aids to the clinicians in rationalizing their decision-making process and justifying their final choices. For example, if a clinician is armed with a set of data that contains information about the probability of the onset of multiple sclerosis (MS) for a particular patient, it may be important for the clinician to consult with a decision aid to guide the prognosis before jumping into conclusion based simply on short-sighted self-evaluation of the complex dataset. Hence, the term “decision support tools” is often used.

Decision analysis is popularly used to aid clinical decisions under uncertainty and risk. Two simple examples are the use of decision tables and decision trees. The computation essentially generates the expected outcome of each alternative among a given set of alternatives on the basis of available (unavailable) information about the environment and converts the information on uncertainty into risk estimates by means of Bayesian technique, Markov chains or other more esoteric mathematical or probabilistic models.

In situations where there can be many more alternatives and it is not possible to generate a manageable set of alternatives, mathematical programming takes the approach that reality can be simplified and represented as a set of mathematical equations or relationships. These relationships represent the constraints and limitations on the number of inputs as well as the relationships between inputs and the outcome variables. The commonly employed linear programming technique used in providing optimal solutions to many well-structured, mostly single goal (criteria) problems is a simple illustration of mathematical programming. Other more sophisticated mathematical programming techniques used to solve complex, semi-structured, multi-criteria problems include non-linear programming, dynamic programming, 0-1 programming, and data envelopment analysis (DEA).

If the complexity of the problem situation increases to the extent that the relationships among the variables cannot be conveniently simplified into a series of mathematical relationships, then computer simulations, certainty factors and stochastic modeling may be used. In "survival analysis" where prognostication is to be predicted by a clinician, for instance, methods such as life tables, Kaplan-Meier analysis, and Cox regression may be used. In a computer simulation, either the real distribution of variables can be used, or a probabilistic distribution may be applied to model the variable distribution to be simulated. In this respect, using simulation to model reality can allow the relationships between variables to be kept either very simple or closer to reality. Moreover, data that have been collected in the past may be used to test and validate the simulation model. When the validity of the simulated model can be demonstrated, further experiments can be constructed to compare various alternatives. These experiments have the advantage that time can be compressed significantly, allowing several months or years to be modeled quickly within a single simulation run that may last only a few minutes. In contrast to mathematical programming, however, only good enough ("satisficing") solutions rather than optimizing solutions can be expected from the use of computer simulation models.

In highly complex situations where the problems are considered somewhat ill-structured and even simulation cannot be applied, heuristics or rules-of-thumbs such as rule-based systems, frame-based systems as well as neural networks (NNs) methodology are often employed by decision makers to aid problem solving. These heuristics may be incorporated into a computer model of the situation; thus, the term, heuristic programming. An example of heuristic programming is the use of expert methods such as neural computing (networks). NNs are experimental computer designs that purport to build intelligent algorithms. NNs operate in a manner modeled after our human brains, in particular our cognitive ability to recognize patterns. Another class of heuristic models is that of genetic algorithms. Genetic algorithms randomly generate initial solutions to specific procedures, which can then be further recombined and mutated at random just as in an evolutionary process to produce offspring solutions that may yield better offspring and parent solutions or new algorithms. For example, a set of generic operators can be used to generate specific procedures for developing routing and scheduling heuristics for solving an ambulance dispatching problem. These generic operators can then be stored for generating new algorithms. In this way, new customers can be added, routes can be merged, and the sequence of customers can be modified using different sets of generic operators. A form of visual interactive modeling can then be used to allow the user to see the results and intervene to change the procedure if the results are not as experienced. Other examples of heuristic programming include the use of fuzzy logic, case-based reasoning and rough-set methods since these techniques often incorporate experts' heuristics in generating problem solutions.

At the farthest end of the decision complexity spectrum lie the non-quantitative (qualitative analysis) models. The field of non-quantitative analysis is very young and there is a need for considerable research to examine the applications of different approaches and use of computerized models for such analysis. The rationale for clinicians to employ MI thinking such as using an ES, AI or even human-controlled robots instead of relying on knowledge and/or experience gained through the traditional case-by-case bedside training is to achieve a high level, combined human-automated intelligent symbiosis in the art of using and interpreting medical evidence.

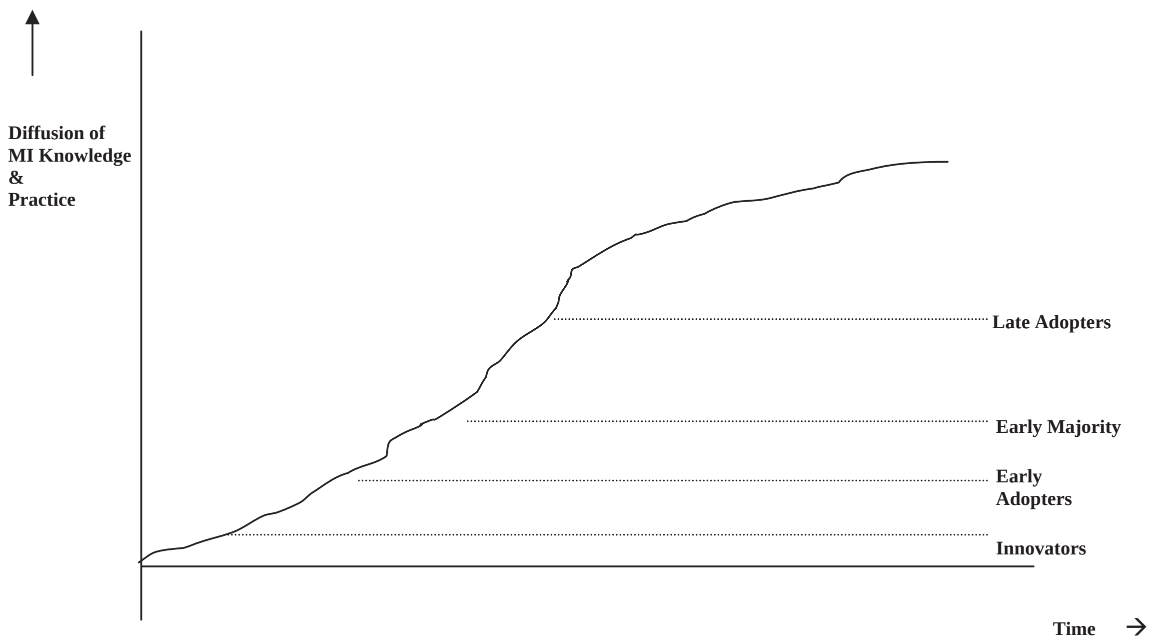
Social & Organizational Theories

As with any emergent discipline, the need for innovation, continuing growth and development encourages MI researchers and practitioners to draw selectively from theories proposed and tested in many related but previously established disciplines. Specific to MI diffusion strategies, the social and organizational theories have provided insightful guidance for MI research and practice.

In the context of the innovation diffusion theory, for example, the rate and patterns of MI growth as a discipline, in and of itself, are expected to follow the well-known S curve of innovation diffusion (Rogers, 1983). The innovation diffusion growth pattern typifies the stages of adoption of knowledge and practice in a new field. For instance, the researchers who adopt a new form of analytic method on the basis of their own recognition and awareness of the desirability of this method as applied to similar problems identified in another more establish field will adopt it before others in the MI research and practice community. The first adopters of the new model are known as the innovators. Others who follow will then become early adopters. They will not have access to all of the information that the innovators have about the new model so that they often tend to regard the innovators as experts and will readily adopt the practices advocated by these experts. An additional link to this chain is classified as the early majority, who are then followed by late adopters.

Figure 1 shows the diffusion pattern of MI knowledge and practice over the last few decades. As shown, we are entering an era between the so-called early majority and the late adopters for the diffusion of MI knowledge and practice. This implies that the continuing growth and diffusion of MI knowledge and practice in the coming years is expected to still be relatively high as more and more scholars from various clinical sub-specialties enter the MI discipline. In other words, it appears timely for us to begin consolidating our past knowledge and experience in the MI field and to identify key gaps for future MI research and practice. The release of this particular work is a testament to the rapid growth and expansion of MI field in recent years.

Figure 1. The diffusion of medical informatics (MI) knowledge and practice (© 2008 J. Tan. Used with permission.)



l

As well, the diffusion theory further implies that successful MI implementation can be achieved through critical support provided by a champion who could act as the innovator or expert, followed by grass-root user groups such as nurses and physician residents (who are willing to become early adopters) to break down the resistance that may be inherent to medical technological innovation and the additional need for attitudinal and behavioral change towards such an innovation. As a critical mass of user support gathers strength, success is eventually achieved due to the proliferation of early majority and late adopters.

More specifically, there have been recurring themes and increasing interests among both the academic and practitioner communities on the potential impact of effective technology management strategies among healthcare organizations and modeling the technology acceptance process. In their contribution on Health Care Technology Management (HCTM), Eisler, Tan and Sheps (in Tan, 2008) took the perspective of technology as the *extension of human and organizational ability*. Based on the results of their study, they found that the major critical dimensions for successful HCTM include the following: (a) Strategic Management factors; (b) Management of Change and Innovation factors; and (c) Organizational Management factors. Among these factors, the functions of a chief technology officer (CTO) were found to be have the largest differences in reported perception based on implementation ratings and gap scores between high performing and low performing teaching hospitals. This result strongly confirmed the message from the literature about the necessity of executive attention and leadership for HCTM. Besides leadership, coordination, and facilitation, the responsibilities of the CTO include such activities as gatekeeping, advocacy, funding, sponsorship, policy and procedure development, promotion, capacity building, and overseeing the technology management system.

Another set of organizational factors that were shown to be critical for successful HCTM include identifying customer needs and priorities; the organizations having effective strategies to respond flexibly and readily to technological change; and the routine evaluation and benchmarking of organization's performance as a function of technology management activities. In summary, organizational theories indicate that there are key differences in HCTM sophistication among healthcare institutions. Major differences occur in areas of strategic technology management, followed by change management, and to a lesser extent, organizational management.

At this point, we would like to steer the readers to a key conceptualization on MI that brought about the informatics revolution in medicine. We called this, the MI "central hypothesis."

The Central Hypothesis

The informatics revolution in medicine during the 1960s and the 70s brought about a surging interest in MI, especially among biostatisticians, public health advocates and epidemiologists, occupational health and environmental scientists, radiologists, general physicians, dentists, pharmacists, nurses and other health professionals. Essentially, many of these clinicians were questioning the use of medical information/knowledge derived merely from the traditional bedside training paradigm. This eventually led to a movement in support of more intelligent applications of effective computerized decision aids that incorporated complex statistical techniques and probabilistic models for improved medical information analysis. In the 1980s, evidence-based medicine emerged to further guide medical decisions and substantially changed the traditional approach to medical reasoning.

Moreover, one should familiarize oneself with a number of commonly used terms in generating diagnostic, therapeutic and prognostic decisions in MI. For instance, public health informaticians describe the impact of diseases on a specific population using terms such as incidence and prevalence rates. "Incidence rates" refers to new occurrences in terms of the number of people newly infected with the disease, usually over a year whereas "prevalence rates" refers to the proportion of people having already been infected with the disease at a particular point in time relative to the size of the population who have already been exposed to the disease.

When testing a particular patient for a specific disease, the sensitivity-specificity dimension of the test is another pair of closely related terms used in MI. A test is said to be as sensitive as it is able to identify those people who truly have the disease (*true positive rate*) among those who have tested positively while a test is said to be as specific as it is able to rule out those people who truly do not have the disease (*true negative rate*) among those who have tested negatively.

Related key concepts used in determining the health status of a patient in diagnosing for the presence of a disease is the positive vs. negative predictive values as derived from the sensitivity-specificity dimension of the test. On the one hand, positive predictive value uses the underlying concepts of incidence-prevalence rates and sensitivity of the test to help a clinician predicts with some assurance the likelihood of the patient having been infected by a disease. On the other hand, negative predictive value uses the same concepts of incidence-prevalence rates and specificity of the test to help a clinician estimates more confidently the increased likelihood of the patient not having been infected by the disease being tested.

When determining the health status of a person with respect to a certain disease, clinicians must therefore recognize that each medical diagnosis is, after all, only an estimation process. What is important is to determine if more tests are truly needed to increase the likelihood estimates and if particular treatments are to be planned and evaluated. In most instances, the patient is subjected to a range of medical tests with the intent that the clinician will arrive at an increasingly comfortable measure of likelihood that the patient will indeed have (or not have) the disease. This, in turn, ensures that the appropriate planning is adopted based on the best treatment protocols available and that the patient will actually benefit from the treatment regiment, considering the fact that the patient indeed do have the disease. In medicine, both false negatives and false positives are costly, not only to the clinicians but even more so to the patients. This is where MI plays a key role and can have a significant impact.

The central hypothesis of the MI philosophy, thus, is that the entire cycle of medical information gathering, storage, analysis, comparison and estimation from one medical test to another can be better managed through the intelligent application of an "informatics" ideology. If this goal is achieved on an ongoing basis, both the clinician and the patient can ultimately contribute to a higher quality of life and an improved wellness to the society. A deep appreciation of this central hypothesis is critical for those yearning to be next generation medical/health informaticians.

MI DEVELOPMENT & DESIGN METHODOLOGIES

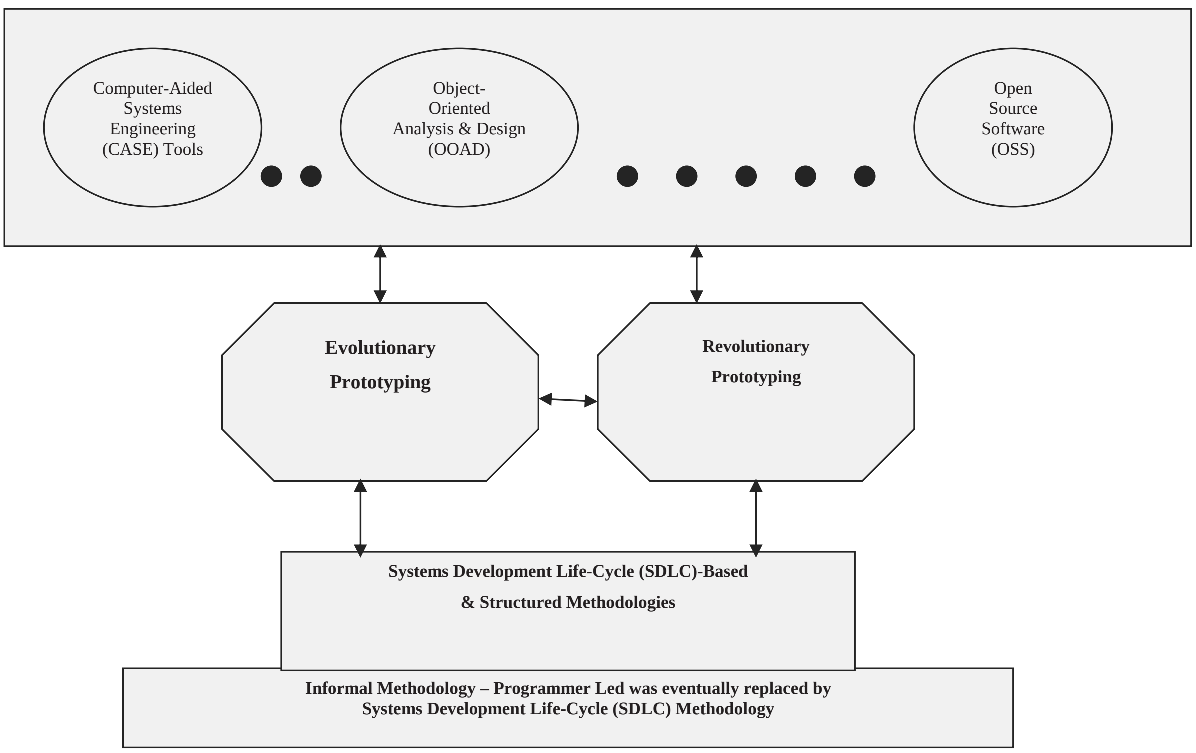
As we have noted previously, a high failure rate of MI deployment has been attributed largely to the lack of attention on the applications of social and organizational theories, in particular, as these theories relate to MI development and design context. It has been estimated that medical systems have an almost fifty percent higher chance of failure than most other information systems deployment due to the specific nature of resistance faced in health care – for example, both physicians and nurses have their own professional cultures and belief systems that may not be aligned with those of managers, systems analysts and informaticians. Hence, new methods for system design and evaluation in this area are badly needed.

Essentially, MI development and design methodologies should bring together the different pieces of the MI puzzle – data, people, hardware and software, network and other resources – towards achieving a vision of an effectively deployed MI systems. Indeed, the challenges of working with physicians who not only have their own professional culture, but may also be considered as much customers as employees of the health care institution in which they practiced is a critical issue for MI technology diffusion. To realize the vision of an accepted MI system throughout the health care institution, the informatics analyst must therefore be able to create a shared vision among key stakeholders with unwavering support from top management. Acceptance of the MI system from key users through active participation in the MI design and development process will be paramount, especially for physician and nurse users.

Once the specification of the system has been fully vetted by the relevant stakeholders, the next step will then be executing a set of related but critical activities including systems analysis, design and evaluation. Systems analysis relates to activities involving the review of current information architecture and the organization needs from the users' perspectives. In contrast, systems design relates to activities involving the specification of new information architecture and systems requirements. Finally, systems evaluation relates to activities involving the continuing testing and endorsing the designed MI system – that the designed MI system is exactly what is needed and will be used once implemented.

The early period of MI development and design has been characterized by the absence of any formalized system development methods. Typically, it was the programmer who bore the major responsibility. Over the years, as user needs became more complex, the need for a formal development process became evident; subsequently, various design methodologies were championed to help minimize the problems of uncoordinated MI development. Each methodology has been based on a philosophical view, which can range from an exclusive focus on the technical aspects to a focus on the user side. Most prominent among the MI development methodologies include the traditional systems development life cycle (SDLC) approach, the structured techniques, prototyping and more contemporary models as shown in Figure 2.

Figure 2. The evolution of medical informatics (MI) design & development methodologies (© 2008 J. Tan. Used with permission)



The SDLC approach involves following a rigid set of procedures beginning with feasibility study. This is followed by systems investigation, analysis, design, and evaluation. The final step entails systems maintenance and review and the entire life cycle is then repeated. While this approach was a significant improvement over the informal leaving-it-to-the-programmer attitude, it nevertheless failed to marshal strong support and acceptance of MI systems among key stakeholders often because of the lack of active user involvement throughout the analysis, design and evaluation phases of the process.

Structured techniques provided an entirely new perspective to systems development with a greater concentration on systems analysis (SA) and systems design (SD) phases – phases where input from end-users are most critical to the eventual success of the implemented system. Use of diagramming techniques, for example, data flow diagrams to depict the data flows between one process and another helps to bring in the views of the users. Other methodology, for example, Structured Systems Analysis and Design Methodology, extends the soft system methodology (SSM) and emphasizes the analysis and design stages of the SDLC model to better capture the views of the end-users.

Nonetheless, both SDLC-based and structured techniques required that users are able to specify in advance the information requirements in the MI system. However, users are often unable to verbalize what they want, and even if they do, their wants may not exactly reflect their real needs. This is evidenced by the many revisions most newly developed MI systems must undergo before these are finally accepted. One philosophical approach that addresses this problem creatively is rapid vs. evolutionary prototyping.

In evolutionary prototyping, the proven "SDLC" and structured methodologies are essentially incorporated into prototyping merely to fine-tune the MI development process, to engage greater user participation, and to enhance user acceptance of the final product. In rapid (revolutionary) prototyping, the designer often uses a creative trial-and-error process to generate an initial prototype quickly. In this instance, fourth generation languages are often applied to produce these rapid initial prototypes. The users as well as programmers are then encouraged to test out, validate and fine-tune these prototypes. Such iterations are then repeated until a final acceptable MI product is achieved.

Today, the evolution of MI design methodologies has moved from the rigid procedural SDLC-based and structured approaches to rapid prototyping to more contemporary models. Prime examples include Computer-Aided Systems (Software) Engineering (CASE) tools, Object-Oriented Analysis and Design (OOAD) and Open Source Software (OSS). CASE tools automate different parts of software or systems development and can assist with any or all aspects of the SA and SD processes. OOAD focuses on the objects that are acted upon in the development process. The methodology is based on the premise that software should be developed out of standard, reusable components wherever possible. Finally, OSS supports the ideology of making certain licensing of the source code for particular working prototypes freely accessible and generally available for other intending users' adoption and modification within the programming community.

As the number of OSS products in the MI field such as VistA system (a computerized patient records system developed in MUMPS by the US Department of Veteran's Affairs), OpenEMed (another patient record system), and OSCAR (a family practice office management and medical record system) becomes popularized over time, the OSS model of software development have continued to spread in the MI community. Indeed, recent studies (e.g, Erickson et al. 2005; DeLano, 2005; Scarsbrook, 2007) have shown that the OSS development model can be beneficial in reducing software development costs effectively and allowing for very rapid scientific advancement due to the open sharing of information and software as a way to overcome certain barriers of standardization. A case where OSS adoption is largely responsible for a number of low-cost products implemented successfully throughout the hospital is the Beaumont Hospital in Ireland (Fitzgerald & Kenny, 2004). OSS adoption has also been noted to reduce the need for frequent vendor turnover, lower development costs, and lessen the impact due to the lack of standards affecting electronic patient records in the primary care area (Kantor et al. 2003).

Altogether, the application of social and organizational theories and the adoption of a practical approach to MI design and development that integrates the different user characteristics, technological and organizational components within the larger context of technology change management, evolving user demands and new organizational arrangements will ensure the successful implementation of future increasingly complex MI systems.

MI TOOLS & APPLICATIONS

In this section, our focus on MI tools and applications will encompass the emergence of the evidence-based medicine (EBM) movement in the 1980s, which has significantly changed the way in which medical information is now used in routine healthcare decisions as discussed in Tan (2005), a taxonomy of the decision-aiding strategies or techniques as advocated by Zachary (1986), and the expert knowledge elicitation methods as discussed in Tan with Sheps (1998). Throughout these discussions, the concepts of the electronic medical records (EMR) and the electronic health records (EHR) will be used to illustrate the applicability of these tools.

Evidence-Based Medicine

Evidence-based medicine practice requires the clinician to articulate a clear statement of a patient's medical problem, to screen actively through the extent literature relating to the patient's problem, to evaluate the evidence critically, and from the resulting knowledge, to make the best treatment decisions with regard to the patient's problem.

Markovitz in Tan (2005) described EBM as a five-step process: (1) identify the "clinical question" relevant to the patient by means of asking the following question: "In a given patient population, does a particular intervention, compared to controls or standard therapy, result in an improved outcome?"; (2) use bibliographic databases such as MEDLINE which can be accessed through convenient interfaces like OVID (http://www.ovid.com) and other EBM resources on the Internet to search for problem-related evidence; (3) critically appraise the findings, validity, and applicability of the evidence; (4) incorporate the appraised evidence into the values, preferences and beliefs uphold by the patient; and (5) self-evaluate the process on a continuing basis to further enhance its efficiency and effectiveness. With a wave of advances in biostatistical and clinical research methods, EBM can be achieved typically from sourcing multiple information such as relevant clinical practice guidelines, clinical trials, systematic reviews, and meta-analysis of past studies found in high quality, peer-reviewed journals that focus on the patient's problem.

To illustrate, a trained clinician who is asked to check on the EHR or EMR of a patient (with all the annotated progress notes) should also be expected to be familiar with the terminologies of clinical epidemiology, biostatistics, clinical trial methodology and clinical research designs in order to confidently perform a critical appraisal and rating of the published literature as pertaining to the health status of the examined patient. In many instances, in order to develop acceptable and trusted treatment guidelines for best practices in MI for the patient, the trained physician may consult with a committee of scientists and expert clinicians who may have collectively rated the strength of recommendations for a particular treatment based on the evidence for efficacy or for adverse outcome supporting a recommendation vis-à-vis the quality of evidence supporting the recommendation. The National Guideline Clearinghouse (NGC: http://www.guideline.gov/), sponsored by the US Agency for Healthcare Research and Quality, offers a trusted/authoritative source for such guidelines. In addition, the trained physician may also consult with an increasing number of full-text or summary form guidelines drawn from various professional societies and/or organizations that have used a documented EBM approach have been included and made available through this unique site.

In the realms of medical research, the following is a sampling of key terms that should be clearly understood as these form the building blocks of many research designs and are the basis for a proper evaluation of the MI research literature: informed consent, the internal review board (IRB), debriefing, bench vs. clinical research, prospective vs. retrospective research, control group vs. treatment group, focus groups, blinding vs. double-blinding, randomized control trials, cohort or longitudinal studies vs. case-control studies, attrition, sampling frame, sampling size, power, effect size, statistical vs. clinical significance, qualitative vs. quantitative research, and many other terms (e.g., relative risk, relative risk reduction vs. absolute risk reduction), all of which can be easily found in a standard text on health informatics (see Jordan, 2002) or research methodology (see Kerlinger, 1973;1986). Table 2 provides a glossary of key terms in MI research.

A more detailed explanation of these terms is outside the scope of this review. At this point, we turn to the Zachary (1986) taxonomy of cognitively based decision-aiding strategies.

A Taxonomy of Decision-Aiding Techniques

Several decision-aiding tools and techniques in MI and other healthcare informatics domains have become popular among MI analysts and researchers. A taxonomy of these ranges from information control techniques to process models and/or choice models to representational aids to analysis and reasoning methods to judgment refinement and amplification. Outside of the scope of this discussion would be more complex MI tools and new methodologies such as rough-set analysis, neural networks, and fuzzy-set theory, all of which have been presented in great detail in Tan with Sheps (1998).

Information control techniques involve the storage, retrieval, and organization of data, information and knowledge. Common examples include clinical databases such as an EMR or an EHR or even a comprehensive pharmaceutical database that will permit a physician to look up on comparable prescriptions for use in treating a diagnosed illness. A model-based management system which incorporates the application of a data envelopment analysis (DEA) methodology, for example, to profile the relative efficiency and effectiveness of best practices among a group of selected physicians awaiting the renewal of certain hospital privileges or a knowledge base management system such as an expert system programmed to provide online medical consultation to a physician resident on a specific clinical intervention procedures when faced with a complex MI decision may well be added as enhancement to the basic functioning EMR, EHR or other health database management system.

Process and choice models are computational models that may be applied respectively to help predict the behavior of real-world processes such as through the use of “what-if” capabilities or to solve multi-criteria problems with such techniques as those involving the integration of individual criteria across different aspects and/or alternative choices, for example, the application of multiattribute and multialternative utility models. In other words, these models could be programmed as simply as a single Excel function with “what-if” analysis or could be enriched with a series of increasingly complex computational models. Use of these models can be enriching for physicians who need to make complex decisions beyond just abstracting information from an EMR or EHR.

Representational aids refer to expressions of manipulation of specific representations of decision problems. For example, a visual representational aid could be graphics-based where the clinician reading a set of EMR and EHR progress notes is supplemented with supporting evidence from digital radiological images and reports to help interpret the captured information on the patient wellness recovery. At the

Table 2. Glossary of Key Terms for MI Research Literature (©J. Tan, 2008. Used with permission)

Term	Definition
Attrition	The number of participating subjects dropping out of a study
Bench Research	Laboratory research performed on chemical or biological elements such as cells, genes, and other cultures rather than living human subjects
Blinding	Participants are ignorant of their group treatment assignments during a study
Case-control Studies	The use of naturally occurring vs. control groups to classify study participants
Clinical Research	Research conducted with living human participants
Clinical Significance	The applicability and importance of a study's findings in real-world situations
Cohort Studies	Group (cohort) assignments are not randomized but based on known exposure
Control Group	A group that does not receive the test stimulus but is observed for comparison
Debriefing	A process to correct subjects' wrong/unethical perceptions at end of a study
Double Blinding	Both participants and investigators are ignorant of study group assignments
Effect Size	The degree for which a study variable impacts on the study outcome(s)
Focus Groups	Recruited subjects to draw opinions/observations on various research artifacts
Informed consent	Let subjects know of the study's purpose, risks and credentials of researchers
Internal Review Board (IRB)	An institutional ethics committee to evaluate the potential benefits vis-à-vis the risks of a proposed research
Power Analysis	Method to guide study design/sampling so as to achieve a desired effect size
Prospective Research	A study where subjects' data are followed through "looking forward" in time
Qualitative Research	Research methodology that reflects an objective and positivist legacy
Quantitative Research	Research methodology that reflects a subjective, interpretative approach
Randomized Control Trials (RCT)	Relatively short (prospective) clinical experiments that use randomization as a means of group assignments and controls, requiring lengthy IRB approvals
Retrospective Research	A study where data on subjects are examined by "looking backward" in time
Sampling Frame	The population or list of objects (sampling units such as persons) in a sample
Sampling Size	The number of subjects needed in a study to attain valid scientific evidence
Statistical Significance	Level to which study sample results mimics results expected in a population
Treatment Group	Participants assigned into a group who will be given the intervention studied

community level, a geographical information system (GIS) to depict specific population health hazard distribution or an epidemic among neighboring communities of a county would be a spatial representation of the captured environmental and population health data. Other representational tools could include matrix data and model representational methods.

Analysis and reasoning methods are means of performing problem-specific reasoning on the basis of a representation of a decision problem. For instance, in abstracting information from an EMR or EHR on a patient diagnosed with multiple sclerosis, the clinician may use the symbolic reasoning embedded in an expert system to aid him or her to make a prognosis for this patient. Another example would be the application of mathematical programming techniques for clustering various cases from different cohorts of patients exhibiting similar symptoms and requiring comparable treatments.

Finally, judgment refinement and amplification techniques are formalization and quantification of heuristic judgment processes, for example, bootstrapping and Bayesian theorem application. As discussed throughout Tan with Sheps (1988), medical informaticians have typically employed these analytic tools and techniques as well as automated intelligence applications to enhance decisions rather than simply make a judgment based on EMR or EHR data. Furthermore, these applications can be useful for various clinical areas such as cancer, emergency medicine, utilization review, population health monitoring, and nursing. This brings us to knowledge elicitation methods, a key MI tool and application in building and designing expert systems.

Knowledge Elicitation Methods

Expertise in MI and other healthcare informatics domains (e.g., health information management, nursing informatics, and pharmacy informatics) can generally be grouped along a continuum ranging from laypersons to experts. On the one hand, a layperson may be considered simply as someone who has only common sense or everyday knowledge of the domain. An expert, on the other hand, is someone who has gained specialized knowledge of the domain after many years of training. In between these extremes are beginners (novices), intermediates, and sub-experts. Beginners/novices are those with pre-requisite knowledge assumed by the domain; intermediates are, by default, those whose domain knowledge are just above the beginner level but just below that of the sub-expert level while sub-experts are those with generic but not substantive, in-depth knowledge of the domain. In this section, we discuss approaches to the knowledge extraction problem encountered in designing and developing ES, intelligent DSS and other equally intelligent forms of integrated and emerging MI technologies and applications.

A wide range of techniques has evolved from studies in diverse fields and disciplines (e.g., medical cognition, cognitive psychology, artificial intelligence, organizational science, computer science and linguistic) in terms of eliciting knowledge from humans for the purpose of incorporating such expertise into specialized computer software. Among the more established techniques discussed in the literature are interviews, computer-based interactive techniques, methods involving rating and sorting tasks, protocol analysis, and cognitive task analysis (CTA). Owing to limited space, we will only highlight briefly each of these techniques and their implications for MI development and applications.

Apparently, interviews are direct means of acquiring knowledge from experts; they can be structured or unstructured. In structured interviews, the expert or non-expert may be probed, based on a structured protocol, to describe in sequence how specific cases are normally dealt with when they are performing a certain task, particularly under uncertain or complex situations. In unstructured interviews, the same expert (non-expert) may be asked similar type and related questions in no particular order, depending on answers previously provided. It is also possible for the interviewer to probe for further clarification or to lead the study subject to talk on particular aspects of the problems if the interviewer feels that previous answers provided are inadequate. The disadvantage of using interviews is that it is difficult to expect anyone to be able to articulate precisely the hidden knowledge that is to be extracted because most people will tell us what they think they would be expected to do in performing a certain task then to state what they actually do.

Reading the data abstracted from an EMR or EHR, an expert will anchor on certain important information that a novice may well miss out. Novices also tend to have difficulty differentiating what may be considered the most important pieces of information recorded and this is where eliciting knowledge from a novice vs. an expert in real-world case management helps to generate a high quality consultative expert system such as MYCIN.

Computer-based interactive approaches to knowledge extraction involve having the study subjects use interactive tools or computer programs known as knowledge-based editors to assist them in directly generating computer-assisted knowledge acquisition. OPAL is an example of a graphical knowledge acquisition system for a cancer therapy management program whereas INTERMED is a collaborative project which uses experimental tools for extracting knowledge of medical practice guidelines based on experts' interpretations of written guidelines. Another application of computer-based interactive approach is the use of software designed to analyze case data, thereby automatically inducing the inference rules. In this case, the interactive approach used is essentially an indirect means of knowledge acquisition, that is, a method for which inferences are made about the nature of the expert knowledge from computer analysis of case data. These case data may be abstracted from an EMR or EHR.

Psychological research in judgment and personality studies has contributed to the elicitation of knowledge via use of rating and sorting tasks. Here, the attempt is to create a classification scheme, thereby identifying the domain elements along certain meaningful taxonomies. For instance, experts can be asked to sort concepts printed on cards into meaningful clusters. Similarly, these experts may be asked to rate concepts along a certain continuum or among different categories. As an illustration, experts may be asked to rate different whiplash cases involving rear-ended motor vehicle accidents into "mild", "intermediate" and "severe" categories based on varying reports and injuries. In this way, the hidden knowledge based on the experts' general opinions as well as interpretations of these cases can be elicited. A standardization of these terms to describe the different clusters for which the cases may be discriminated can then be captured into the EMR or EHR for the respective patient.

Protocol analysis or thinking aloud may be considered a critical direct approach for knowledge elicitation. This technique has received considerable attention in cognitive psychology. The question has been whether experts are better able to articulate the knowledge they possess in thinking aloud when asked to solve a problem than less-than-experts and what are the notable differences between the thinking strategies of experts versus less-than-experts with regards to solving the same problem in specialized domains. One application of this method in the field of medical cognition is to record down the interactions of experienced physicians (or residents) with patients in terms of diagnosing the patients' problem. The analysis of these differing interactions would provide researchers with insights into the different thinking strategies of residents versus expert physicians when faced with similar diagnostic problem cases. The intent is to use the extracted knowledge for programming intelligent computer software that can serve as useful decision support tools for training residents in real-world settings. Just as interviews, a major problem with protocol analysis is the ability of the experts (or non-experts) to accurately verbalize what may be hidden in their respective thinking/reasoning processes. In many cases, protocol analysis need not be taken under real-life physician-patient interactions, these could also be generated from expert or less-than-expert abstraction of EMR or EHR data on particular patients.

Finally, CTA extends most, if not all, of the above traditional task analysis approaches to include ways of capturing higher-level cognitive processes in task performance as well as physical behaviors (see Kushniruk and Patel in Tan with Sheps, 1998). CTA refers to a set of methodologies, including the use of structured interviews, video analysis of work situations, and protocol analysis or other approaches, that can be applied separately or jointly to capture the knowledge, skills, and processing strategies of experts versus less-than-experts when given complex tasks to solve. CTA generally involves six steps: (1) identify decision problems to be studied in the analysis; (2) generate decision tasks or cases; (3) obtain a record of expert problem solving for the task(s); (4) obtain a record of the problem solving of novices and intermediates for the same task(s) that was (were) presented to the experts in step 3; (5) analyze the performance of experts versus less-than-experts; and (6) recommend the systems requirements, design specifications, and knowledge base contents for the MI development. After repeated and careful investigations as well as rigorous analyses of the data derived from the application of these several steps can we only expect to gain proper and valuable insights towards achieving the right mix of information-knowledge elements by the less-than-experts needed to support their decision making and complex problem solving. In fact, even experts will find an intelligent ES developed from CTA approach useful as the system can serve as a sounding board to situations involving critical life-death decisions. Again, these complex problem situations can be abstracted as typical EMR or EHR cases.

Our review of MI tools and applications has focused on how experts vs. less-than-experts can be aided in making complex medical decision. With patient data captured in EMR and EHR, it is possible to generate powerful tools by capturing and creating expert systems to aid these decisions in complex situations. This, in turn, leads us to the need for recognizing major themes, utilization and applications of the rapidly expanding MI field, which we will now explore.

MI MAJOR THEMES, UTILIZATION & APPLICATIONS

The increasing complexities of the MI "identity" as a result of its explosive growth in inter-disciplinary and multi-disciplinary directions call for an analysis of its major themes, utilization and applications. Over the years, a number of sub-fields, major themes, utilization and applications within the broader MI movement have emerged.

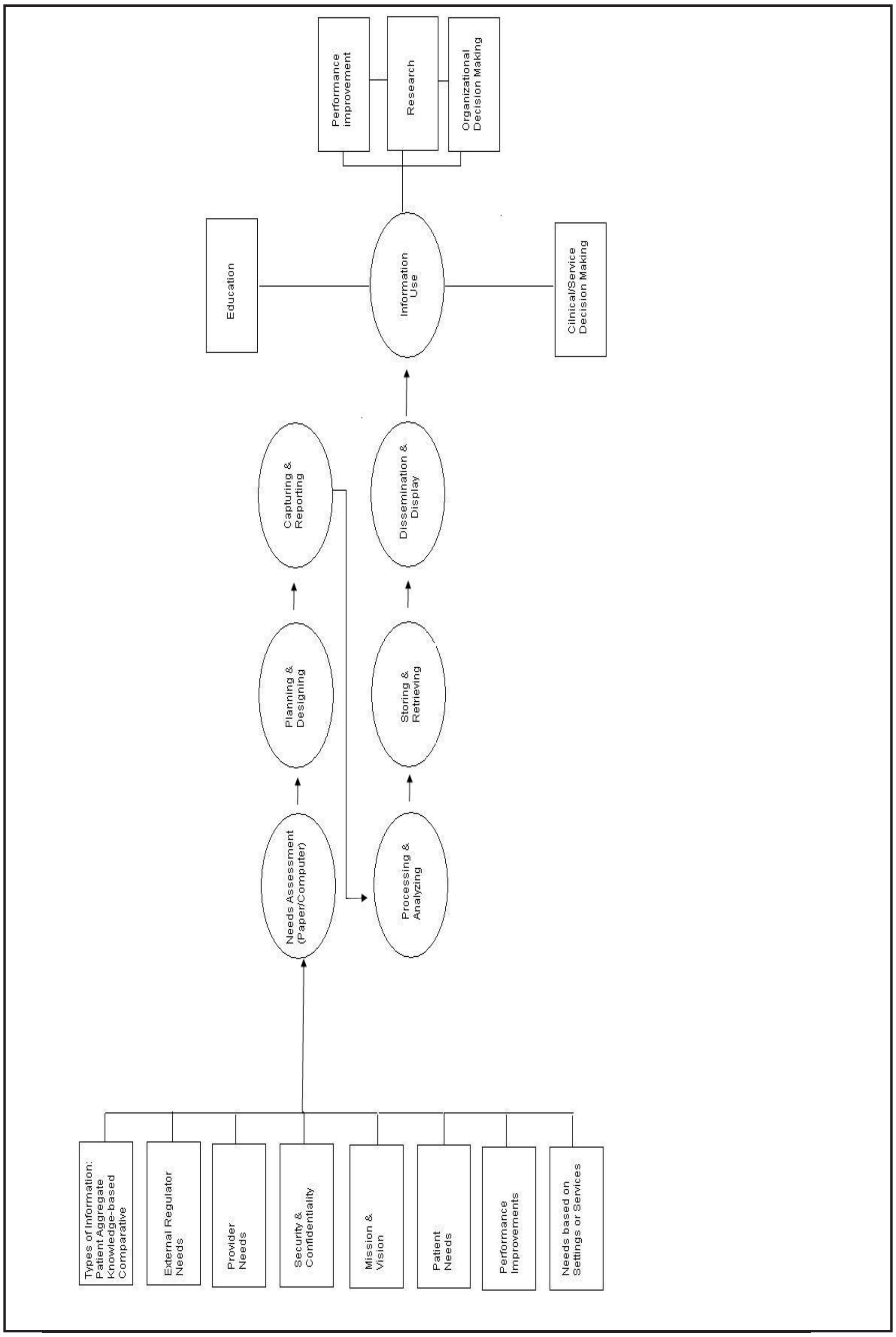
With the limitation on space for this overview, we will briefly cover three key themes encompassing various current utilization and applications that cut popularly across many of the MI sub-fields:

1. Healthcare Information Management (HIM)
2. Health Decision Support Systems (HDSS)
3. Telemedicine

Healthcare Information Management (HIM)

Encompassing a major branch of applied healthcare informatics, HIM professionals are trained to safeguard the accessibility, timeliness, integrity and security of high quality patient records for use by various health professionals. As such, these professionals are expected to be competent in many aspects of applied healthcare informatics, including best practices in health data recording, clinical coding standards and methodologies, Health Insurance Portability and Accountability Act (HIPAA) rules, health laws and regulations pertaining to the preservation of high quality healthcare datasets, healthcare data computing and analytic methods, critical issues on health information systems design and development, and the need to maintain an organizational view of HIM in an evolving and modernized healthcare environment. The American Health Information Management Association (AHIMA) and Health Information Management and Systems Society (HMISS) are the key professional organizations offering certification to HIM professionals and most textbooks written for these certification examinations rely heavily on the 2005 JCAHO (Joint Commission on Accreditation of Healthcare Organizations) model for managing information, as shown in Figure 3.

Figure 3. JCAHO Management of Information Model, 2005 adapted from http://www.jointcommission.org/NR/rdonlyres/656F7F4B-2ECA-4847-8811-B7B54DDA1584/0/LAB2008IMChapter.pdf p.3 of 18



Parallel to the HIS curriculum advocated in Tan (1995; 2001), the JCAHO model depicts a data-information-knowledge processing strategy with collected patient-specific data being aggregated to produce useful and comparable information to accumulate knowledge for supporting the HIM professional activities.

One cornerstone HIM technology that is now being seriously touted to improve patient care delivery is Electronic Health Records (EHRs). Like most HIM applications, EHRs contain both hardware and software components interacting with the human component as a catalyst to bring about customized, efficient and effective sought-after solutions to problem tasks (i.e., generate information-knowledge for patient care decisions) in the different areas of health services delivery. In the US, initial efforts to standardize health information infrastructure were made by ONCHIT, the Office of the National Co-ordinator for Health Information Technology formed under the US Department of Health and Human Services (HHS) in 2004. Headed by Dr. David Brailer, ONCHIT's mission was to achieve a US-wide adoption and diffusion of interoperable EHRs by 2014. When Dr. Brailer subsequently resigned, the HHS separately funded a nonprofit, private group, the Certification Commission for Healthcare Information Technology (CCHIT), to complete the work on setting standards for EHRs and supporting networks. Vendors meeting these standards were to be certified by CCHIT. By 2006, for example, CCHIT had certified a list of 22 ambulatory EHR products.

Beyond HIM, strong interests in MI among health researchers, educators and practitioners during the late 1980s and early 1990s quickly diversified into areas of expert method applications, clinical decision support systems, nursing decision support systems and other forms of health decision support systems.

Health Decision Support Systems (HDSS)

Researchers have always wanted to add intelligence to computer systems, and its extension to medical IS/IT quickly became noted as a most valuable and noteworthy MI application domain. Intelligent medical IS, it is thought, should be able to mimic the thinking processes of expert clinicians. How such thinking processes may be engineered and expert knowledge programmed into automated systems are issues related to ES development process, an issue we have discussed earlier in the review.

Cognitive scientists believe that the human brain may often be conceived as a "living super-computer." Experiments conducted on the human stimuli-response system inform us of the familiar stories of how human perception, pattern recognition, biases and judgments are formed when exposed to various forms of external stimuli (information) in structured vs. less-than-structured task situations. Findings from many of these studies have stimulated the design of various forms of DSS. Tan (1998) differentiated among data-based DSS, model-based DSS, knowledge-based DSS and graphics-based DSS. The application of a database management approach in which access to various organizational data sources can be achieved via a single, multidimensional data depository is basic to all forms of DSS. A model-based DSS, then, provides the decision maker with the added capability of drawing from a variety of models to fit the problem task situation via the use of a model-base management component. In contrast, the knowledge-based DSS, or more simply, an ES, makes use of embedded rule-based algorithms, frames, neural networks and fuzzy logic to augment human decisions. Finally, the graphics-based DSS, such as a geographical information system (GIS), capitalizes on the power of human visualization through the use of a flexible, graphics-driven interface component. A GIS application in healthcare services may be the digital mapping of a certain epidemic such as the HIV infection among a subpopulation group across various Sub-Sahara African countries to study the spread of the prevailing ailment. This knowledge can then be used to target regional interventions effectively for specific population groups rather than focusing treatments on isolated individuals.

A group HDSS combines analytic modeling, network communications, and decision technology to support group thinking, problem formulation, and goal seeking solutions. It aims essentially at easing the group decision-making processes, and among other things, its use will not only reduce the cognitive burden, but also the mental effort associated with groups of decision makers. A major benefit of group HDSS is its potential for increasing efficiency, effectiveness and productivity of group interactions through the application of asynchronous board meetings, online forums or special group chat facilities. Put together, in conceptualizing and designing any health group/organizational DSS, an important strategy is to achieve a good mix and match of the many different forms of DSS that will best support the combined decision needs of the group or organization.

At this point, we turn to another major theme of MI, telemedicine.

Telemedicine

Telemedicine, telematics or telehealth, according to the American Telemedicine Association (ATA), has a history of moving from the linking of two geographically separate points to achieve a medical service with the potential to administer invasive telesurgical procedures at a distance (first wave), to the use of digital networks to perform virtual consultations, including diagnoses of disease and disorders (second wave), to the empowerment of patients with medical expertise to achieve a higher level of wellness (third wave). In other words, telehealth has continued its mission to expand the boundaries and capabilities of medical services delivery to the underserved populations regardless of their geographical locations as it has the advantage of quickly transferring the medical expertise and capability concentrated in an area to another with low accessibility. The focus of telemedicine has been on medical services, not the changing technologies although changing technologies have accelerated the telehealth movement.

As the cost of such expertise transfer continue to reduce substantially with advancing low-cost telecommunications technology and increased competition, the demands for all forms of telemedical services are growing rapidly and efforts have been underway to seek Federal and third party approval for reimbursing telemedical services. Teleradiology and telephathology are among the first successful applications of telemedicine services as far as gaining third party approval for payment reimbursements. In teleradiology, x-rays or scanned images of patients are digitized and stored electronically for sharing among multiple health providers at geographically distant sites without the need for physician-patient or physician-physician interactions. Like teleradiology, no patient interaction is also needed in telephathology as the phathologists are able to provide diagnosis and consultation remotely via exchange of digitized microscopic or routine surgical images. So far, studies have documented the relative accuracy of telephathological diagnosis, which has spurred its approval for Medicare and Medicaid reimbursements, besides teleradiological services. We expect that with greater interactions among subspecialty consultants and the referring pathologist, telepathology readings would continue to improve.

Teledermatology, teleneurology and telepsychiatry are other examples of telemedicine applications where teleconsultation plays a key and active role. Beyond teleconsultations, more dynamic interactions may also be achieved in telemedicine, for example, telecare as in telehomecare and telenursing healthcare, telelearning as in videoconferencing and online medical education, and telesurgery. Telesurgery is apparently much more interactive in nature. Telegastroscopy, for instance, involves both remote consultations and surgical procedures. Indeed, the rapid development of robotics, sensor networks and sensor-based remote-activated and monitoring devices to support telesurgical and virtual reality (VR) interventions all promise a brighter future for telemedicine.

Next, we shift focus to social and organizational implications on MI diffusion.

MI ORGANIZATIONAL & SOCIAL IMPLICATIONS

A major aspect to successful MI diffusion and implementation has to do with overcoming the barriers that are related to social and organizational factors. In terms of organizational and social implications for MI diffusion, not only is it critical to underscore its impacts on individual users, groups and organizations, but also it is important to relay how the MI technology may ultimately affect the larger context of our health services delivery system and on society.

As we enter the next era of the MI field, knowledge of MI impacts will pave new directions for MI technology development and innovation. Still, impediments on the future growth and development in MI will surface such as the growing complexity of legal and security issues in MI.

Impacts on Individuals, Groups and Organizations

MI impacts significantly on the individual user, especially the clinician with respect to work habits, information processing efficiency and decision-making capability, as well as reliance on automated intelligence; on a group level, it impacts on the quality of teleconsultations among clinical experts, volume of information exchanges and effectiveness of group decisions; and on the organization level, it impacts on the quality of organizational patient care delivery and global competitiveness.

More specifically, at the individual level, it is critical to know if use of MI applications will result in speed, accuracy, greater clinical significance and decision-making effectiveness for the clinical user. For example, it may be argued that physicians who are equipped with a Blackberry or an iPhone that will serve both as a cellular phone with an automated reference drug directory, and a Net appliance with the capability to access a built-in e-prescription system will be able to better perform an intelligent online prescription request order for their patients anytime, anywhere than physicians who do not adopt such a technological innovation. At the group level, MI will impact on the ability of expert consultants to share and exchange data, to coordinate teamwork and progress group diagnostic and therapeutic activities. A virtual patient record (VPR) system for use by multiple care providers, for example, is one that will integrate all of the information provided by the different clinicians regarding the patient at anyone time.

At the organizational level, MI diffusion will impact on changes such as organizational reporting structure, work habits and culture. Not only will MI be expected to improve clinical productivity and enhance intra-organizational communications, but also it will redefine clinician-patient relationship, improve organizational image, alter organizational culture, increase market competitiveness and open up new avenues for new collaborations, organizational arrangements and partnership services. Health organizations will adopt a more sensitive attitude and mindset towards information sharing and an evolving culture with the experience of having gone through a major MI diffusion; for example, an organization may completely change the way it performs patient care because of automated intelligence, online training, telework habits and virtual network capabilities. When MI implementation fails, the resulting consequences can also be devastating for an organization not only in terms of costs but also personnel turnover, changes in technical leadership and reporting structure and the potential for future MI resistance from users.

Impacts on Health Services Delivery System and Society

The same impacts discussed previously may now be extended to the entire health services delivery system as well as society. For example, the use of the Internet to transfer massive amount of media-rich patient data and the availability of knowledge systems such as robots and automated intelligent systems may induce many legal and ethical questions about privacy, security, as well as individual and institutional intellectual property rights issues. One contention, for example, is: Who owns all the different pieces of the stored medical information about a particular patient? Another question may be: What information should (should not) be kept online about an individual and who has the right to alter the information stored on a MI database? In the past, many of these and other similar questions have prevented desired progress in the MI field. With the enactment of HIPAA rules, answers to many of these questions are becoming clearer but not definite until test cases move forward and authoritative court decisions have been pronounced.

Critical societal impacts of MI advances include changes in how work may be performed (e.g., telecommuting, virtual patient visits, self-diagnosis), changes in the availability, accessibility and affordability of medical expertise to the underserved patient population, new opportunities for cyber-crime and new ways of purchasing telehealth services for consumers, new gadgets and automated devices for helping seniors and the disabled, the construction of healthy "smart" houses and new methods in self-care, wellness promotion and lifestyle changes, and in redefining the quality of one's life. Imagine how MI applications can not only improve the efficient and effective functioning of a group of doctors and nurses, but patients who have access to specialized MI equipment and devices may also be aided to perform self-care and reduce the burden on hospital emergency services.

One sensitive question, for example, has to do with the determination of technological resource allocation to decide who among the "surviving" patients should be saved in the aftermath of an epidemic or major health hazards? Take the September 11 scenario and the fact that there just are not enough first responders and facilities around to save all of the victims who were immediately affected. How does society go about deciding where to allocate the limited MI and other resources to achieve the most justifiable results? These and other aspects of social impacts such as digital divide, cost effectiveness and the productivity paradox relating to ethical, security, legal and political concerns still need to be worked out and answers to these questions are never easy. They will be further explored in the section on critical issues.

MI MANAGERIAL IMPACT & KEY IMPLEMENTATION ISSUES

Any innovation, including MI, will become obsolete if not diffused. In practice, MI diffusion issues relate to MI strategy and vision as well as many other critical issues, which have been discussed partly in the theoretical section of this review such as MI technology management, leadership and acceptance-resistance experiences in MI implementation and will be further explored in the next section.

Here, our focus is on activities dealing with the need for management to create an MI strategy and vision for diffusing competitive, cost-effective MI applications in key areas throughout a healthcare organization. These issues summarize the MI managerial impact.

MI Strategy

While there may be good reasons to support MI innovation such as enhanced patient care, more intelligent clinician decision outcomes, and greater efficiencies in the clinical evaluation process, new efforts may fail for lack of a successful MI strategy.

In essence, MI strategy entails aligning the business or organizational mission (the goals) with the information needs of the users (particularly those of the clinicians and nurses for health care organizations) in the delivery of patient care (the tasks). In other words, three key questions should be asked when structuring an MI innovation diffusion strategy: First, “Is the MI strategy targeted to fulfill the core mission of the organization?” For example, in the case of a center for cancer treatment and research, plan should be in place to position this particular center to become one of the best places in the world to seek cancer treatment owing to its MI innovation diffusion strategy. Second, “What is the clinical significance of the MI innovation?” Here, the center clinicians should actually be benefiting significantly in their practices due to employing MI innovation – the effects should be evidenced in more medical breakthroughs and error prevention due to improved treatment decision outcomes. Finally, “Is the MI innovation serving its ultimate purpose of augmenting patient care delivery?” Herein lies the significance of the MI diffusion equation – are these systems functioning appropriately to save lives, improve the treatment center’s image, and strengthen the provider-patient relationships so cherished by the organization.

A major trend in MI strategic planning is the shifting of responsibilities and power from traditional IS/IT professionals to end users, and more appropriately, to top management. This trend is justified because of the growing acceptance of the notion that IT is a corporate and strategic resource and should be properly managed just like any other organizational assets including land, labor and capital. More recent approaches call for a shortening of the time span among MI strategic planning sessions and more attention paid to the changing marketplace. Regardless of the time span in-between strategic planning sessions, there is always the need to conduct environmental assessment before moving onto the formulation of a MI strategy.

One approach to MI strategic planning is scenario planning. Here, competing multiple futures may be first envisioned and strategies are then developed and tested against these possible futures. The MI vision is then set within these possible futures and further reconciled with current reality. For example, if the MI vision is one in which intelligent, sensor-based medical devices are to be used to monitor cancer patient round-the-clock while the current reality involves only regular treatment visits and emergency monitoring, then the transition to the new MI-intensive environment will not only call for changes in work practices and habits, but new ways of connecting patients to the cancer center electronically. Indeed, no single approach to MI strategy formulation is considered superior; instead, a blending of approaches is often recommended as the various approaches deal with different aspects of creating a feasible MI strategy.

MI Implementation

The challenges of MI implementation are often interwoven with many other social and organizational challenges, including the integration of quality planning, quality control and quality improvement processes to evolve a secure, well-managed, and quality-focused MI environment, the integration of information management technology, organization management technology, and user-interface technology for building an efficient, organization-wide infrastructure to support MI innovation and expansion, and the integration of data, model, and knowledge elements for designing effective MI applications.

Any MI implementation will bring about some form of organizational change, for example, changes in the organizational reporting structure, changes in the level of computing competence required of current and future clinical users, and changes in the information flow processes and decision-making functions of the organization. To ensure that clinicians will have the appropriate knowledge, skills and attitudes that are critical for addressing concerns arising from these MI related changes, it is important that health organizations also address staffing and training issues such as having an aggressive recruitment program for attracting valuable MI professionals from among competitors, creating opportunities for training and development, and employing winning strategies for the retention of knowledgeable and well-trained MI experts.

At the core of MI implementation is overcoming user resistance and cultivating user acceptance through a continuous process of buy-in, quality improvement, participatory decision making such as efforts to involve the users in identifying their information needs, involving them in systems specification, system analysis, systems design, vendor selection, system testing and providing them with flexible and extensive user training opportunities. Remember that physicians and most other clinicians have complex schedule and will not work around their schedule just to attend a training meeting. Moreover, opportunities for input from all end-users of the MI systems will only ensure that the designed system meet the expectation and needs of the users. Most importantly, the users will only be willing to most likely use the system on an increasing basis if they find that there are ways to tell them that the systems benefited them in lightening their workload and in improving their productivity, giving them more time to do patient care. As well, availability, accessibility, efficiency, effectiveness and ease of use of the systems all play a part towards getting the users hooked on the systems.

Together, these key challenges typically point towards the need for an integration of environmental, technological, and organizational components for driving and directing the implementation of various MI technologies and applications successfully within the larger intra-organizational or inter-organizational systems context. Other MI administrative issues may include policies dealing with privacy, security and confidentiality of patient records, legal and ethical considerations in clinical data collection, analysis, and distribution, and organizational policies regarding authorized use of MI.

CRITICAL ISSUES

As noted, the primary objectives of most, if not all, MI implementation are reduced operational costs, greater patient satisfaction, and better quality clinical decisions. As MI systems promise to make information and decision handling more efficient and effective in reducing costly errors and unnecessary delays, as well as projecting an improved professional image, it is important to implement MI applications that are accepted by its users (clinicians).

Critical Success Factors

What factors determine the success of an MI innovation diffusion effort? Critical Success Factors (CSFs) for MI diffusion are those factors that will drive the success of MI diffusion. Based on past studies, commonly held managerial factors for success include effective technology management (TM) and executive leadership, as well as infrastructural factors whereas organizational and social factors include implementation factors and culture-specific factors. The focus of this section will be on the managerial factors as the organizational and social factors are elaborated in Section VII on organizational and social impacts.

Briefly, technology leadership and the management of the MI capacity within the health organization are key success factors for MI innovation and implementation. Top management must work to ensure that appropriate infrastructural support is in place to develop the MI expertise needed to support current and future managerial and organizational functions and activities. In many cases, executive leadership is expected, including directives and policy setting by the chief information officer (CIO) in order to guide necessary organizational restructuring for efficient information processing and effective decision making, the revamping of decision processes, the bridging of MI innovation with changing environment, and the rapid adaptation of new MI capabilities to support increasingly sophisticated clinical information services in this age of information and knowledge explosion.

As we have noted, it is important for an effective MI strategy to impact on the business of the organization. The resistance to MI implementation experienced within an organization is largely due to poor alignment between MI strategy and organizational strategic plan, inadequate communications, lack of user training, and support from top management. A champion for MI to forming a task force that would marshal the needed support and user-analyst relationship and/or user participation throughout the development of MI strategy is an important ingredient to achieve MI diffusion, utilization and acceptance. Without a champion at the senior management level, it is likely that MI resources will not be well allocated and its needs for efficient and effective transformation of the organizational work habits will not be supported. This is especially true for health care organizations, with its established tradition of intensive paper-based information processing, professionally oriented physicians and nurses, and traditional management habits.

Other Issues

Other issues concern computer literacy among clinicians and all health organizational workers catering to patient care functions, the digital divide, the productivity paradox, ethical, legal and security issues. Again, an in-depth treatment of any of these issues is more properly the devotion of another chapter. Here, we will only brief touch on some of them.

Computer literacy, which can often translate into resistance, can be a threat even for highly trained professionals, as they do not want to be embarrassed for showing ignorance about their lag in technological expertise in front of their colleagues. Thus, it may be wise to conduct separate training sessions for executives, physicians, nurses, middle managers and staff.

A broader impact of computer literacy is the digital divide, where it is apparent that we need to pay special attention to the underserved population who may find accessibility to information technology and use of Internet a real challenge. Especially vulnerable are some special groups such as seniors who are also classified as minority and those living in impoverished parts of many of the inner cities. In this instance, the emerging field of e-health is making an impact. Tan and colleagues, for example, are working on educating a group of these seniors in selected urban Detroit areas on healthy aging and lifestyle changes through the use of Internet-based educational software known as the eHealthSmart program. A major aim of this program is to provide Internet-based tutorials on smoking cessation and weight management to the elderly and seniors who are eager to use computers to help themselves in changing their unhealthy lifestyle behaviors.

Another critical issue is the productivity paradox, which refers to the phenomenon that, to date, we have yet to witness the same level of productivity improvements achieved with the automation of processes in the service industry since the advent of computers in the way we have witnessed the effects of automation on productivity experienced in the manufacturing industry. In other words, information technologies, with its increasing applications across all service sector industries, have not been widely productive as anticipated. Apparently, it is argued that, on the one hand, we are not out of the woods yet in terms of maturing from old aged manufacturing productivity and efficiency while, on the other hand, we have yet to reach the stage of being fully immersed in the digital economy to leverage on technological productivity and efficiencies on a massive scale. Time can only tell when this generation will reach the threshold of a critical mass of technology usage to achieve the productivity and efficiency that we have already experienced in manufacturing, albeit the gradual slowdown. What this implies literary is the need for greater and higher MI diffusion in the current century, especially with health care taking a lead role. After all, which industry would thrive with automation in this economy when we are faced with an aging population, an accelerating health care expenditure of approximately 16% to17% of our

GNP? Hopefully, many of us will live to see the promise land of technological productivity that MI diffusion and automation will usher us in.

In terms of security and privacy issues, the health care organization is held responsible as the gatekeeper for securing private and critical information on hundreds and thousands of individual patients. Imagine someone finding a jump drive or disks containing private hospitalization records of certain patients who may also be important political figures, legendary athletes, famous entertainers or key legislative and policy makers. Imagine also that intruders are able to hack into the various MI systems and access key information on patient and hospital personnel identification, laboratory test results and digitized radiological images of different patients, or information on storage facilities for key pharmaceutical products and controlled substances. The ramification for losing any such information can be devastating for both the individuals whose information has been stolen as well as for the institution serving as the gatekeeper of our most private and confidential information.

EMERGING TRENDS

Moving on to emerging trends of MI thinking and sub-specialties that have surfaced in recent years, we observe that there are basically two streams of innovation. One group of paradigm shift includes consumer health informatics, electronic healthcare (e-health) and all its sub-fields such as electronic medicine (e-medicine), e-homecare, and patient-centered e-health (PCEH). Another group of paradigm shift concerns mobile healthcare (m-health) technologies, ubiquitous healthcare (u-health), and virtual healthcare (v-health). Here, it is possible only to provide readers with brief glimpses of these sometimes overlapping but emerging MI trends. Readers who are interested in pursuing further details can refer to the readings provided in the bibliography and the rest of this 4-volume work.

Consumer health informatics is bringing about a consumer-driven health care model in that consumers will play a more active and decisive role regarding managing their own health and well being. For the educated consumer, greater access to quality health information is, in and of itself, a form of therapy. E-health is a paradigm shift to re-channel clinical expertise into the hands of the less-than-experts through the applications of e-technology. E-technologies encompass complex MI network design and consumer-oriented informatics, the Internet, wireless communications and emerging technological applications such as web services and remote clinical monitoring devices and management systems. E-medicine, another MI sub-fields, refers to affordable and acceptable use of MI to support health-related services, surveillance, information, education, knowledge, and research dissemination. It is the use of advanced information and communications technology (ICT) to bridge geographic gaps and improve care delivery and education. Patient-centered e-health (PCEH), in a nutshell, represents one of the paradigm shifts that is part of the larger e-health movement where not only is the role of healthcare providers being redefined, but where the expectation bar for consumers to participate actively in decisions leading to their own health as well as the overall quality and acceptability of e-healthcare informatics and services are being raised. Moreover, we see that the various disciplines of e-clinical care, e-public health care, e-nursing and e-homecare overlap to a considerable extent. Apparently, it has not become easier, but more fuzzy, to distinctively separate the different knowledge domains of “e-health,” which is the umbrella concept to encompass most, if not all, of these emerging MI sub-disciplines and sub-fields.

M-health, an extension of the e-health concept, refers to mobile-health or use of remote, satellite-based networks, cellular-based networks, and other wireless networks (e.g., sensor-based networks) to effect health data exchange services as well as clinical information and decision support services. Ubiquitous healthcare (u-health), and virtual healthcare (v-health) are the next evolution of MI where concepts such as virtual community networks, ES/HDSS embedded in walls, table-tops, refrigerators, and other objects in the surrounding environment, as well as VR such as use of head mounted displays, smart houses and wearable computing are becoming commonplace. Examples include a doctor checking the status of a patient's wellness in real time by connecting to a nano-sensor device installed in a patient's home or attached physically in some shape, form or manner to the patient or a monitoring system, which automatically contacts the designated health agency to initiate treatment procedure while alerting on an impending heart attack for a patient. Such an evolution can be expected to provide an ubiquitous MI resource for delivering patient care services.

The trending of the MI movement towards achieving u-health and v-health may soon become a reality when personalized information exchanges, integrative knowledge sharing, and decisional support and communications can happen in real time between an institutional care provider and the patient. Electronic health records will be available, accessible, affordable and adaptable to our needs. Nano-science and gene-based therapies will be possible because of intelligent mining of these personal records not only to serve us in terms of our wellness recovery whenever we are injured, but also brings about a sense of equity among the community, serving even those who are unable to insure themselves. Notwithstanding these changes, we have seen how a young discipline such as the MI field can grow and expand quickly to affect every aspects of our daily living, in particular, our health care system, our community and our health.

CONCLUSION

In bringing this discussion to a close, we will use the example of designing a expert diagnostic support system (E-DSS) useful for both physicians and patients to show how the different major facets of MI themes, its theories, central hypothesis, tools/methods, diffusion and impacts can all be combined to achieve effective health care. First, this E-DSS may have to be divided into two components, one to support experts (i.e., physicians, nurses and clinical therapists) and the other for layperson (i.e., consumer, patient) support. This MI tool will, say, be focused on wellness promotion areas such as smoking cessation, weight management, and nutrition advice. On the one hand, the expert component of this E-DSS will mimic the hypothetico-deductive approach used by most clinicians in deriving diagnostic decisions, which is, essentially, an iterative process of data collection, interpretation and hypothesis generation and refinement until a satisfied level of certainty is reached. On the other hand, the layperson component of the E-DSS will mimic an online tutorial for layperson to be educated with the expertise programmed into the different health promoting educational modules. Thus, this E-DSS concept illustrates the central hypothesis of MI philosophy, which is simply to enhance decision making, based on the GST that underlies also many of the evolving MI themes that we have reviewed so far, where synergy effects can be achieved with the combination of individual components in the design of MI applications.

Second, for most of the clinical decisions associated with the E-DSS such as knowing if a female patient is at risk for diabetes due to unhealthy lifestyles, the sensitivity-specificity methodology forms the basis for hypothesis generation. For the E-DSS to "think" like an expert clinician, the system must therefore be taught the different likelihood of specific symptoms to a health challenge, such as the likelihood of having diabetes as a result of being obese, unchanged smoking habits or poor lifestyle behaviors. Such likelihood, better expressed in terms of probability, is a measurement derived from health statistics. The cutoff point in which a patient is diagnosed with a particular challenge is the result of continuous data collection and analysis, for example, a female patient 50 years old is diagnosed with diabetes when her blood glucose reading reaches 5.5. If this cutoff point is reduced, then the sensitivity of the blood glucose test increases while the specificity decreases. Thus, if the cutoff point for diagnosing diabetes for the above clinical group is reduced to 5.4, then more patients will be diagnosed with diabetes. The cutoff point is then a guideline to diagnosing diabetes. However, a patient with a blood glucose reading of 5.6 does not necessarily have diabetes. This is because every clinical test has false positives as every clinical test result is made up true positives, true negatives, false positives, and false negatives. As we noted previously, the true-positive rate (TPR) is the sensitivity probability of a test result and measures the likelihood that the patient being suspected of having the disease does have a positive test result. Conversely, the true-negative rate (TNR) is the specificity probability of a test result and measures the likelihood of a non-diseased patient having a negative test result.

The initial hypothesis of a patient having a certain disease or condition is the pretest probability (i.e., prevalence). This judgment may be based on prior experience or on knowledge of the medical literature. When tests and examinations are conducted, the initial hypothesis is verified, thereby yielding the predictive value, or resulting in the posttest probability. Predictive value is the probability that a disease is present based on a test result. The posttest probability then becomes the pretest probability of the next hypothesis. This leads to the Bayes' theorem to calculate the posttest probability using the pretest probability and the sensitivity and specificity of the test. Herein lies the foundation of MI theory and methodology on which our clinical reasoning model is built upon and programmed into the expert component of the E-DSS to support clinical and diagnostic decisions. It is possible, of course, to apply other methodologies such as neural networks, heuristic programming and case-based reasoning, but our intent here is merely to illustrate a common and specific MI methodology based on a well-known decision theory, the Bayes' theorem. On the layperson component, the E-DSS may be supported with a knowledge base where expert knowledge on weight management, smoking cessation or healthy nutrition can be made available for online consultation.

In terms of MI diffusion and impacts, the E-DSS as we have illustrated must be supported by healthcare organizations that would promote wellness such as third-party payers and accepted by the intending clinical and layperson users in order to be effective and have a positive impact on the individual, group, organization and society. Specifically, the clinician will have a better control of predicting the risk for particular patient to have, say, obese-related diseases, the patient will have a better management of their own lifestyles with access to expert knowledge, the health maintenance organizations (HMOs) or the government will be able to reduce payments on treatment costs due to healthier subscribers, our society will have less burden to bear as people age more gracefully and the list goes on. Further, most HMOs and governmental agencies have numerous other patient record systems with varying computer platforms and data storage structures. How this E-DSS can be integrated or embedded into the larger organizational computing infrastructure is a key challenge. In other words, rather than having patient data stored redundantly in the E-DSS as well as other separate systems such as computerized patient records (CPR), it may be possible to network these different systems so that the same data are stored only once and can be shared virtually among the different sub-systems such as in the HMO's EHRs that was discussed previously. Other related challenges include the need to align the goals of the organization (the business) with the goals of the different sub-systems, for example, the clinician's aim to treat his or her patients speedily, and the aims of the IS/IT support departmental staff to coordinate data collection and standardization, to improve computer-user interfaces, to train users and to encourage their active participation in managing and harnessing the technology. Until a seamless alignment is achieved, the diffusion of MI such as the E-DSS technology will be limited due to lack of acceptance among its supporters.

Finally, the greatest impact of a system like the E-DSS we discussed here will be felt when it can be made available to any clinician as well as any layperson worldwide and not just those living in a certain area or country. In this sense, the development of sensor networks and ubiquitous computing is an ideal platform for MI development. When software applications are run through these wireless networks, the limitations on geographical location and time disappeared. In other words, the necessary health expertise can be made accessible, available and affordable 24/7 anywhere in the world to less-than-expert physicians and underserved patients. When the access interface becomes ubiquitous, the dynamic interactions between the clinician and the patient may not only be limited to diagnostic decision making and treatment planning, but it will be 24/7 monitoring even for disease prevention, lifestyle changes and wellness promotion. As people live longer, such new demands on the next generation MI innovation will certainly be on the rise. Nonetheless, security is always the major issue when MI applications are to be accessed ubiquitously. The integrity, privacy and confidentiality of clinical data must not be compromised when MI development moves in this direction.

REFERENCES

Bertalanffy, L. V. (1968). *General Systems Theory*, New York: Braziller.

Blois M, Shortliffe E. (1990). The computer meets medicine: emergence of a discipline. In: E. Shortliffe, L. Perreault (eds). *Medical Informatics: Computer Applications in Health Care*. v. 20. Reading, MA: Addison-Wesley.

Collen M. (1977). Preliminary announcement for the Third World Conference on Medical Informatics. *Medinfo* 80.

DeLano, W. L. (2005). The case for open-source software in drug discovery. *Drug Discovery Today, 10*(3).

Erickson, B. J., Langer, S., & Nagy, P. (2005). The Role of Open-Source Software in Innovation and Standardization in Radiology. *Journal of the American College of Radiology, 2*(11), 927-931.

Fitzgerald, & Kenny. (2004). Developing an information systems infrastructure with open source software. *Software, IEEE, 21*(1), 50-55.

Hersh, W. (2002). Medical informatics: Improving health care through information. *Journal of the American Medical Association*, 288:1955-1958.

Joint Commission on Accreditation of Healthcare Organizations. (2005). *Comprehensive Accreditation Manual for Hospitals*. Oakbrook Terrace, IL:JCAHO.

Jordan, Theresa J. (2002). *Understanding medical information: A user's guide to informatics & decision making*, New York: McGraw-Hill.

Kantor, G. S., Wilson, W. D., & Midgley, A. (2003). Open-source Software and the Primary Care EMR. *Journal of the American Medical Informatics Association, 10*(6), 616-

Kerlinger, F.N. (1986). *Foundations of Behavioral Research*, 3rd Ed., New York: Holt, Rinehart & Winston.

Newell, A., & Simon, H. A. (1971). *Human Problem Solving*, Englewood Cliffs, NJ: Prentice-Hall.

ACM. (1987, November 5-6). *Proceedings of ACM Conference on History of Medical Informatics*, Bethesda, MD, USA.

Rogers, E. M. (1983). *Diffusion of Innovations, 3rd Edition.* New York: Free Press.

Scarsbrook, A. F. (2007). Open-source software for radiologists: A primer. *Clinical Radiology, 62*(2), 120-130.

Shortliffe, E.H., Perreault, L.E. et al. (Eds.) (2001). *Medical Informatics: Computer Applications in Health Care.* New York: Springer-Verlag.

Stead, W.W. (1998). Medical informatics: On the path toward universal truths. *Journal of the American Medical Informatics Association, 5*(6), 583-584.

Tan, J. (Ed.) (2005). *E-Health information systems: Introduction to students and professionals*. San Francisco: Jossey-Bass Inc.

Tan, J. (2003). Information Systems (IS) in Health Care. In H. Bidgoli et al. (eds.) *Encyclopedia of Information Systems* (Vol 2, pp. 519-536). Academic Press.

Tan, J. (2001). *Health Management Information Systems: Methods and Practical Applications* Second Edition. Gaithersburg, MD: Aspen Publishing Inc.

Tan, J., & S. Sheps (eds.) (1998). *Health Decision Support Systems*. Gaithersburg, Maryland: Aspen Publishing Inc.

van Bemmel J.H. (1984). The structure of medical informatics. *Medical Informatics*, *3*(4), 175-180.

van Bemmel, J.H. and Musen, M.A. (Eds.) *Handbook of Medical Informatics*. Heidelberg, Germany: Springer, 1997.

Zachary, W. A Cognitively Based Functional Taxonomy of Decision Support Techniques. Human-Computer Interaction 2 (1986):25-63.

About the Editor

Joseph Tan, Dip, BA, MS, PhD, holds a professional diploma in civil engineering from Singapore Polytechnic, an undergraduate degree in mathematics and computer science from Wartburg College, IA, a master's degree in industrial & management engineering from the University of Iowa, and a PhD in management information systems from the University of British Columbia (UBC). He has been a tenured associate professor, teaching in the Department of Healthcare & Epidemiology at UBC for many years prior to serving as professor and head of the Information System and Manufacturing (ISM) Department, School of Business, Wayne State University.

Professor Tan publishes widely in numerous computing, ergonomics, information systems, health informatics, health education, e-health and e-business journals and has served as guest editor and member of various journal editorial boards. He sits on key organizing committees for local, national, and international meetings and conferences. Professor Tan's research, which has enjoyed significant support in the last several years from local, national and international funding agencies and other sources, has also been widely cited and applied across a number of major disciplines, including healthcare informatics and clinical decision support, health technology management research, human processing of graphical representations, ergonomics, health administration education, telehealth, mobile health, and e-health promotion programming.

Section I
Fundamental Concepts and Theories

This section serves as the foundation for this exhaustive reference tool by addressing crucial theories essential to the understanding of medical informatics. Chapters found within these pages provide an excellent framework in which to position medical informatics within the field of information science and technology. Individual contributions provide overviews of knowledge management in healthcare, health portals, and health information systems, while also exploring critical stumbling blocks of this field. Within this introductory section, the reader can learn and choose from a compendium of expert research on the elemental theories underscoring the research and application of medical informatics.

Chapter 1.1
Evaluation of Health Information Systems:
Challenges and Approaches

Elske Ammenwerth
University for Health Sciences, Medical Informatics & Technology, Austria

Stefan Gräber
University Hospital of Saarland, Germany

Thomas Bürkle
University of Erlangen, Germany

Carola Iller
University of Heidelberg, Germany

ABSTRACT

This chapter summarizes the problems and challenges that occur when health information systems are evaluated. The main problem areas presented are the complexity of the evaluation object, the complexity of an evaluation project, and the motivation for evaluation. Based on the analysis of those problem areas, the chapter then presents recommendations of how to address them. In particular, it discusses in more detail what benefits can be obtained from applying triangulation in evaluation studies. Based on the example of the evaluation of a nursing documentation system, it shows how both the validation of results and the completeness of results can be supported by triangulation. The authors hope to contribute to a better understanding of the peculiarities of evaluation in healthcare, and to provide information how to overcome them.

INTRODUCTION

It is hard to imagine healthcare without modern information and communication technology (ICT). It is evident that the use of modern information technology (IT) offers tremendous opportunities to reduce clinical errors, to support healthcare professionals, and to increase the efficiency of care, and even to improve the quality of patient care (Institute of Medicine, 2001).

However, there are also hazards associated with ICT in healthcare: Modern information systems (ISs) are costly, their failures may cause negative effects on patients and staff, and possibly, when inappropriately designed, they may result in healthcare professional's spending more time with the computer than with the patient. This all could have a negative impact on the efficiency of patient care. Therefore, a rigorous evaluation of IT in healthcare is recommended (Rigby, 2001) and is of great importance for decision makers and users (Kaplan & Shaw, 2002). Evaluation can be defined as the decisive assessment of defined objects, based on a set of criteria, to solve a given problem (Ammenwerth et al., 2004).

The term ICT refers to technologies as such. Whether the use of these technologies is successful depends not only on the quality of the technological artifacts but also on the actors (i.e., the people involved in information processing and the organizational environment in which they are employed). ICT embedded in the environment, including the actors, is often referred to as an IS in a sociotechnical sense (Berg, Aarts, & van der Lei, 2003; Winter et al., 2001).

Many different questions can lead the evaluation of IT. Within evaluation research, two main (and often rather distinct) traditions can be found: The objectivist (positivistic) and the subjectivistic tradition (Friedman & Wyatt, 1997), which are related to the dominant use of either quantitative or qualitative methods (for details, see Chapter XII).

Despite a large amount of published evaluation studies (e.g., van der Loo, 1995) found over 1,500 citations on evaluation of healthcare IT between 1967 and 1995, and Ammenwerth and de Keizer (2004) found 1,035 studies between 1982 and 2002; many authors report problems during evaluation. One of the main problems frequently discussed is the adequate choice of evaluation methods. While objectivistic researchers tend to concentrate on quantitative methods, subjectivistic researchers mainly rely on qualitative methods. Sometimes, a mixture of methods is applied. For example, qualitative methods are used to prepare quantitative studies, or quantitative measurements are used to support qualitative argumentation. However, there is still usually one tradition which dominates typical evaluation studies, leading to a focus either on quantitative or qualitative methods.

Many researchers point to the fact that this domination of one method or tradition may not be useful, but that a real integration of various methods from both traditions can be much more helpful to get comprehensive answers to given research questions. The integration of the complementary methods (and even beyond this, of data sources, theories and investigators), is discussed under the term triangulation.

In this chapter, we first want to review some of the underlying reasons that make evaluation of healthcare IT so difficult. We will structure the problems into three main problem areas: the complexity of the object of evaluation, the complexity of the evaluation project, and the motivation to perform evaluation. We will discuss means how to overcome the discussed problems.

As one more detailed example, we then discuss what benefits can be obtained from applying triangulation in an evaluation study. Based on the example of the evaluation of a nursing documentation system, we show how both the validation of results and the completeness of results can be supported by triangulation.

TYPICAL PROBLEMS IN EVALUATION OF IT IN HEALTHCARE

First Problem Area: Complexity of the Evaluation Object

When understanding IT as part of the IS of an organization, it is clear that evaluation requires not only an understanding of computer technology, but also of the social and behavioral processes that affect and are affected by the technology. This complexity of the evaluation objects has some important consequences. First, the introduction of IT takes time. It is not enough to implement the technology and then to immediately measure the effects. Users and workflow need a lot of time to get used to new tools and to completely exploit the new possibilities (Palvia, Sharma, & Conrath, 2001). Thus, evaluation results can develop and change during this first period of use. Then, even after an introduction period, the evaluation object may steadily change (Moehr, 2002; moving evaluation target). For example, the use of IT may be affected by changes in work organization, or in staff. It is nearly impossible to reach a stable situation in a flexible healthcare environment which makes evaluation results dependant of the point in time where the evaluation took place. In addition, each IS in our definition is quite unique. While the IT may be similar in various departments, workflow, users and used functionality may be different. In addition, the organization of its introduction as well as the overall user motivation may differ. Thus, even when the same IT is introduced, its effects may be varying (Kaplan & Shaw, 2002). The influence of such factors on the results of an evaluation study is often hard to disentangle (Wyatt, 1994), posing the problem of external validity (Moehr, 2002): Many evaluation studies may be valid only for the particular institutions with their specific IS.

The complexity of the evaluation object is an inherent attribute in healthcare IT evaluation and cannot be reduced. However, there are some ways to handle this problem in evaluation studies. To address the problem of external validity, the IT and its environment that is going to be evaluated should be defined in detail before the beginning of the study. Not only the software and hardware used should be described, but also the number of users and their experience and motivation, the way IT is introduced and used, the general technical infrastructure (e.g., networks) and any further aspects that may influence the usage of IT and its effects. The functionality and the way it is really used should also be of importance. Only this information may allow interpretation of the study results and comparison of different locations. Then, to address the problem of the moving evaluation target, all changes in the IT and its interaction with the users should be carefully documented during the study. For example, changes in workflow, in staffing, or in hardware or software should be documented with reference to the ongoing evaluation. This permits the explanation of changes and differences in effects measured during the study period. Another approach to address the problem of the moving evaluation target may be to define smaller evaluation modules. This would allow the evaluation design or evaluation questions to be adapted to changes in the environment. Each module answered a question related to a defined phase of the introduction of the IT. In addition, an evaluation must be planned in a long-term perspective in order to allow the users and the environment to integrate the new IT. Hence enough resources for long-term evaluation (e.g., over several months or even years) should be available.

Second Problem Area: Complexity of the Evaluation Project

Evaluation of IT is performed in the real and complex healthcare environment, with its different professional groups, and its high dependency on external influences such as legislation, economic constraints, or patient clientele. This poses problems to the evaluation projects, meaning the planning, executing and analyzing of an IT evaluation study. For example, the different stakeholders often have different conceptions and views of successful IT (Palvia et al., 2001). The different stakeholder requirements can serve as a frame of reference for evaluation during the early phases of the IT life cycle, but also guide evaluations during later phases. In each case, multiple-stakeholder views may lead to a multitude of (possibly conflicting) evaluation questions (Heathfield et al., 1999).

Depending on the point of view adopted, the evaluation will require different study designs and evaluation methods. The evaluation researcher must decide, for example, on the evaluation approach, on the adequate evaluation methods (e.g., quantitative vs. qualitative), and on the study design (e.g., RCT vs. observational study). Each has its own advantages and drawbacks (Frechtling, 1997; Moehr, 2002), making their selection a rather challenging endeavor. This multitude of possible evaluation questions and available evaluation methods makes the planning of an evaluation study quite complex.

The complexity of the evaluation project has several consequences. First, the overall success of IT is elusive to define (Palvia et al., 2001), and it is therefore often difficult to establish clear-cut evaluation criteria to be addressed in a study (Wyatt, 1994). Each stakeholder group may have individual questions, and a universal evaluation in terms of absolute or relative benefits is usually not feasible (or, from a more subjectivistic point of view, not even possible). It is also unrealistic to expect that the IT itself will have a direct and easy to measure effect on the outcome quality of patient care, like in a drug trial (Wyatt, 1994). Thus, indirect measures are often used such as user satisfaction or changes of clinical processes, which, however, do not give a really complete picture of the benefits of IT. Often, changes in the evaluation questions may occur during the study (e.g., based on intermediate evaluation results, new insights, changes in stakeholders' opinions, or changes of the IT [scope creep]; Dewan & Lorenzi, 2000). Changes in study questions, however, may be difficult to balance with study resources. Finally, the selection of adequate evaluation designs and evaluation methods is often regarded as a problem during evaluation studies. Evaluators may not be sufficiently aware of the broadness of available approaches, or be too deeply embedded in either the qualitative or the quantitative paradigm, neglecting the possible contributions of the complementary approach. Thus, inadequate methods or study designs may be chosen which may not be able to answer the original study questions.

The following suggestions may be useful in order to deal with the complexity of the evaluation project. First, it is recommended that the general intention of the evaluation and the starting point should be agreed early on. In principle, evaluation should start before the new IT is implemented, in order to allow for early gathering of comparative data, and then continue during all phases of its life cycle (VATAM, 2000). Then, the areas of evaluation should be restricted to aspects which are of most importance to the involved stakeholders, and which can be measured with the available resources. A complete evaluation of all aspects of a system (such as economics, effectiveness, and acceptance) is usually not feasible. A balance between the resources of a study and the inclusion of the most relevant aspects has to be found. In addition, sufficient time should be invested into the definition of relevant study questions. All involved stakeholder groups should discuss and agree on the goals of evaluation (VATAM, 2000). The selected study questions should be relevant for decision-making with regard to introduction, operation or justification of IT (Ammenwerth et al., 2004). Conflicting goals should be discussed and solved, as they are not only problematic for an evaluation, but for the overall management of new IT. Fourth, when new evaluation questions emerge during the study, they should only be included in the study design when it is possible without creating problems. Otherwise, they should be tackled in consecutive studies. Each shift in evaluation questions must thoroughly be documented. For each study question, adequate methods must be chosen. A triangulation of methods may be useful to best answer the study questions (Heathfield, Pitty, & Hanka, 1998). For example, to address the effects of a nursing documentation system, both quantitative methods (time measurement, user acceptance scales, documentation quality measurement) as well as qualitative methods (focus group interviews) were used. We will discuss this example later on in more detail.

Third Problem Area: Motivation for Evaluation

An evaluation study can normally only be conducted when there is sufficient funding, and a sufficient number of participants (e.g., staff members, wards). Both these variables depend on the motivation of stakeholders (e.g., hospital management) to perform an evaluation. Sometimes, this motivation is not very high, because, for example, of fear for negative outcome, or of fear for revealing deficiencies of already implemented technology (Rigby, Forsström, Roberts, & Wyatt, 2001). In addition, the introduction of IT in an organization is a deep intervention that may have large consequences. It is thus often very difficult to organize IT evaluation in the form of an experiment, and to easily remove the system again at the end of the study in case the evaluation was too negative.

Even with a motivated management, it may be difficult to find suitable participants. Participating in a study usually requires some effort from the involved staff. In addition, while the users have to make large efforts to learn and use a new, innovative system, the benefit of joining a pilot study is usually not obvious (the study is conducted in order to investigate possible effects), but participation may even include some risks for the involved staff such as disturbances in workflow. In summary, due to the given reasons, the hospital management as well as involved staff members is often reluctant to participate in IT evaluation studies.

The described problem has consequences for the study. Without the support and motivation of the stakeholders to conduct an evaluation study, it will be difficult to get sufficient resources for an evaluation and sufficient participants willing to participate. Second, due to the given problems, the study organizer tends to recruit any participant who volunteers to participate. However, those participants may be more motivated to participate than the "normal" user. This leads to the well-known volunteer effect, where results are better when participants are motivated. In addition, evaluation results are not only important for the involved units, but also for the overall organization or for similar units in other organizations. To allow transfer of results, the pilot wards or pilot users must be sufficiently representative for other wards or users. But, as each IT within its environment is quite unique (see Problem Area 1); it is difficult to find comparable or representative participants.

To increase the number of participants, two approaches should be combined. First, the responsible management should be informed and motivated to support the study. The result of an evaluation study may be important to decide on new IT, and to support its continuous improvement. Then, the possible participants could be directly addressed. It should be made clear that the study provides the opportunity to influence not only the future development of IT in healthcare but also the own working environment. User feedback of study results may act as an important driving force for users to participate in the study. Offering financial compensation or additional staff for the study period may help to gain support from participants and from management. As in clinical trials, multicentric studies should be considered (Wyatt & Spiegelhalter, 1992). This would largely increase the number of available participants. This means however, that study management requires much more effort. A multicentric study design is difficult when the environment is completely different. In addition, the variation between study participants will be bigger in multicentric trials than in single-center ones. This may render interpretation and comparison of results even more difficult (cp. discussion in Problem Area 1).

Summary of General Recommendations

The above discussed problems and approaches will now be summarized in a list of 12 general recommendations for IT evaluation in healthcare:

1. Evaluation takes time; thus, take your time for thorough planning and execution.
2. Document all of your decisions and steps in a detailed study protocol. Adhere to this protocol; it is your main tool for a systematic evaluation.
3. Strive for management support, and try to organize long-term financial support.
4. Clarify the goals of the evaluation. Take into account the different stakeholder groups. Dissolve conflicting goals.
5. Reduce your evaluation questions to an appropriate number of the most important questions that you can handle within the available time and budget. If new questions emerge during the study, which cannot easily be integrated, postpone them for a new evaluation study.
6. Clarify and thoroughly describe the IT object of your evaluation and the environment. Take note of any changes of the IT and its environment during the study that may affect results.
7. Select an adequate study design. Think of a stepwise study design.
8. Select adequate methods to answer your study questions. Neither objectivist nor subjectivist approaches can answer all questions. Take into account the available methods. Consider being multimethodic and multidisciplinary, and consider triangulation of methods, data sources, investigators, and theories. Strive for methodical (e.g., biometrics) advice.
9. Motivate a sufficient number of users to participate. Consider multicentric trials and financial or other compensation.
10. Use validated evaluation instruments wherever possible.
11. Be open to unwanted and unexpected effects.
12. Publish your results and what you learned to allow others to learn from your work.

One of the most discussed aspects is the selection of adequate methods and tools (Point 6) and, here especially, the adequate application of multimethodic and multidisciplinary approaches (Ammenwerth et al., 2004). The interdisciplinary nature of evaluation research in medical informatics includes that a broad choice of evaluation methods is available for various purposes. In Sections II and III of this book, several distinct quantitative and qualitative evaluation methods have been presented and discussed in detail. All of them have their particular application area. However, in many situations, the evaluator may want to combine the methods to best answer the evaluation questions at hand. Especially in more formative (constructive) studies, a combination of methods may seem necessary to get a more complete picture of a situation. To support this, the method of triangulation has been developed and will now be presented in more detail.

THE THEORY OF TRIANGULATION

The term triangulation comes from navigation and means a technique to find the exact location of a ship base on the use of various reference points. Based on this idea, triangulation in evaluation means the multiple employments of data sources, observers, methods, or theories, in investigations of the same phenomenon (Greene & McClintock, 1985). This approach has two main objectives: First, to support a finding with the help of the others (validation); second, to complement the data with new results, to find new information, to get additional pieces to the overall puzzle (completeness; Knafl & Breitmayer, 1991).

Triangulation is, based on work by Denzin (1970), usually divided into the following four types, which can be applied at the same time:

- **Data triangulation:** Various data sources are used with regard to time, space, or persons. For example, nurses from different sites are interviewed, or questionnaires are applied at different times.
- **Investigator triangulation:** Various observers or interviewers with their own specific professional methodological background take part in the study, gathering and analyzing the data together. For example, a computer scientist and a social scientist analyze and interpret results from focus group interviews together.
- **Theory triangulation:** Data is analyzed based on various perspectives, hypotheses or theories. For example, organizational changes are analyzed using two different change theories.
- **Methods triangulation:** Various methods for data collection and analysis are applied. Here, two types are distinguished: within-method triangulation (combining approaches from the same research tradition), and between-method triangulation (combining approaches from both quantitative and qualitative research traditions, also called across-method triangulation). For example, two different quantitative questionnaires may be applied to access user attitudes, or group interviews as well as questionnaires may be applied in parallel.

It should be noticed that the term triangulation is only used when one phenomenon is investigated with regard to one research question.

The term triangulation is often seen strongly related to the term multimethod evaluation; because methods triangulation is seen as the most often used triangulation approach. However, as we want to stress, it is not limited on the combination of methods, but also describes combination of data sources, investigators, or theories.

Example: Triangulation during the Evaluation of a Nursing Documentation System

Background of the Study

Nursing documentation is an important part of clinical documentation. There have been some attempts and discussions on how to support the nursing documentation using computer-based documentation systems.

In 1997, Heidelberg University Medical Center started to introduce a computer-based nursing documentation system in order to systematically evaluate preconditions and consequences. Four different (psychiatric and somatic) wards were chosen for this study.

In the following paragraphs, we will concentrate on those parts of the study that are relevant for the triangulation aspects of the study. Please refer to other publications for more details on methods and results, such as (Ammenwerth, Mansmann, Iller, & Eichstädter, 2003; Ammenwerth et al., 2001).

Three of the four study wards had been selected by the nursing management for the study. On all three wards, the majority of nurses agreed to participate. Ward B volunteered to participate. The four study wards belonged to different departments. Wards A and B were psychiatric wards, with 21 resp. 28 beds; Ward C was a pediatric ward for children under two years of age, with

15 beds; Ward D was a dermatological ward, with 20 beds.

Our study wards were quite different with regard to nursing documentation. In Wards A and B, a complete nursing documentation based on the principles of the nursing process–for details on nursing process, see, for example, Lindsey and Hartrick (1996)–had been established for several years. In contrast, in Wards C and D, only a reduced care plan was documented; documentation was mostly conducted in the ward office. Only in Ward C, major parts of documentation were also conducted in the patients' rooms. The youngest staff member could be found in Ward D; the staff least experienced in computer use was in Ward C.

Study Design

The software PIK (Pflegeinformations-und Kommunikationssystem, a German acronym for "nursing information and communication system") was introduced on those four wards. The functionality covered the six phases of the nursing care process. The study period was between August 1998 and October 2001. Wards A and B started in 1998 with the introduction of the documentation system; Wards C and D joined in 2000.

The study consisted of two main parts: The objective of the more quantitative study was to analyze the changes in the nurses' attitudes with regard to nursing process, computers in nursing, and nursing documentation system, after the introduction of the computer-based system. Standardized, validated questionnaires were applied based on Bowman, Thompson, and Sutton (1983), for nurses' attitudes on the nursing process; on Nickell and Pinto (1986), for computer attitudes; on Lowry (1994), for nurses' attitudes on computers in nursing; and on Chin (1988) and Ohmann, Boy, and Yang (1997), for nurses' satisfaction with the computer-based nursing documentation system. We carefully translated those questionnaires into German and checked the understandability in a prestudy. We used a prospective intervention study with three time measurements: approximately three months before introduction ("before"); approximately three months after introduction ("during"); and approximately nine months after introduction ("after").

The second part of the study was a more qualitative study. Here, the objective was to further analyze the reasons for the different attitudes on the wards. The quantitative study exactly described these attitudes, and the qualitative study was now intended to further explain those quantitative results. The qualitative study was conducted in February 2002, after the analysis of the quantitative study was finished. In this qualitative study, open-ended focus group interviews were conducted with up to four staff members from each ward (most of them already have taken part in the quantitative study), with the three project managers from each department, and with the four ward managers from the wards. Open-ended means that the interviews were not guided by predefined questions. We used two general questions that started the interviews (e.g., "How are you doing with PIK?" "How was the introduction period")? The rest of the interview was mostly guided by the participants themselves, with relatively little control exerted by the interviewers.

All interviews were conducted by a team of two researchers. They took about one hour each. The interviews were audio taped and analyzed using inductive, iterative content analysis based on Mayring (1993). This means that the transcripts were carefully and stepwise analyzed, using the software WinMaxProf98.

In the following paragraphs, only those results of the quantitative and qualitative study relevant for the triangulation aspects of the study will be presented. Please refer to the already mentioned study publications for more details.

Results of Quantitative Analysis of User Attitudes

All in all, 119 questionnaires were returned: 23 nurses answered all three questionnaires, 17 nurses answered two, and 16 nurses answered one questionnaire. The return rates were 82% for the first questionnaire, 86.5% for the second questionnaire, and 90.2% for the third questionnaire. A quantitative analysis of the individual items of the questionnaires revealed unfavorable attitudes, especially in Ward C. In both Wards C and D, the nurses stated that the documentation system does not “save time” and does not “lead to a better overview on the course of patient care.” In addition, in Ward C, the nurses stated that they “felt burdened in their work” by the computer-based system and that the documentation system does not “make documentation easier.” In Wards A and B, the opinions with regard to those items were more positive.

The self reported daily usage of the computer-based documentation system was quite similar among all wards: about 1 to 2 hours a day during the second and third questionnaires, with highest values in Ward B and lowest values in Ward A. The self-confidence with the system, as stated by the nurses, was rather high on all wards during both the second and third questionnaire. The mean values were between 3 and 3.7 during the second questionnaire and between 3.4 and 3.8 during the third questionnaire (1=minimum, 4=maximum).

Statistical analysis revealed that the overall attitude on the documentation system during the third questionnaire was positively correlated to the initial attitude on the nursing process, to the attitude on computers in general and to the attitude on computers in nursing. Both computer attitude scores were in turn positively correlated to the years of computer experience. For details, see (Ammenwerth, Mansmann, et al., 2003).

Overall, the results of quantitative analysis pointed to a positive attitude on the computer-based nursing documentation already shortly after its introduction, which significant increase on three of the four wards later on. However, on ward C, the quantitative results revealed negative reactions, showing a heavy decline in the attitude scores during the second questionnaire. On ward C, the overall attitude of the computer-based system remained rather negative, even during the third questionnaire. What could be the reasons? In order to answer this question, a subsequent qualitative study was conducted.

Results of Qualitative Analysis of User Attitudes

This part of the study was conducted as planned. Overall, about 100 pages of interview transcript were analyzed. Details of the interviews are published elsewhere (Ammenwerth, Iller, et al., 2003); we will summarize only the main points.

In Ward C, some distinct features came up in the interviews that seem to have lead to low attitude scores at the beginning. For example, the nursing process had not been completely implemented before, thus the documentation efforts now were much higher. Documentation of nursing tasks covered a 24 hour day, due to the very young patients and their high need for care. Thus, the overall amount of documentation on Ward C was higher. Patient fluctuation was also highest in ward C. Nurses found it time-consuming to create a complete nursing anamnesis and nursing-care plan for each patient. The previous computer experience and number and availability of motivated key users was seen as rather low in Ward C. Then, during the introduction of the nursing documentation system, the workload was rather high in Ward C due to staff shortage, which increased pressure on the nurses. Finally, and most important, nursing documentation had previously at least partly been carried out in the patients’ rooms. However, during our study, computers were installed only in the ward office. No mobile computers were available, which, according to the nurses, lead to time-consuming and inefficient double documentation.

Interesting differences were found between the nurses and the project management. For example, the nurses stated in the interviews that they were not sufficiently informed on the new documentation system, while the project management stated to have offered information that had not been used. Another example is that the nurses felt that training was insufficient. In the opinion of the project management, sufficient opportunities had been offered. We will later see how this divergent information helps to complete the overall picture.

In Ward D, the attitude on the documentation system was high in the interviews. The nurses saw benefits, especially in a more professional documentation, which would lead to a greater acknowledgment of nursing. Standardized care planning was seen to make care planning much easier, without reducing the individuality of the patient. Overall, the ward felt at ease while working with the new documentation system.

In Wards A and B, the attitudes were also positive. The nurses stressed the better legibility of nursing documentation in the interviews. They said that time effort for nursing care planning was lower, but overall, time effort for nursing documentation was much higher than before. The interviews showed that the introduction period had been filled with anxiety and fear about new requirements for the nurses. Now, after some time, the nurses felt self-confident with computers. An interesting discussion arose on the topic of standardization. Most nurses felt that standardized care plans reduced the individuality of the care plans, and that they did not really reflect what is going on with the patient. Finally, those wards, too, mentioned insufficient teaching and support in the first weeks.

These rather short summaries, from the interviews, should highlight some distinct features of the wards, showing similarities (e.g., on insufficient teaching and fears at the beginning), but also differences (e.g., on the question on standardized care plans or time effort).

APPLICATION OF TRIANGULATION IN THIS STUDY

After analysis of the quantitative study and the qualitative study, we now want to see how the different results can be put together to get a broader picture of the effects and preconditions of a nursing documentation system. We thus applied all four types of triangulation as described by Denzin (1970):

- **Data triangulation:** Various data sources were used: Within the quantitative study, data triangulation with regard to time was used as the questionnaires were submitted three times to the same users (data triangulation with regard to time). In addition, in the interviews, not only nurses but also project management and ward management were interviewed (data triangulation with regard to persons).
- **Investigator triangulation:** Within the qualitative study, the two interviewers had different backgrounds (one more quantitative coming from medical informatics, the other, more qualitative, coming from social science). Both acted together as interviewers, analyzed the transcript together, and discussed and agreed on results and conclusions.
- **Theory triangulation:** We learned from various complementing theories to better understand the results of our studies. For example, to explain the implementation phases, we took ideas both from the book of Lorenzi and Riley (1995; first-, middle-, and second-order change) as well as from the change theory of (Lewin, 1947; unfreezing, moving, refreezing phase). With regard to user evaluation, we used the technology

acceptance model (TAM) of Davis (1993), and the task-technology-fit model (TTF) of Goodhue (1995).

- **Methods triangulation:** We applied between-methods triangulation by applying both quantitative questionnaires and qualitative focus-group interviews to investigate user's attitudes.

As stated in the introduction, triangulation has two main objectives: To confirm results with data from other sources (validation of results) and to find new data to get a more complete picture (completeness of results). We will now briefly discuss whether triangulation helped to achieve those goals.

Validation of Results

Validation of results is obtained when results from one part of the study are confirmed by congruent (not necessarily equal) results from other parts of the study. In our example, some parts of the study showed congruent results:

First, both the questionnaire and the interviews focused on attitudes issues. In this area, both approaches lead to congruent results, showing, for example, favorable attitudes in three wards. In addition, both the questionnaires and the interviews showed problems with regard to the user satisfaction with the nursing documentation system in Ward C. However, as the interviews were conducted later, they could better show the long-term development in the wards. Hence, both data sources thus showed congruent results.

Second, we found congruent results of the two scales attitudes on nursing process and attitude on the computer-based nursing documentation system within the standardized questionnaires. Both focus on different attitude items, both showed comparable low results in Ward C and higher results on the other wards, pointing to congruent measurements.

Those two selected examples show how results of some parts of the study could be validated by congruent results from other parts of the studies.

Completeness of Results

Besides validation, triangulation can increase completeness when one part of the study presents results which have not been found in other parts of the study. By this new information, the completeness of results is increased. The new information may be complementary to other results, or it may present divergent information.

In our study, both questionnaires and interviews presented partly complementary results, which led to new insights. For example, impact of the computer-based documentation system on documentation processes and communication processes had not been detected by the questionnaire (this aspect had not been included in the questions). However, the documentation system seems to have influenced the way different healthcare professionals exchanged patient-related information. This led to some discussion on this topic on all wards in the interviews and seems to have had an impact on the overall attitude. Those effects only emerged in the group interviews (and not in the questionnaires); enlarging the picture of the effects of the nursing documentation system and helping to better understand the reactions of the different wards.

Another example is the complementarity of the results in the interviews and questionnaires in Ward C. The interviews were done some time after the questionnaires. Thus, during this time, changes may have occurred. The change theory of Lewin (1947) stated that organizational changes occur in three phases: unfreezing (old patterns must be released, combined with insecurity and problems), moving (new patterns are tested), and refreezing (new patterns are internalized and seen as normal). The low attitude scores in Ward C, even at the last measurement point, indicate that the ward was in the moving phase during this time. During the interviews, the stress articulated by the nurses seems to be less severe. This can be interpreted as Ward C's slowly changing from the moving into the refreezing phase.

Triangulation can thus help to get a more complete picture of the object under investigation. Often, especially when applying various methods during the investigation, the results will not be congruent, but they may be divergent (e.g., contradictory). This is an important aspect of triangulation, as divergent results can especially highlight some points, present new information, and lead to further investigation.

In our study, we found some divergent results. For example, during the group interviews, nurses from one ward stressed that they do not see a reduction in effort needed for documentation by the computer-based system. However, in the questionnaires, this ward indicated strong time reductions. This differences can lead to the questions of whether time efforts are judged with regard to the situation without the nursing documentation system (where the amount of documentation was much lower, and so was the time effort), or with regard to the tasks that have to be performed (the same amount of documentation can be done much quicker with the computer-based system). This discussion can help one better understand the answers. Interesting differences of point of view could also be found between the staff and the project management of one ward in the group interviews. While the nurses of this ward claimed in the interviews that training was suboptimal, the project management stated that sufficient offers had been made. Those apparent contradictions may point to different perceptions of the need for training by the different stakeholders. Those insights may help to better organize the teaching on other wards.

As those (selected) examples show, triangulation helped us to obtain a better picture of the reaction of the four wards. The evaluation results also led to some decision on how to improve the technical infrastructure as well as how to better organize the teaching and support in some wards. All wards are still working with the computer-based nursing documentation system.

DISCUSSION

Medical informatics is an academic discipline and, thus, evaluation is an important part of any system development and implementation activity (Shahar, 2002; Talmon & Hasmann, 2002). However, many problems with regard to healthcare IT evaluation have been reported. Wyatt and Spiegelhalter (1992) as well as Grémy and Degoulet (1993) already discussed the complexity of the field, the motivation issue, and methodological barriers to evaluation. Examples of meta-analysis of IT evaluation studies confirm those barriers (e.g. Brender, 2002; Johnston, Langton, Haynes, & Mathieu, 1994; Kaplan, 2001).

In this chapter, we elaborated on a number of important problems and categorized them into three areas: the complexity of the evaluation object, the complexity of the evaluation project with its multitude of stakeholders, and the motivation for evaluation.

A kind of framework to support evaluation studies of ISs may be useful to address the problem areas discussed in this chapter. In fact, many authors have formulated the necessity for such a framework (e.g., Grant, Plante, & Leblanc, 2002; Shaw, 2002). Chapter XIV will present a framework for evaluation in more detail. One important part of such a framework is the call for a multimethod evaluation. While triangulation has long been discussed and applied in research (one of the first being Campbell & Fiske, 1959), the idea of the possible advantages of multimethod approaches or triangulation in more general terms is not really reflected in medical informatics literature.

The background of multimethod approaches has been more deeply discussed in Chapter XII.

In general, both quantitative and qualitative methods have their areas and research questions where they can be successfully applied. By triangulating both approaches, their advantages can be combined. We found that both complementary and divergent results from the different sources gave important new information and stimulation of further discussion.

In the past, there has been a more basic discussion about whether intermethods triangulation is possible at all. It is discussed that the epistemological underpinnings between quantitative and qualitative research paradigms may be so different that a real combination may not be possible (Greene & McClintock, 1985; Sim & Sharp, 1998). However, this argumentation is not taking into account that a tradition of research has formed beyond subjectivistic and objectivistic paradigms. Evaluation methods are chosen accordingly to research questions and the research topic. Thus, the question of which methods to apply and how to combine them only can be answered with respect to the research topic and the research question and not on a general basis. Thus, as important as this discussion might be in the light of progress in research methods, evaluation researchers in medical informatics may be advised to start to select and combine methods based on their distinctive research question. This gives evaluation researchers a broad range of possibilities to increase both completeness and validity of results, independent of his or her research tradition.

CONCLUSIONS

Evaluation studies in healthcare IT take a lot of time, resources, and know-how. Clearly defined methodological guidelines that take the difficulties of IS evaluation in healthcare into account may help to conduct better evaluation studies. This chapter has classified some of the problems encountered in healthcare IT evaluation under the three main problem areas of a) complexity of the evaluation object, b) complexity of the evaluation project, and c) limited motivation for evaluation. We suggested a list of 12 essential recommendations to support the evaluation of ISs. A broadly accepted framework for IT evaluation in healthcare that is more detailed seems desirable, supporting the evaluator during planning and executing of an evaluation study.

Focusing on methodological aspects, we have presented some basics on triangulation and illustrated them in a case study. The correct application of triangulation requires–as other evaluation methods–training and methodological experience. Medical informatics evaluation research may profit from this well-established theory.

ACKNOWLEDGMENT

This chapter is based on two earlier publications, published by Elsevier (Ammenwerth, Gräber, Herrmann, Bürkle, & König, 2003; Ammenwerth, Mansmann, & Iller, 2003).

REFERENCES

Ammenwerth, E., Brender, J., Nykänen, P., Prokosch, U., Rigby, M., & Talmon, J. (2004). Visions and strategies to improve evaluations of health information systems-Reflections and lessons based on the HIS-EVAL workshop in Innsbruck. International Journal of Medical Informatics, 73(6), 479-91.

Ammenwerth, E., & de Keizer, N. (2004). An inventory of evaluation studies of information technology in healthcare: Trends in evaluation research 1982-2002. Methods of Information in Medicine, 44, 44-51.

Ammenwerth, E., Eichstädter, R., Haux, R., Pohl, U., Rebel, S., & Ziegler, S. (2001). A randomized evaluation of a computer-based nursing documentation system. Methods of Information in Medicine, 40(2), 61-8.

Ammenwerth, E., Gräber, S., Herrmann, G., Bürkle, T., & König, J. (2003). Evaluation of health ISs-Problems and challenges. International Journal of Medical Informatics, 71(2/3), 125-135.

Ammenwerth, E., Iller, C., Mahler, C., Kandert, M., Luther, G., Hoppe, B., et al. (2003). Einflussfaktoren auf die Akzeptanz und Adoption eines Pflegedokumentationssystems-Studienbericht (2/2003). Innsbruck: Private Universität für Medizinische Informatik und Technik Tirol.

Ammenwerth, E., Mansmann, U., & Iller, C. (2003). Can evaluation studies benefit from a triangulation of quantitative and qualitative methods? A case study. International Journal of Medical Informatics, 70(2/3), 237-248.

Ammenwerth, E., Mansmann, U., Iller, C., & Eichstädter, R. (2003). Factors affecting and affected by user acceptance of computer-based nursing documentation: Results of a two-year study. Journal of the American Medical Informatics Association. 10(1), 69-84.

Berg, M., Aarts, J., & van der Lei, J. (2003). ICT in healthcare: Sociotechnical approaches. Methods of Information in Medicine, 42, 297-301.

Bowman, G., Thompson, D., & Sutton, T. (1983). Nurses attitudes towards the nursing process. Journal of Advanced Nursing, 8(2), 125-129.

Brender, J. (2002). Methodological and methodical perils and pitfalls within assessment studies performed on IT-based solutions in healthcare. Aalborg: Virtual Centre for Health Informatics.

Campbell, D., & Fiske, D. (1959). Convergent and discriminant validity by the muli-trait, multi-method matrix. Psychological Bulletin, 56, 81-105.

Chin, J. (1988). Development of a tool measuring user satisfaction of the human-computer interface. Proceedings: Human factors in Computing, Chi'88 Conf (pp. 213-218). New York: Association for Computing Machinery.

Davis, F. (1993). User acceptance of information technology: System characteristics, user perceptions and behavioral impacts. International Journal of Man-Machine Studies, 38, 475-487.

Denzin, N. (1970). Strategies of multiple triangulation. In N. Denzin (Ed.), The research act (3rd ed., pp. 297-331). Chicago: Aldine.

Dewan, N., & Lorenzi, N. (2000). Behavioral health information systems: evaluating readiness and user acceptance. MD Computing, 17(4), 50-52.

Frechtling, J. (1997). User-friendly handbook for mixed method evaluation [Ausdruck im Eval-Ordner]. Retrieved May 2005, from http://www.ehr.nsf.gov/EHR/REC/pubs/NSF97-153/start.htm

Friedman, C., & Wyatt, J. C. (1997). Evaluation methods in medical informatics. New York: Springer.

Goodhue, D. (1995). Understanding user evaluations of information systems. Management Science, 41(12), 1827-44.

Grant, A., Plante, I., & Leblanc, F. (2002). The TEAM methodology for the evaluation of information systems in biomedicine. Computers in Biology and Medicine, 32(3), 195-207.

Greene, J., & McClintock, C. (1985). Triangulation in evaluation: Design and analysis issues. Evaluation Review, 9(5), 523-545.

Grémy, F., & Degoulet, P. (1993). Assessment of health information technology: Which questions for which systems? Proposal for a taxonomy. Medical Informatics, 18(3), 185-193.

Heathfield, H., Hudson, P., Kay, S., Mackay, L., Marley, T., Nicholson, L., et al. (1999). Issues in the multi-disciplinary assessment of healthcare information systems. Assessment of Healthcare Information Technology & People, 12(3), 253-275.

Heathfield, H., Pitty, D., & Hanka, R. (1998). Evaluating information technology in health care: Barriers and challenges. British Medical Journal, 316, 1959-1961.

Institute of Medicine. (2001). Crossing the quality chasm: A new health system for the 21st century. Washington: National Academy Press.

Johnston, M., Langton, K., Haynes, R., & Mathieu, A. (1994). Effects of computer-based clinical decision support systems on clinician performance and patient outcome: A critical appraisal of research. Annuals of Internal Medicine, 120, 135-142.

Kaplan, B. (2001). Evaluating informatics applications: Clinical decision support systems literature review. International Journal of Medical Informatics, 64, 15-37.

Kaplan, B., & Shaw, N. (2002). People, organizational and social issues: Evaluation as an exemplar. In R. Haux & C. Kulikowski (Eds.), Yearbook of medical informatics 2002 (pp. 91-102). Stuttgart, Germany: Schattauer.

Knafl, K., & Breitmayer, B. (1991). Triangulation in qualitative research: Issues of conceptual clarity and purpose. In J. Morse (Ed.), A contemporary dialogue (pp. 226-239). Newbury Park, CA: Sage.

Lewin, K. (1947). Frontiers in group dynamics: Concepts, methods, and reality of social sciences: Social equalization and social change. Human Relations, 1, 5-14.

Lindsey, E., & Hartrick, G. (1996). Health-promoting nursing practice: The demise of the nursing process? Journal of Advanced Nursing, 23(1), 106-112.

Lorenzi, N., & Riley, R. (1995). Organizational aspects of health informatics - Managing technological change. New York: Springer.

Lowry, C. (1994). Nurses' attitudes toward computerised care plans in intensive care. Part 2. Intensive and Critical Care Nursing, 10, 2-11.

Mayring, M. (1993). Einführung in die qualitative Sozialforschung. Weinheim: Psychologie-Verlag-Union.

Moehr, J. R. (2002). Evaluation: Salvation or nemesis of medical informatics? Computers in Biology and Medicine, 32(3), 113-125.

Nickell, G., & Pinto, J. (1986). The computer attitude scale. Computers in Human Behaviour, 2, 301-306.

Ohmann, C., Boy, O., & Yang, Q. (1997). A systematic approach to the assessment of user satisfaction with healthcare systems: Constructs, models and instruments. In C. Pappas (Ed.), Medical Informatics Europe '97. Conference proceedings (Vol. 43, Pt. B, pp. 781-5). Amsterdam: IOS Press.

Palvia, S., Sharma, R., & Conrath, D. (2001). A socio-technical framework for quality assessment of computer information systems. Industrial Management & Data Systems, 101(5), 237-251.

Rigby, M. (2001). Evaluation: 16 powerful reasons why not to do it-And 6 over-riding imperatives. In V. Patel, R. Rogers, & R. Haux (Eds.), Proceedings of the 10th World Congress on Medical Informatics (Medinfo 2001,Vol. 84, pp. 1198-1202). Amsterdam: IOS Press.

Rigby, M., Forsström, J., Roberts, R., & Wyatt, W. (2001, September 8). Verifying quality and safety in health informatics services. British Medical Journal, 323, 552-556.

Shahar, Y. (2002). Medical informatics: Between science and engineering, between academia and industry. Methods of Information in Medicine, 41, 8-11.

Shaw, N. (2002). CHEATS: a generic information communication technology (ICT) evaluation framework. Computers in Biology and Medicine, 32, 209-200.

Sim, J., & Sharp, K. (1998). A critical appraisal of the role of triangulation in nursing research. International Journal of Nursing Studies, 35(1/2), 23-31.

Talmon, J., & Hasmann, A. (2002). Medical informatics as a discipline at the beginning of the 21st century. Methods of Information and Medicine, 41, 4-7.

van der Loo, R. (1995). Overview of published assessment and evaluation studies. In E. M. S. J. van Gennip & J. S. Talmon (Eds.), Assessment and evaluation of information technologies (pp. 261-82). Amsterdam: IOS Press.

VATAM. (2000). The VATAM Websites. Validation of health telematics applications (VATAM). Retrieved May 2005, from http://www-vatam.unimaas.nl

Winter, A., Ammenwerth, E., Bott, O., Brigl, B., Buchauer, A., Gräber, S., et al. (2001). Strategic information management plans: The basis for systematic information management in hospitals. International Journal of Medical Informatics, 64(2/3), 99-109.

Wyatt, J. (1994). Clinical data systems, part 3: Development and evaluation. The Lancet, 344, 1682-1688.

Wyatt, J., & Spiegelhalter, D. (1992). Field trials of medical decision-aids: Potential problems and solutions. In P. Clayton (Ed.), 15th Annual Symposium on Computer Applications in Medical Care (pp. 3-7). New York: McGraw-Hill.

This work was previously published in E-Health Systems Diffusion and Use: The Innovation, the User and the USE IT Model, edited by T. A. M. Spil and R. W. Schuring, pp. 212-236, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 1.2
Assessing E-Health

Nilmini Wickramasinghe
Illinois Institute of Technology, USA

Elie Geisler
Illinois Institute of Technology, USA

Jonathan Schaffer
The Cleveland Clinic, USA

ABSTRACT

While healthcare is the biggest service industry on the globe, it has yet to realize the full potential of the e-business revolution in the form of e-health. This is due to many reasons, including the fact that the healthcare industry is faced with many complex challenges in trying to deliver cost-effective, high-value, accessible healthcare and has traditionally been slow to embrace new business techniques and technologies. Given that e-health to a great extent is a macro-level concern that has far reaching micro-level implications, this chapter first develops a framework to assess a country's preparedness with respect to embracing e-health (i.e., the application of e-commerce to healthcare) and from this, an e-health preparedness grid to facilitate the assessment of any e-health initiative. Taken together, the integrative framework and preparedness grid provide useful and necessary tools to enable successful e-health initiatives to ensue by helping country and/or organization within a country to identify and thus address areas that require further attention in order for it to undertake a successful e-health initiative.

INTRODUCTION

Technology has the potential to help solve many of the problems faced by developed and developing countries alike; from improving healthcare delivery to opening up commerce opportunities. The number of nations offering e-commerce solutions is increasing every year, and consequently the number of workplaces that have Internet connection for business activities has been rapidly growing worldwide. Table 1 provides three estimates of e-commerce forecasts. These projections clearly highlight the important role of e-commerce on the global economy. However, as the Secretary General of the United Nations notes in his forward to the e-commerce and development report (UNCTAD, 2002), "knowing that an instrument is powerful is not enough to ensure that it will be put to the best use. We need to understand how it works, and how and when it should be used … and maximize its power."

Within the umbrella of e-commerce, one area, e-health, has yet to reach its full potential in many developed countries, let alone developing countries. Each country is positioned differently and has varying potential and preparedness regarding embracing e-commerce technologies generally and e-health in particular. Given the macro-level nature of many issues pertaining to the development of e-health (Alvarez, 2002), in order to be more effective in their e-health initiatives, it is important for countries to assess their potential, identify their relative strengths and weaknesses, and thereby develop strategies and policies to address these issues to effectively formulate and implement appropriate e-health initiatives. To do this effectively, it is valuable to have an integrative framework that enables the assessment of a country's e-health preparedness. This chapter proposes to develop such a framework that can be applied to various countries throughout the globe and from this to generate an e-health preparedness grid. In so doing, we hope to facilitate better understanding of e-health initiatives and thus maximize their power.

E-HEALTH

Reducing healthcare expenditure as well as offering quality healthcare treatment is becoming a priority globally. Technology and automation have the potential to reduce these costs (Institute of Medicine, 2001; Wickramasinghe, 2000), thus, e-health, which essentially involves the adoption and adaptation of e-commerce technologies throughout the healthcare industry (Eysenbach, 2001), appears to be a powerful force of change for the healthcare industry worldwide.

Healthcare has been shaped by each nation's own set of cultures, traditions, payment mechanisms, and patient expectations. Therefore, when looking at health systems throughout the world, it is useful to position them on a continuum ranging from high government involvement (i.e., a public healthcare system) at one extreme, to little government involvement (i.e., private healthcare system) at the other extreme, with many variations of a mix of private–public in between. However, given the common problem of exponentially increasing costs facing healthcare globally, irrespective of the particular health system one examines, the future of the healthcare industry will be partially shaped by commonalties such as the universal issue of escalating costs and the common forces of change including a) empowered consumers, b) e-health adoption and adaptability, and c) shift to focus on the practice of preventative versus cure driven medicine, as well as four key implications,

Table 1. Worldwide e-commerce estimates and forecasts (in billion $) (Source: From United Nations Conference on Trade and Development, 2002. Available online from http://r0.unctad.org)

Source	2000	2001	2002	2003	2004	2005	2006
Forester			**2,293.50**	**3,878.80**	**6,201.10**	**9,240.60**	**12,837.30**
IDC	**354.90**	**615.30**				**4,600.00**	
Emarketer (B2B only)	**278.19**	**474.32**	**823.48**	**1,208.57**	**2,367.47**		

namely a) health insurance changes, b) workforce changes and changes in the roles of stakeholders within the health system, c) organizational changes and standardization, and d) the need for healthcare providers and administrators to make difficult, yet necessary choices regarding practice management.

E-health is a very broad term that encompasses various activities related to the use of many e-commerce technologies and infrastructures, most notably the Internet, for facilitating healthcare practice. The World Health Organization (WHO, 2003), a major world health body, defines e-health as "being the leveraging of the information and communication technology (ICT) to connect provider and patients and governments; to educate and inform health care professionals, managers and consumers; to stimulate innovation in care delivery and health system management; and, to improve our healthcare system". In contrast, a technologically oriented definition of e-health is offered by Intel, which refers to e-health as "a concerted effort undertaken by leaders in healthcare and hi-tech industries to fully harness the benefits available through convergence of the Internet and healthcare".

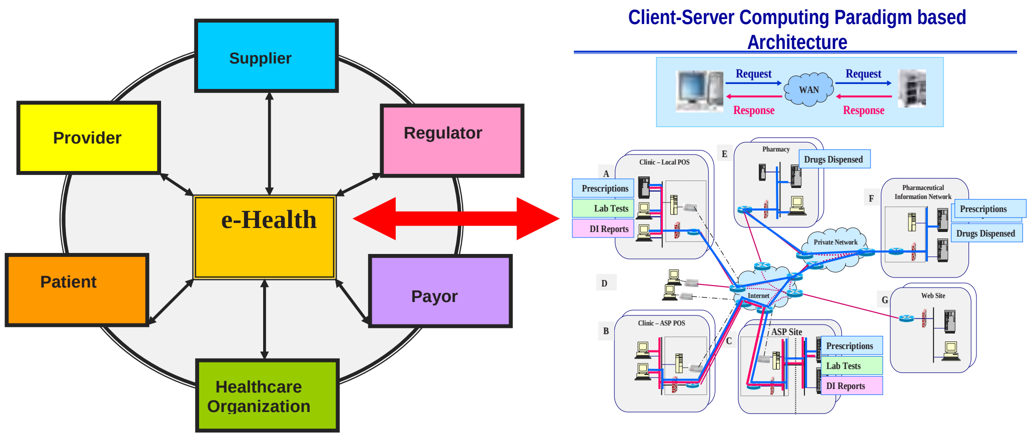
Health professionals are increasingly being drawn into evaluating the Internet as a source of consumer information for health and medicine. Practitioners report that a growing number of patients arrive at their offices either with questions related to appropriate Web sites to visit or a large variety of health-related content gathered from the Internet. Some of this content may prove extremely helpful to the health and/or recovery of a patient. Because the Internet has created new opportunities and challenges to the traditional healthcare information technology industry, the advent of e-health seems fitting to address both these opportunities and challenges. The new possibilities for facilitating effective healthcare delivery fall primarily into the following main categories: a) the capability of healthcare consumers to interact with their providers online (B2C = "business to consumer"), b) the possibility to improve healthcare institution-to-institution transmissions of data (B2B = "business to business"), and c) The new possibilities for peer-to-peer communication of healthcare consumers (C2C = "consumer to consumer").

A more comprehensive definition of e-health would need to incorporate the healthcare, business, and technological perspectives; hence we define e-health as an emerging field at the intersection of medical informatics, technology, public health, and business. Thereby, e-health entails the delivery of health services and health information through the Internet and other related e-commerce technologies. In a broader sense, the term characterizes not only a technical development, but also a state-of-mind, a paradigm shift, and a commitment for networked global thinking to improve healthcare locally, regionally, and globally by using information and communication technologies (Geisler,1999).

WEB OF PLAYERS IN HEALTHCARE

Figure 1 depicts the Web of healthcare players and the key elements of the e-health architecture that serves to support the interactions between and within this Web of players. What is of particular importance when trying to assess e-health initiatives is a) it is imperative to understand the implications of any e-health initiative on all the players who make up this Web, and b) to realize the full potential of the Internet or Web-based technologies that make up the IT architecture so that the most effective and efficient e-health initiative is, in fact, designed and implemented. To get a better appreciation for this we must first understand the traditional competitive forces affecting any organization and then the role of the Internet in impacting these forces (Geisler, 2001).

Figure 1. Web of e-health players (Source: Wickramasinghe et al., 2004)



TRADITIONAL COMPETITIVE FORCES

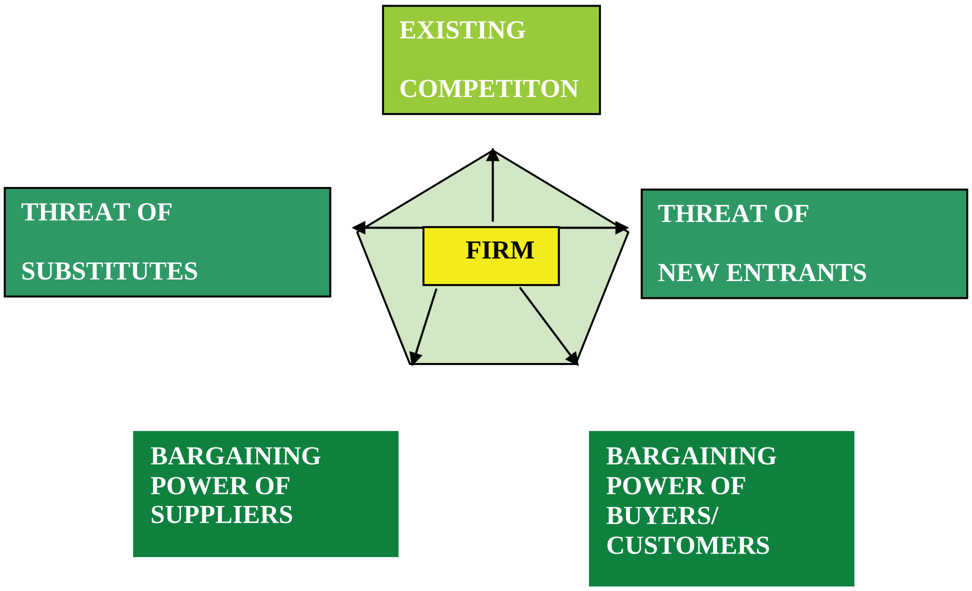
The strategy of an organization has two major components (Henderson & Venkatraman, 1993). These are a) formulation-making decisions regarding the mission, goals and objectives of the organization, and b) implementation-making decisions regarding how the organization can structure itself to realize its goal and carryout specific activities. For today's healthcare organizations, the goals, mission and objectives all focus around access, quality and value and realizing this value proposition for healthcare then becomes the key (Wickramasinghe et al., 2004). Essentially, the goal of strategic management is to find a "fit" between the organization and its environment that maximizes its performance (Hofer, 1975). This then describes the market-based view of the firm, developed by Michael Porter (1980). The first of Porter's frameworks is the generic strategies.

The use of technology must always enable or enhance the businesses objectives and strategies of the organization. This is particularly true for 21st century organizations, where many of their key operations and functions are so heavily reliant on technology and the demand for information and knowledge is so critical. A firms' relative competitive position (i.e., its ability to perform above or below the industry average) is determined by its competitive advantage. Porter (1980) identified three generic strategies that impact a firm's competitive advantage. These include cost, focus, and differentiation. Furthermore, Porter (1980) noted that two, and only two, basic forms of competitive advantage typically exist: cost leadership and differentiation.

Firms can use these two forms of competitive advantage to either compete across a broad scope of an industry or to focus on competing in specific niches, thereby leading to three generic strategies. Porter (1980) noted that firms should be cautious about pursuing more than one generic strategy, namely cost, differentiation, and focus. For example, if a cost leadership strategy is adopted, it is unlikely that a firm can also maintain and sustain differentiation, because it would not be possible to simultaneously pursue the costly capital investment or maintain high operating costs required for differentiation and, thus, in the long run, the firm has a confused strategy which leads to failure.

In order to design and develop one's strategy, an organization should first perform an industry analysis. Porters five forces (or competitive forces) model is most useful (Porter, 1980, 1985). Figure 2 depicts this model. Essentially, Porter has taken concepts from microeconomics and modeled them

Figure 2. Porter's competitive (five) forces model



in terms of five key forces that together outline the rules of competition and attractiveness of the industry.

The forces are as follows:

1. **Threat of new entrant:** a company new to the industry that could take away market share from the incumbent firms
2. **Threat of substitute:** an alternative means that could take market share from product/ service offered by the firms in the industry
3. **Bargaining power of buyers:** the strength of buyers or groups of buyers within the industry relative to the firms
4. **Bargaining power of suppliers:** the strength of suppliers relative to the firms in the industry
5. **Rivalry of existing competition:** relative position and market share of major competitors

The collective strength of these five forces determines the attractiveness of the industry and thus the potential for superior financial performance by influencing prices, costs, and the level of capital investment required (Porter, 1985). Once a thorough industry analysis has occurred, it is generally easier for a firm to determine which generic strategy makes most sense for it to pursue and enables the firm to exploit most of its core competencies in its existing environment.

In this chapter we propose the application of Porter's (1985) five forces model to the Web of e-health players shown in Figure 1. Healthcare delivery organizations and the industry that supports them are competing and cooperating for patients and resources. They compete as suppliers of services, technology, and other industrial resources. This competition is among the entities in the healthcare sector as well as with organizations in other industries. The Web of e-health serves as an instrument that mediates the relations among the five forces (Geisler, 1999). This in turn creates several opportunities for the organizations involved to improve and enhance their abilities to navigate the competitive environment of their industry.

Table 2. The three e-opportunity domains and their components

Domain	Components
e-operations	• Automation of administrative processes • Supply-chain reconfiguration • Reengineering of primary infrastructure • Intensive competitive procurement • Increased parenting value
e-marketing	• Enhanced selling process • Enhance customer usage experience • Enhanced customer buying experience
e-services	• Understanding of customer needs • Provision of customer service • Knowledge of all relevant providers • Negotiation of customer requirements • Construction of customer options

ROLE OF THE INTERNET OF THE COMPETITIVE FORCES

Feeny (2001) presented a framework that highlights the strategic opportunities afforded to organizations by using the Internet. In particular, he highlights three e-opportunity domains. Table 2 details these domains and their respective components.

E-OPPORTUNITIES IN HEALTHCARE

Given the three areas of e-opportunities discussed previously, Glaser (2002) identified several key e-opportunities for healthcare, detailed in Table 3.

The Goals of E-Health

The preceding definition of e-health is broad enough to capture the dynamic environment of the Internet and at the same time acknowledge that e-health encompasses more than just "Internet and Medicine." Thus, e-health strives to achieve many goals beyond the use of the Internet. These goals, taken together perhaps best characterize what e-health is all about (or what it should be about; JMIR, 2003). Some of these goals of e-health include the following:

- **Efficiency:** One of the promises of e-health is to increase efficiency in heathcare, thereby decreasing costs. One possible way of decreasing costs would be by avoiding duplicative or unnecessary diagnostic or therapeutic interventions, through enhanced communication possibilities between healthcare establishments and through patient involvement (Health Technology Center, 2000). The Internet will naturally serve as an enabler for achieving this goal in e-health.
- **Quality of care:** Increasing efficiency involves not only reducing costs, and thus is not an end in and of itself but rather should be considered in conjunction with improving quality one of the ultimate goals of e-health.

Table 3. The e-opportunities for healthcare organizations

Domain	Components
e-operations	• Internet-based supply purchasing • Prescription writing, formulary checking, and interaction checking using hand-held devices
e-marketing	• Delivery of consumer health content and wellness management tools over the Internet • Use of consumer health profiles to suggest disease management and wellness programs
e-services	• Patient-provider communication and transaction applications • Web-based applications to support the clinical conversation between referring and consulting physicians
Crossing multiple domains	• Increasing the level of information content in the product • Increasing the information intensity along the supply chain • Increase in the dispersion of information

A more educated consumer (as a result of the informational aspects of e-health) would then communicate more effectively with their primary care provider which will, in turn, lead to better understanding and improved quality of care.

- **Evidence-based:** E-health interventions should be evidence-based in the sense that their effectiveness and efficiency should not be assumed but proven by rigorous scientific evaluation and support from case histories. Web-accessible case repositories facilitate the timely accessibility of such evidence and thus help in achieving the necessary support of a diagnosis or treatment decision. The evidence-based medicine goal of e-health is currently one of the most active e-health research domains, yet much work still needs to be done in this area.
- **Empowerment of consumers and patients:** By making the knowledge bases of medicine and personal electronic records accessible to consumers over the Internet, e-health opens new avenues for patient-centered medicine, enables patient education and thus increases the likelihood of informed and more satisfactory patient choice (Umhoff & Winn, 1999).
- **Education:** The education of physicians through online sources (continuing medical education) and consumers (health education, tailored preventive information for consumers) makes it easier for physicians as well as consumers to keep up to date with the latest developments in the medical areas of their respective interests. This, in turn, is likely to have a positive impact on the quality of care vis-à-vis the use of the latest medical treatments and preventive protocols.
- **Extending:** Extending the scope of healthcare beyond its conventional boundaries, in both a geographical sense as well as in

a conceptual sense, leads to enabling such techniques as telemedicine and virtual operating rooms, both of which are invaluable in providing healthcare services to places where it may otherwise be difficult or impossible to do.

- **Ethics:** E-health involves new forms of patient-physician interaction and poses new challenges and threats to ethical issues such as online professional practice, informed consent, privacy and security issues (Healthcare Advisory Board, 2002). However, this is not an intrinsic feature of e-health but rather a feature of the Internet technology which is the foundation for all e-business initiatives, therefore, e-health along with e-government, e-insurance, e-banking, e-finance, and e-retailing must all contend with these ethical issues. Given the nature of healthcare, these issues could be more magnified.
- **Equity:** To make healthcare more equitable is one of the aims of quality generally identified by the American Institute of Medicine (Institute of Medicine, 2001) and is one of the goals of e-health. However, at the same time there is a considerable threat that e-health, if improperly implemented and used, may deepen the gap between the haves and the have nots, hence the need for a robust framework to ensure the proper implementation of e-health initiatives. In particular, some of the key issues for equity revolve around broad access and familiarity with the technology.

Today, a large number of patients and consumers already use the Internet to retrieve health-related information, to interact with health providers, and even to order pharmaceutical products (for example drugstore.com). For example, it has been noted that "the number of Medline searches performed by directly accessing the database at the National Library of Medicine increased from 7 million in 1996 to 120 million in 1997, when free public access was opened; the new searches are attributed primarily to non-physicians" (JMIR, 1999). Thus, the Internet will act as a catalyst for evidence based medicine in particular and e-health generally in two ways: First, it enables health professionals to access timely evidence. Second, it enables consumers to draw from the very same knowledge base, leading to increased pressure on health professionals to actually use the evidence (PWC, 2003).

Key Challenges of E-Health

By 2005, 88.5 million adults will use the Internet to find health information, shop for health products, and communicate with affiliated payers and providers through online channels, according to Cyber Dialogue (PWC, 2003). Today, the e-health consumer demand includes the need for specific health services, such as obtaining information when faced with a newly established diagnosis.

Many key challenges must be addressed to develop optimal partnerships between consumers and other groups of healthcare stakeholders. Some of these include the need for:

- Meaningful collaboration with healthcare recipients,
- Efficient strategies and techniques to monitor patterns of Internet use among consumers,
- Preparation for upcoming technological developments,
- Balancing between connectivity and privacy,
- Better understanding of the balance between face-to-face and virtual interactions, and
- Equitable access to technology and information across the globe.

When we consider the domain of e-health at the macro level, three important issues must be carefully considered (Cyber Dialogue, 2001);

namely, Procurement, Connectivity, and Benefits. We briefly discuss each in turn.

E-Procurement

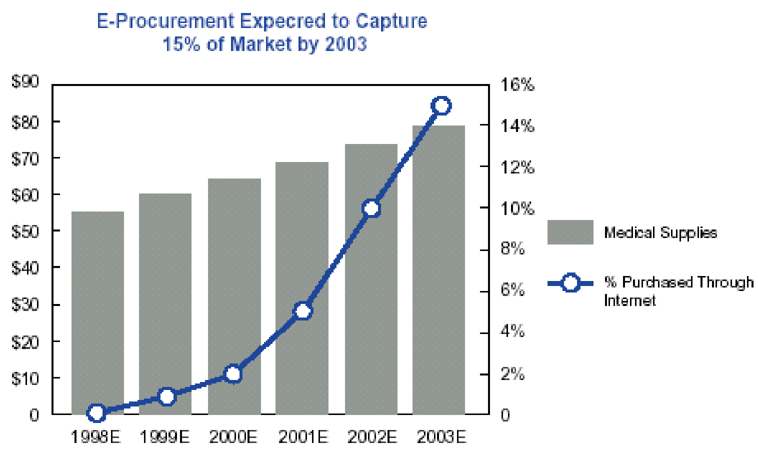
Health systems must begin to contemplate how their organizations will adapt and leverage Internet-based tools to manage their medical supply chains. Procurement in healthcare supplies must move toward an e-business platform for data interchange because of the ubiquity and cost effectiveness of the Internet, the primary e-business platform. The Internet decreases many of the restrictions placed by geographic and time barriers and also facilitates the use of incorporating artificial intelligence solutions, such as intelligent agents that can be deployed to hunt for best buys. Buyers and suppliers also must work together toward standardization, including the development of a universal product numbering system to facilitate e-procurement. As shown in Figure 3, e-procurement of medical supplies is estimated to grow to 15% of medical supply spending in the U.S. by 2003, according to a Deutsche Bankreport (PWC, 2003)

E-Connectivity

Healthcare has traditionally been locally delivered because a patient's usual first port of call is their local primary care physician. To reorient such a tradition, connectivity companies, which on the other hand have a global rather than local focus, must be pragmatic and move in incremental steps when connecting healthcare organizations. Technology is the integral tool, but it will not achieve its full potential or live up to its promise unless healthcare organizations successfully deploy it and then keep track of their clinicians' and administrative workers' effective use of it. To do so, managers must design processes and metrics for productivity; otherwise, it is like expecting someone to drive a car when his or her experience is limited to a 10 speed bicycle. However, healthcare organizations will find that achieving Web-enabled connectivity offers the most opportunity initially, and that other functions such as disease management, outcomes management, and demand management can be Web-enabled later.

Health plans and hospitals are beginning to migrate to the Internet for claims-related transactions as the first step of a broader Internet strategy. This

Figure 3. Estimated U. S. medical e-procurement market sales (Source: Dialog Cybercitizen Health Trend Report [Re. No. 49], by Odyssey Research, 2003. Available online from www.odysseyresearch.org)



is illustrated in Figure 4. Because many organizations continue to use EDI for claims submissions, transactions surrounding claims (e.g., eligibility, referrals) will thus logically be the first to be targeted for e-health connectivity. Those health plans that are adopting Internet connectivity for these functions view them as the foundation on which to build other Internet-enabled partnerships with patients and providers.

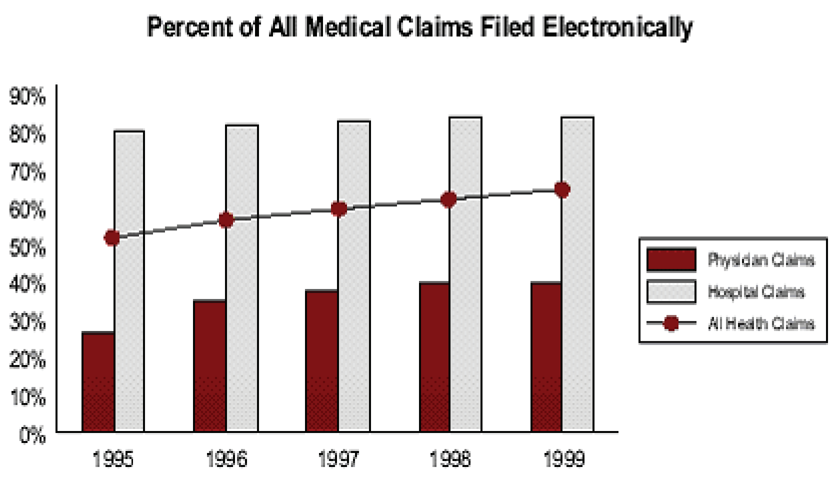
The number of healthcare transactions is outpacing the growth of health spending, creating a critical need to automate the handling of such transactions through the adoption of e-health. The number of claims submitted increased by about 7% during the past five years, according to the Health Data Directory. In contrast, healthcare spending has increased at between 5% and 6% during the same period (ibid). Many of the functions associated with claims submissions and payment are repetitive tasks that are more efficiently done by computers. The most expensive processes are not the claims submissions themselves, but the tasks surrounding the claims processing, such as eligibility checks and referrals. Coupled with increasing labor shortages, there is an urgent need for healthcare organizations to reengineer their processes.

Health plans must understand physicians' needs when designing Internet-based solutions and that first-mover advantage is not as important as a system that works. Many non-healthcare organizations have benefited from first-mover advantage in adopting e-commerce initiatives, which means they had access to the most capital, the best partners and could establish solid brand recognition. However, this is not necessarily an appropriate strategy for healthcare as healthcare is more pragmatic and a "show-me" industry where successful models replicate specialty by specialty.

E–Benefits

American industry has discovered the benefits of e-business. Like the conundrum of the chicken and the egg, many employers and health plans are awaiting the development and implementation of e-benefits and e-insurance products. Health plans do not want to deliver Web-based products if employers are not ready to use them. Employers

Figure 4. Increase in the percentage of electronically filed claims (Source: Dialog Cybercitizen Health Trend Report [Re. No.. 49], by Odyssey Research, 2003. Available online from www.odysseyresearch.org)



cannot deliver e-benefits products until health plans develop them. However, starting with online benefits enrollment, this aspect of e-health is evolving in stages. A by-product of this evolution is the fusion of employee responsibility and empowerment; hence, employers will gradually cede more control for health benefits to employees themselves. One of the primary drivers of e-benefits is delivery of self-service capabilities in which employees can customize their own insurance plans and have ready access to them, just as they do with their brokerage accounts. By putting this information at their fingertips, employees may become more fiscally responsible about those benefits. Ultimately, they may want complete control over more and more aspects of their benefits. To empower employees toward that end, some employers will have to embrace a defined contribution model, also called self-directed or consumer-directed health plans. This builds on the findings of the report "Defined Contribution Healthcare," which specifically discusses the various models of defined contribution health plans (ibid).

The national research conducted by PricewaterhouseCoopers (PWC, 2003) indicated that few employers are willing to adopt self-directed, otherwise known as defined contribution, health plans today, but will more likely move incrementally toward Web-enabling benefits processing and hence shifting more responsibility to employees. As employers adopt more e-health initiatives and Web-enabled functions, they will move more responsibility for choices to employees. As that balance of responsibility tips towards the employees, employers will need to assess how ready their workforce is to accept increasing levels of responsibility (Geisler, 2000).

If a significant portion of the healthcare insurance market moves to self-directed and Internet health benefit accounts, physicians may be most directly and adversely affected. Some contribution health plans have medical savings accounts as a centerpiece. In these accounts, employees pay out of a medical savings account for routine expenses up to $1,500 or $2,000. Physicians will need to contend with patients who are paying cash for their visits and who may shop around for the best value. Figure 5 shows the statistics of the probability of customers using the Internet when they have accessibility to PCs.

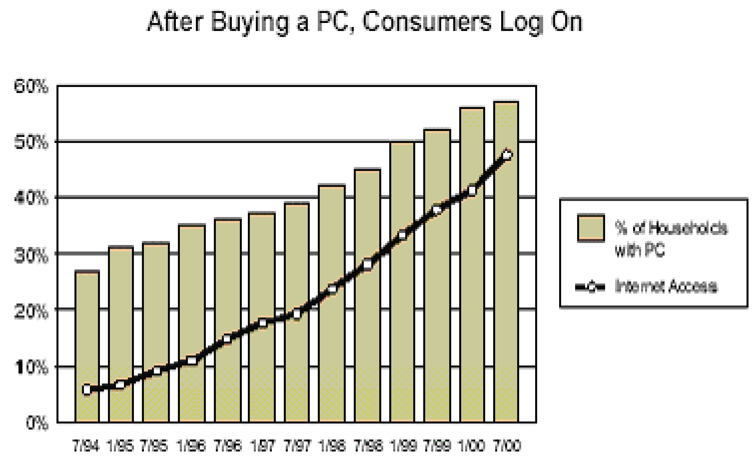
As the utility function of benefits moves toward the Internet, insurers must then develop products that serve new and existing customers, such as e-quote products that assist brokers and other intermediaries. Further, they also must follow the changing needs of consumers whose response to defined contribution health products is to date largely untested.

In summary then, the effective addressing of the key e-health issues of procurement, connectivity, and benefits requires standardization, a pivotal prerequisite for the implementation of successful e-health initiatives. Without such standardization the exchange of documents and other procurement information, connectivity, and e-commerce enabled benefits clearly become more problematic. Unfortunately, standardization is woefully lacking in too many areas of healthcare, let alone e-health. Given the opportunities for e-health to benefit various aspects of healthcare and the far reaching impact of any such e-health initiative, it becomes imperative then to have frameworks and models that not only bring to the forefront the key e-health issues but also provide guidelines for how to effectively bring to bear successful e-health initiatives for all healthcare organizations.

A FRAMEWORK FOR ASSESSING E-HEALTH POTENTIAL

We propose the framework shown in Figure 6 to assess the e-health potential and preparedness of countries. Healthcare polices are generally developed to a large extent at a macro, country level and thus we believe it is also necessary when

Figure 5. Internet usage among customers (Source: Dialog Cybercitizen Health Trend Report (Re. No.. 49), by Odyssey Research, 2003. Available online from www.odysseyresearch.org)



looking at e-health to first take a macro perspective and analyze the level of the country in terms of embracing e-health. The framework highlights the key elements that are required for successful e-health initiatives and therefore provides a tool that allows analysis beyond the quantifiable data into a systematic synthesis of the major impacts and prerequisites. The framework contains four main prerequisites, four main impacts, and the implications of these prerequisites and impacts to the goals of e-health. By examining both the prerequisites and the impacts, we can assess the potential of a country and its preparedness for e-health as well as its ability to maximize the goals of e-health. We discuss the prerequisites next.

PREREQUISITES FOR E-HEALTH

As can be seen in Figure 6, the four critical prerequisites for any successful e-health initiative include information communication technology (ICT) architecture/infrastructure, standardized policies, protocols and procedures, user access and accessibility policies and infrastructure, and, finally, government regulation and control. These will now be briefly discussed in turn.

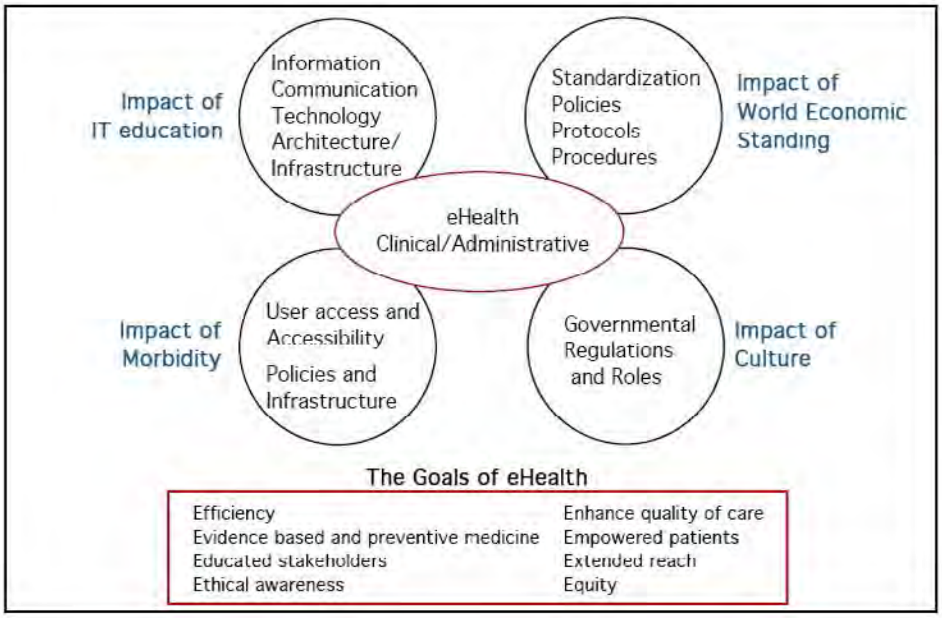
ICT Architecture/Infrastructure

The generic architecture for most e-health initiatives was depicted in Figure 1. To support such a client-server architecture special attention must be paid to the ICT infrastructure.

The ICT infrastructure includes phone lines, fiber trunks, and submarine cables, T1, T3 and OC, ISDN, DSL and other high-speed services used by businesses as well as satellites, earth stations, and teleports. A sound technical infrastructure is an essential ingredient to the undertaking of e-health initiatives by any nation. Such infrastructures should also include telecommunications, electricity, access to computers, number of Internet hosts, number of Internet Service Providers (ISPs), and available bandwidth and broadband access. To offer a good multimedia content and thus provide a rich e-health experience, one would require a high bandwidth. ICT considerations are undoubtedly one of the most fundamental infrastructure requirements.

Networks are now a critical component of the business strategies for organizations to compete globally. Having a fast microprocessor-based computer at home has no meaning unless you have high bandwidth-based communication in-

Figure 6. A framework for assessing a country's/region's e-health potential



frastructure available to connect computers with the ISP. With the explosion of the Internet and the advent of e-commerce, global networks need to be accessible, reliable, and fast to participate effectively in the global business environment. Telecommunications is a vital infrastructure for Internet access and hence for e-commerce. One of the pioneering countries in establishing a complete and robust e-health infrastructure is Singapore which is in the process of wiring every home, office and factory up to a broadband cable network which will cover 98% of Singaporean homes and offices (ibid).

Standardization Policies, Protocols and Procedures

E-health by definition spans many parties and geographic dimensions. To enable such a far-reaching coverage, significant amounts of document exchange and information flows must be accommodated. Standardization is the key for this. Once a country decides to undertake e-health initiatives, standardization polices, protocols and procedures must be developed at the outset to ensure the full realization of the goals of e-health. Fortunately, the main infrastructure of e-health is the Internet that imposes the most widely and universally accepted standard protocols such as TCP/IP and hypertext transfer protocol. It is the existence of these standard protocols that has led to the widespread adoption of the Internet for e-commerce applications.

The transformation to e-health by any country cannot be successfully attained without the deliberate establishment of standardization policies, protocols, and procedures that play a significant role in the adoption of e-health and the reduction of many structural impediments (Samiee, 1998).

User Access and Accessibility Policies and Infrastructure

Access to e-commerce is defined by the World Trade Organization (WTO) as consisting of two critical components: a) access to Internet services and b) access to e-services (Panagariya, 2000); the former deals with the user infrastructure, while the latter pertains to specific commitments to electronically accessible services. The user infrastructure includes number of Internet hosts and number of Web sites, Web users as a percent of the population as well as ISP availability and costs for consumers, PC penetration level, and so forth. Integral to user infrastructure is the diffusion rate of PCs and Internet usage. The United States and the United Kingdom have experienced the greatest penetration of home computers (Samiee, 1998). For developing countries such as India and China there is, however, a very low PC penetration and teledensity. In such a setting it is a considerable challenge then to offer e-health, since a large part of the population is not able to afford to join the e-commerce bandwagon. Countries thus have to balance local call charges, rentals, subscription charges, and so forth, otherwise the majority of citizens will find these costs a disincentive. This is particularly significant for developing and emerging nations where access prices tend to be out of reach for most of the population. Upcoming new technologies hold the promise to increase the connectivity as well as affordability level, and developing countries will need to seriously consider these technologies. In addition to access to PCs and the Internet, computer literacy is important and users must be familiar not only with the use of computers and pertinent software products but also the benefits and potential uses of the Internet and World Wide Web (ibid).

Governmental Regulation and Control

The key challenges regarding e-health use include a) cost effectiveness (i.e., less costly than traditional healthcare delivery), b) functionality and ease of use (i.e., they should enable and facilitate many uses for physicians and other healthcare users by combining various types and forms of data as well as be easy to use), and c) they must be secure. One of the most significant legislative regulations in the United States is the Health Insurance Portability and Accountability Act (HIPAA; Protegrity, 2001).

Given the nature of healthcare and the sensitivity of healthcare data and information, it is incumbent on governments not only to mandate regulations that will facilitate the exchange of healthcare documents between the various healthcare stakeholders but also to provide protection of privacy and the rights of patients. Some countries, such as China and Singapore, even control access to certain sites for moral, social, and political reasons while elsewhere transnational data flows are hindered by a plethora of regulations aimed at protecting domestic technology and related human resource markets (Goff, 1992; Gupta, 1992; Samiee, 1998). Irrespective of the type of healthcare system (i.e., whether 100% government driven, 100% private, or a combination thereof), it is clear that some governmental role is required to facilitate successful e-health initiatives.

The significance of the preceding four prerequisites on e-health initiatives will be modified by the impacts of IT education, morbidity, cultural-social dimensions, and world economic standing, as elaborated upon in the following sections.

Key Impact of E-Health

Figure six highlights four key impacts of e-health, which we now discuss in turn.

Impact of IT Education

A sophisticated, well-educated population boosts competition and hastens innovation. According to Michael Porter, one of the key factors to a country's strength in an industry is strong customer support (Porter, 1990). Thus, a strong domestic market leads to the growth of competition, which leads to innovation and the adoption of technology-enabled solutions to provide more effective and efficient services, such as e-health and telemedicine. As identified previously, the health consumer is the key driving force in pushing e-health initiatives. We conjecture that a more IT-educated healthcare consumer would then provide stronger impetus for e-health adoption.

Impact of Morbidity Rate

There is a direct relationship between health education and awareness and the overall health standing of a country. Therefore, a more health conscious society, which tends to coincide with a society that has a lower morbidity rate, is more likely to embrace e-health initiatives. Furthermore, higher morbidity rates tend to indicate the existence of more basic health needs (WHO, 2003) and hence treatment is more urgent than the practice of preventative medicine and thus e-health could be considered an unrealistic luxury and in some instances such as when a significant percentage of a population is suffering from malnutrition related diseases is even likely to be irrelevant at least in the short term. Thus, we conjecture that the modifying impact of morbidity rate is to prioritize the level of spending on e-health versus other basic healthcare needs.

Impact of Cultural-Social Dimensions

Healthcare has been shaped by each nation's own set of cultures, traditions, payment mechanisms, and patient expectations. While the adoption of e-health, to a great extent, dilutes this cultural impact, social and cultural dimensions will still be a moderating influence on any country's e-health initiatives. Another aspect of the cultural-social dimension relates to the presentation language of the content of the e-health repositories. The entire world does not speak English, so the e-health solutions have to be offered in many other languages. The e-health supporting content in Web servers and Web sites must be offered in local languages, supported by pictures and universal icons. This becomes a particularly important consideration when we look at the adoption and diffusion of evidence-based medicine, as it will mean that much of the available evidence and case study data will not be easily accessible globally due to language barriers.

Therefore, for successful e-health initiatives it is important to consider cultural dimensions. For instance, an international e-commerce study by International Data Corporation (Wilson, 1999) indicated that Web surfing and buying habits differ substantially from country to country and this would then have a direct impact on their comfort to use e-commerce generally and e-health in particular, specially as e-health addresses a more fundamental need. Hence, the adoption of e-health is directly related to ones' comfort with using the technology, and this in turn is greatly influenced by cultural dimensions. Also connected with cultural aspects is the relative entrepreneurial spirit of a country. For example, a study by Hofstede (1980) indicated that in a cultural context, Indians score high on "uncertainty avoidance" criteria when compared to their Western counterparts. As a result, Indians do not accept change very easily and are hostile towards innovation. This, then, would potentially pose a challenge to the starting up of e-health initiatives whose success depends on widespread adoption for their technological innovations. Thus, we conjecture that fear of risk and absence of an entrepreneurial mindset, as well as other cultural-social dimensions, can also impact the success of e-health initiatives in a given country.

Impact of World Economic Standing

Economies of the future will be built around the Internet. All governments are very aware of the importance and critical role that the Internet will play in a country's economy. This makes it critical that appropriate funding levels and budgetary allocations become a key component of governmental fiscal policies so that such initiatives will form the bridge between a traditional healthcare present and a promising e-health future. Thus, this would determine the success of effective e-health implementations and, consequently, have the potential to enhance a country's economy and future growth.

The World Economic Forum's global competitiveness ranking measures the relative global competitiveness of a country. This ranking takes into account factors such as physical infrastructure, bureaucracy, and corruption. Thus we conjecture that weak physical infrastructure combined with high levels of bureaucracy and corruption will lead to significant impediments to the establishment of successful e-health initiatives.

In developing its e-health initiative, a good first step for any country is to assess its standing with respect to the four prerequisites and four impacts discussed previously in this chapter. In this way it will be possible for a country to evaluate its preparedness with respect to these parameters and consequently devise appropriate policies and strategies for an effective and successful e-health initiative. In the following section, we will attempt to provide a guideline that will facilitate such an evaluation.

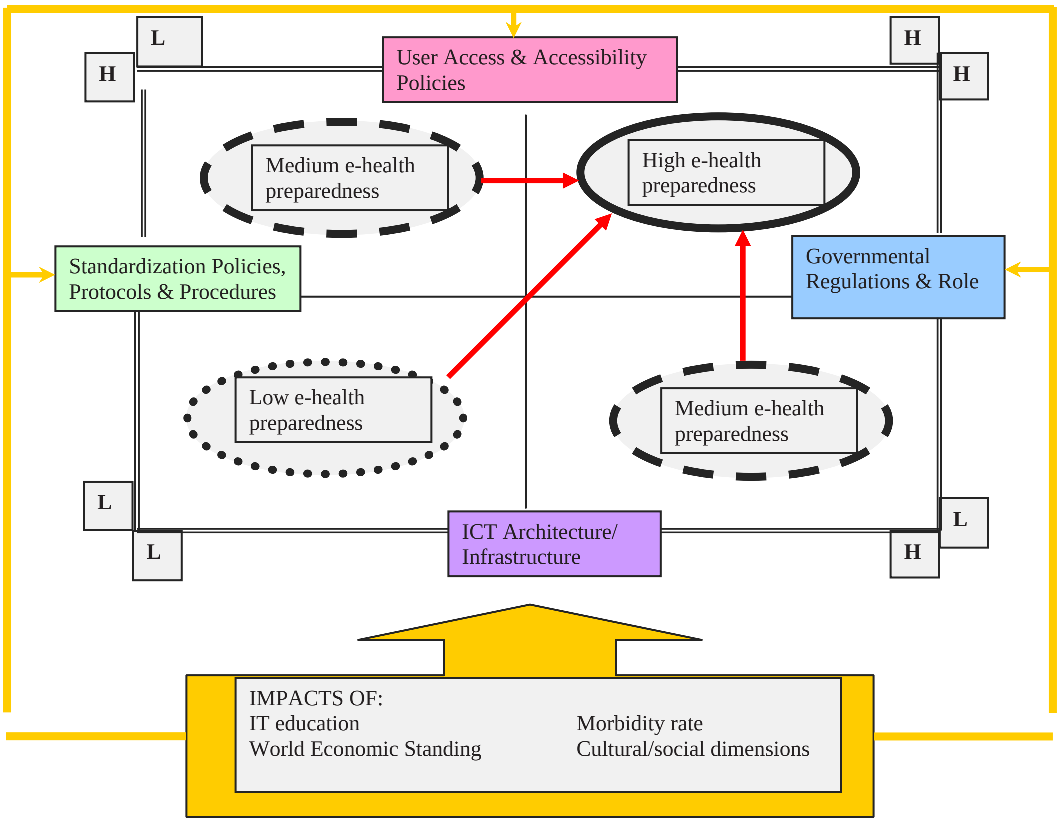
E-HEALTH PREPAREDNESS GRID

By taking the four main prerequisites as well as the four major impacts, identified in our framework in Figure 6 (i.e., information communication technology infrastructure; standardization policies, protocols and procedures; user access and accessibility policies; and infrastructures, governmental regulations and role) as well as the impact of IT education, the impact of morbidity rate, the impact of world economic standing, and the impact of cultural-social dimensions, we developed a grid for assessing e-health preparedness (see Figure 7) in which we can plot various countries with respect to these key parameters. The grid consists of four quadrants that represent the possible states of preparedness with respect to the key parameters for e-health success. The low preparedness quadrant identifies situations that are low with respect to all four prerequisites for e-health potential. The medium preparedness quadrant identifies two symmetric situations; namely, a combination of high and low positioning with respect to the four prerequisites for e-health potential. Finally, the high preparedness quadrant identifies situations that are high with respect to all four prerequisites for e-health potential. This grid not only shows the possible positioning of a given country with respect to its e-health preparedness (i.e., low, medium, or high) but also the path it must take and, more specifically, the prerequisite factors it must focus on to migrate to the ideal state of preparedness; namely, being high with respect to all four prerequisites. The grid also underscores the moderating role of the four impacts irrespective of the relative positioning on the state of preparedness of a given country.

DISCUSSION

From the e-health preparedness grid in Figure 7 we can see several implications. For countries that are low with respect to all four of the e-health prerequisites, much preparatory work is required to be e-health ready and thereby fully realize the goals of e-health (i.e., efficiency, evidence-based and preventive medicine, educated stakeholders, ethical awareness, enhanced quality care, empowered patients, extended reach and equity). For countries that map to the medium preparedness quadrant, more emphasis is needed on upgrading the deficiencies that cause them to score low on some prerequisites while they continue to maintain a high status on other prerequisites so that the full benefits of their e-health initiatives can be realized. Countries that are noted for being pioneers and leaders in the area of e-commerce in general as well as e-health in particular would be expected to map on the high quadrant. The challenge for these countries would be to maintain their high status with respect to all the

Figure 7. E-health preparedness grid



prerequisites. Furthermore, organizations within these countries can also begin to structure their respective e-health initiatives (i.e., micro-level issues) based on the key macro-level issues that both the integrative framework and e-health preparedness grid identify.

It is important to note that irrespective of their positioning on the preparedness grid (i.e., low, medium, or high), all countries must take into account the moderating effect of the four impacts on their e-health initiatives. If countries ignore the moderating effect of these four impacts, it will not be possible to fully realize a successful e-health initiative (and thus would not be able to fully realize the benefits of e-health), even if a country maps into the high-preparedness quadrant. Finally, we conjecture that the type of health system vis-à-vis the public–private continuum would not significantly impact the positioning of a country into a given quadrant. The successful experience of Singapore, albeit an isolated case, somewhat lends credence to our grid, as it would mostly fit the profile of a country in the high-preparedness quadrant and thus would be expected to successfully undertake an e-health initiative (PWC, 2003). Clearly, additional empirical validation of our e-health preparedness framework is required which we leave to future studies.

CONCLUSIONS

E-commerce, as noted by the U.N. Secretary General's address, is an important aspect of business in today's 21st century. No longer is it a luxury for nations; rather it is a strategic necessity in order for countries to achieve economic and business prosperity as well as social viability.

One of the major areas within e-commerce that has yet to reach its full potential is e-health. This is because healthcare generally has been slow in adopting information technologies. Furthermore, there is a shortage of robust frameworks that may be used as guidelines for assessing countries' e-health preparedness and identifying the key areas and deficiencies that need to be addressed in order for successful e-heath initiatives to ensue. In addition, e-health is more than a technological initiative; it also requires a major paradigm shift in healthcare delivery, practice, and thinking. We have attempted to address this gap by developing a framework that identifies the major factors involved in assessing the e-health preparedness of countries and thereby, facilitating countries to focus their efforts on the relevant issues that must be addressed in order that successful e-health initiatives follow (i.e., the goals of e-health are realized). An outcome from our analysis indicates that the relative healthcare system (i.e., whether government driven, public, or two tier) would appear to have less significance in establishing successful e-health initiatives. The first step in the development of any viable e-health strategy is to make an assessment of the current state of e-health preparedness and then how to either move to a state of higher preparedness (i.e., the high quadrant) or focus on maintaining a current high quadrant status–both of these are identified through the use of our framework and thus its value. Finally, we note that with respect to our framework other parameters also exist and could also be considered important, perhaps even as important as the ones we used. However, we believe that the framework will still function the same way (i.e., provide a useful tool for any country trying to determine and develop a successful e-health initiative), irrespective of the number of parameters; in this regard the preference should be simplicity over complexity.

REFERENCES

Al-Qirim, N. (2004). Telemedicine: Building knowledge-based tele-health capability in New Zealand. In N. Wickramasinghe, J. Gupta & S. Sharma (Eds.), *Creating knowledge-based healthcare organizations.* Hershey, PA: Idea Group Publishing (forthcoming).

Alvarez, R. C. (2002). The promise of e-Health: A Canadian perspective. *eHealth International, 1*(1), 4.

Cyber Dialogue. (2000, May 23). *Online health information seekers growing twice as fast as online population.* Cyber Dialogue.

Cyber Dialogue. (2001). *The higher the connection speed, the higher the value: Broadband use as an indication of value amongst e-health users.* Cyber.

Eysenbach, G. (2001). *Journal of Medical Internet Research, 3*(2), e20.

Feeny, D. (2001, Winter). Making business sense of the e-opportunity. *MIT Sloan Management Review, 42*(2), 42.

Gargeya, V. & Sorrell. D. (2004). Moving toward an e-hospital. In N. Wickramasinghe, J. Gupta & S. Sharma (Eds.), *Creating knowledge-based healthcare organizations* (pp. 50-64). Hershey, PA: Idea Group Publishing.

Geisler, E. (1999). Mapping the knowledge base of management of medical technology. *International Journal of Healthcare Technology and Management, 1*(1), 3-10

Geisler, E. (2000). *The metrics of science and technology.* Westport, CT: Greenwood.

Geisler, E. (2001). Organizing for e-Business: The implementation of management principles in electronic commerce. *Journal of e-Commerce and Psychology.*

Glaser, J. (2002). *The strategic application of information technology in health care organizations* (2nd ed.). San Francisco: Jossey-Bass.

Goff, L. (1992). Patchwork of laws slows EC data flow. *Computerworld, 26*(15), 80.

Gupta, U. (1992). Global networks: Promises and challenges. *Information Systems Management 9*(4), 28-32.

Health Technology Center. (2000). A survey conducted for the Health Technology Center (HealthTech) by Harris Interactive in cooperation with Pricewaterhouse Coopers and the Institute for the Future (IFTF). Retrieved from *http://www.ncddr.org/cgi-bin/good-bye.cgi?url=http://www.healthtechcenter.org*

Healthcare Advisory Board. (2002, January). *Use of hospital Web sites to engender community loyalty.*

Hofer, C. (1975). Toward a contingency theory of business strategy. *Academy of Management, 18*(4), 784-810.

Hofstede, G. (1980). *Culture's consequences, international work related values.* Beverly Hills, CA: Sage.

Institute of Medicine. (2001). *Crossing the quality chasm: A new health system for the 21st century.* Washington, DC: Committee on Quality of Health Care in America Institute of Medicine, National Academy Press.

JMIR. Retrieved from *http://www.jmir.org*

Odyssey Research. Dialogue cybercitizen health trend report (Yr. 2001, No. 49). Retrieved from *www.odysseyresearch.org*

Panagariya, A. (2000). E-commerce, WTO and developing countries. *The World Economy, 23*(8), 959-978.

Porter, M. (1990). The competitive advantage of nations. New York: Free Press.

Porter, M., & Teisberg, E. (2004). Redefining competition in healthcare. *Harvard Business Review,* pp. 65-76.

Porter, M. (1980). *Competitive strategy.* New York: Free Press.

Porter, M. (1985). *Competitive advantage.* New York: Free Press.

Porter, M. (1990). *The competitive advantage of nations.* New York: Free Press.

Protegrity. (2001, May). Health Insurance Portability and Accountability Act (HIPPA) privacy compliance executive summary. Author.

PWC. (2003). PricewaterhouseCoopers healthcare practice. Retrieved from *www.pwchealth.com*

Samiee, S. (1998). The Internet and international marketing: Is there a fit? *Journal of Interactive Marketing, 12*(4), 5-20

Sharma, S. & Wickramasinghe, N. (2004). e-Health with knowledge management: Areas for tomorrow. In N. Wickramasinghe, J. Gupta & S. Sharma (Eds.), *Creating knowledge-based healthcare organizations* (pp. 110-124). Hershey, PA: Idea Group Publishing.

Sieving PC Factors Driving the Increase in Medical Information on the Web: One American Perspective. *J. Med. Internet Res.,* 1. Retrieved from *www.jmir.org/1999/1/e3*

Umhoff, B., & Winn, J. (1999, May 1). The healthcare profit pool: Who stands to gain and lose in the digital economy. *Health Forum Journal.*

UNCTAD. (2002). United Nations Conference on Trade and Development. Retrieved from *http://r0.unctad.org/ecommerce/ecommerce_en*

Wickramasinghe, N. (2000). IS/IT as a tool to achieve goal alignment: A theoretical framework. *International Journal Healthcare Technology Management, 2*(1-4), 163-180.

Wickramasinghe, N., et al. (2004). A framework for assessing e-health preparedness. *International Journal of E-Health*.

Wilson, T. (1999). Not a global village after all? Consumer behavior varies widely by country. *Internetweek*, *792*, 13.

World Health Organization. (2003). Retrieved from *http://www.emro.who.int/ehealth/*

ENDNOTE

[1] An earlier version of this chapter is to appear in a volume of the *Internation J. e-Health*, 2005.

This work was previously published in E-Health Systems Diffusion and Use: The Innovation, the User and the USE IT Model, edited by T.A.M. Spil and R.W. Schuring, pp. 294-323, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 1.3
The Competitive Forces Facing E-Health

Nilmini Wickramasinghe
Stuart Graduate School of Business, USA

Santosh Misra
Cleveland State University, USA

Arnold Jenkins
Johns Hopkins Hospital, USA

Douglas R. Vogel
City University of Hong Kong, China

ABSTRACT

Superior access, quality, and value of healthcare services has become a national priority for healthcare to combat the exponentially increasing costs of healthcare expenditure. E-Health in its many forms and possibilities appears to offer a panacea for facilitating the necessary transformation for healthcare. While a plethora of e-health initiatives keep mushrooming both nationally and globally, there exists to date no unified system to evaluate these respective initiatives and assess their relative strengths and deficiencies in realizing superior access, quality and value of healthcare services. Our research serves to address this void. This is done by focusing on the following three key components: 1) understanding the Web of players (regulators, payers, providers, healthcare organizations, suppliers, and last but not least patients) and how e-health can modify the interactions between these players as well as create added value healthcare services, 2) understand the competitive forces facing e-health organizations and the role of the Internet in modifying these forces, and 3) from analyzing the Web of players combined with the competitive forces for e-health organizations we develop a framework that serves to identify the key forces facing an e-health and suggestions of how such an organization can structure itself to be e-health prepared.

INTRODUCTION

E-health is a broad term that encompasses many different activities related to the use of the Internet for the delivery of healthcare service. Healthcare professionals are extending the use of the Internet to include a source of evidence-based consumer information as well as to facilitate the research of protocols for healthcare delivery, accessing laboratory and medical records, and performing second opinion consults (Sharma & Wickramasinghe, 2005; Sharma, Wickramasingeh, Xu, & Ahmed, 2006). Moreover, the Internet is being used by patients to become more knowledgeable about health practices as seen from their questions to their physicians (Gargeya & Sorrell, 2005).

Although, a relatively new term and unheard of prior to 1999, e-health has now become the latest "e-buzzword" used to characterize not only "Internet medicine," but also virtually everything related to computers and medicine (Sharma et al., 2006; Von Lubitz & Wickramasinghe, 2006). The scope and boundary of e-health, as well as e-heath organizations, is still evolving. However one can only imagine it will grow rapidly especially given that governments in both U.S. and Europe, and organizations such as WHO (World Healthcare Organization) are advocating that e-health be on the top of all healthcare agendas and an integral component of any healthcare delivery initiative (Von Lubitz et al., 2006).

Given the growth and variety of e-health initiatives, it becomes important to examine the forces affecting these initiatives and factors leading to the success of e-health. To date, little research examines metrics of measurement pertaining to e-health initiatives or their economic value. What are the forces of competition affecting e-heath? Are the competitive forces constrained by external considerations? Is the issue of competition an appropriate concern for e-health? If so, what are the strong and weak competitive forces? We argue that analysis of these forces would lead us to understand the long-term sustainability of any e-health initiative.

TRADITIONAL COMPETITIVE FORCES

The starting point for understanding the competitive forces facing any e-health initiative lies in understanding the fundamentals of traditional competitive forces that impact all industries and then how the Internet as a disruptive technology has impacted these forces.

The strategy of an organization has two major components (Hendersen & Venkatraman, 1993). These are 1) formulation--making decisions regarding the mission, goals, and objectives of the organization and 2) implementation--making decisions regarding how the organization can structure itself to realize its goal and carryout specific activites. For today's healthcare organizations the goals, mission, and objectives all focus around access, quality, and value and realizing this value proposition for healthcare then becomes the key (Wickramasinghe, Fadlalla, Geisler, & Schaffer, 2005). Essentially, the goal of strategic management is to find a "fit" between the organization and its environment that maximizes its performance (Hofer, 1975). This then describes the market-based view of the firm and has been predominantly developed and pushed by the frameworks of Michael Porter. The first of Porter's famous frameworks is the generic strategies (Porter, 1980).

The use of technology must always enable or enhance the businesses objectives and strategies of the organization. This is particularly true for 21st Century organizations where many of their key operations and functions are so heavily reliant on technology and the demand for information and knowledge is so critical. A firms' relative competitive position (i.e., its ability to perform above or below the industry average is determined by its competitive advantage). Porter (1980) identified three generic strategies that impact a firm's competitive advantage. These include cost, focus, and differentiation. Furthermore, Porter himself notes that two and only two basic forms of competitive advantage typically exist:

1. Cost leadership.
2. Differentiation.

Firms can use these two forms of competitive advantage to either compete across a broad scope of an industry or to focus on competing in specific niches; thereby, leading to three generic strategies. Porter (ibid) notes that firms should be cautious about pursuing more than one generic strategy; namely cost, differentiation, and focus. For example, if a cost leadership strategy is adopted it is unlikely that a firm can also maintain and sustain differentiation since it would not be possible to simultaneously pursue the costly capital investment or maintain high operating costs required for differentiation and thus in the long run the firm has a confused strategy which leads to failure.

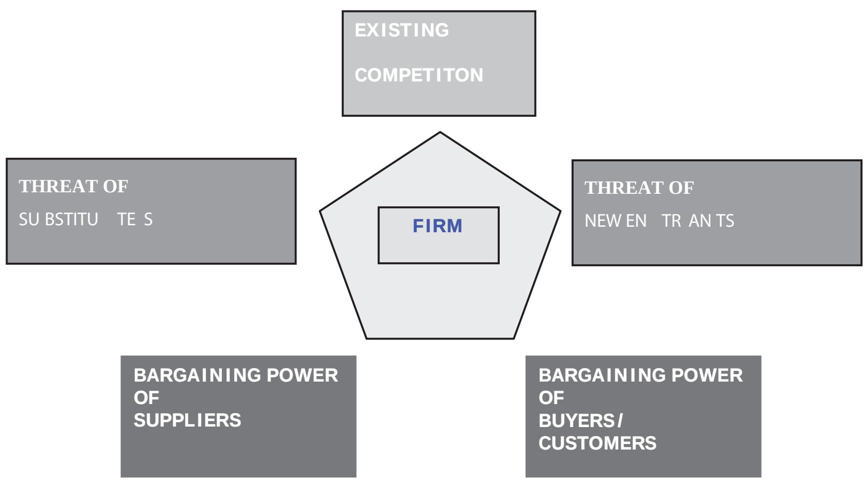
In order to design and develop ones strategy, an organization should first perform an industry analysis. Porters Five Forces or Competitive Forces model is most useful (Porter, 1980, 1985). Figure 1 depicts this model. Essentially, Porter has taken concepts from micro-economics and modeled them in terms of five key forces that together outline the rules of competition and attractiveness of the industry.

The forces are as follows:

1. **Threat of new entrant:** A company new to the industry that could take away market share from the incumbent firms.
2. **Threat of substitute:** An alternative means that could take market share from product/service offered by the firms in the industry.
3. **Bargaining power of buyers:** The strength of buyers or groups of buyers within the industry relative to the firms.
4. **Bargaining power of suppliers:** The strength of suppliers relative to the firms in the industry.
5. **Rivalry of existing competition:** Relative position and market share of major competitors.

The collective strength of these five forces determines the attractiveness of the industry and thus the potential for superior financial performance by influencing prices, costs, and the level of capital investment required (Porter, 1985). Once a thorough industry analysis has occurred, it is generally easier for a firm to determine which generic strategy makes most sense for it to pursue

Figure 1. Porter's Competitive (Five) Forces Model



and enables the firm to exploit most of its core competencies in its existing environment.

Role of the Internet of the Competitive Forces

Feeny (2001) presents a framework that highlights the strategic opportunities afforded to organizations by using the Internet. In particular, he highlights three e-opportunity domains. Table 1 details these domain and their respective components.

E-Opportunities in Healthcare

Given the three areas of e-opportunities previously discussed, Glaser (2002) identifies several key e-opportunities for healthcare. Table 2 details these.

WEB OF PLAYERS IN HEALTHCARE

Figure 2 depicts the Web of healthcare players and the key elements of the any e-health architecture that serves to support the interactions between and within this Web of players. In order to fully capture the flows of information it is necessary to first identify the primary producers and consumers of data and information within the healthcare system. At the center of the information flows is the HCIS (healthcare information system, i.e., the e-health network) because not only does it connect the key players within the healthcare system in an efficient and effective manner, but also it forms the central repository for key information such as patient medical records, billing, and treatment details. Hence, the HCIS provides the foundation for supporting the information flows and decision making throughout the healthcare system. Figure 2 then represents a macro view of the inter-relationships between the key players within this system as well as the sources, destinations and flows of information between these players and the pivotal role of the HCIS.

Healthcare procedures such as medical diagnostics, treatment decisions, and consequent effecting of these decisions, prevention, communication, and equipment usage can be thought of as iatric in nature (Wickramasinghe & Fadlalla, 2004). Integral to these iatric procedures is the generating and processing of information (Wickramasinghe & Fadlalla, 2004). The patient naturally provides key information at the time of a clinical visit or other interaction with his/her provider. Such a visit also generates other information including insurance information, medical history, and treatment protocols (if applicable) which must satisfy regulatory requirements, payer directives and, obviously, the healthcare organization's informational needs. Thus, we see that from a single intervention many forms and types of information are captured, generated, and then disseminated throughout the healthcare system. All this information and its flows must satisfy some common integrity characteristics such as accuracy, consistency, reliability, completeness, usefulness, usability, and manipulability. Consequently, generating a level of trust and confidence in the information's content and processes. Since the information flows across various organizational boundaries, the challenge of ensuring information integrity is further compounded because any integrity problems will propagate with ripple effects following the same trajectory as the information itself. Given the high degree of inter-relatedness between the various players, the consequences of poor quality information (such as the cost of information integrity problems) are multiplied and far reaching. This highlights the need for robust, well designed, and well managed HCIS (Wickramasinghe & Fadlalla, 2004). Such a perspective should not be limited to new systems, but rather, equally and perhaps of even more importance should be applied to existing systems as well.

Table 1. The three e-opportunity domains and their components

	Components
e-operations	• Automation of administrative processes • Supply-chain reconfiguration • Reengineering of primary infrastructure • Intensive competitive procurement • Increased parenting value
e-marketing	• Enhanced selling process • Enhance customer usage experience • Enhanced customer buying experience
e-services	• Understanding of customer needs • Provision of customer service • Knowledge of all relevant providers • Negotiation of customer requirements • Construction of customer options

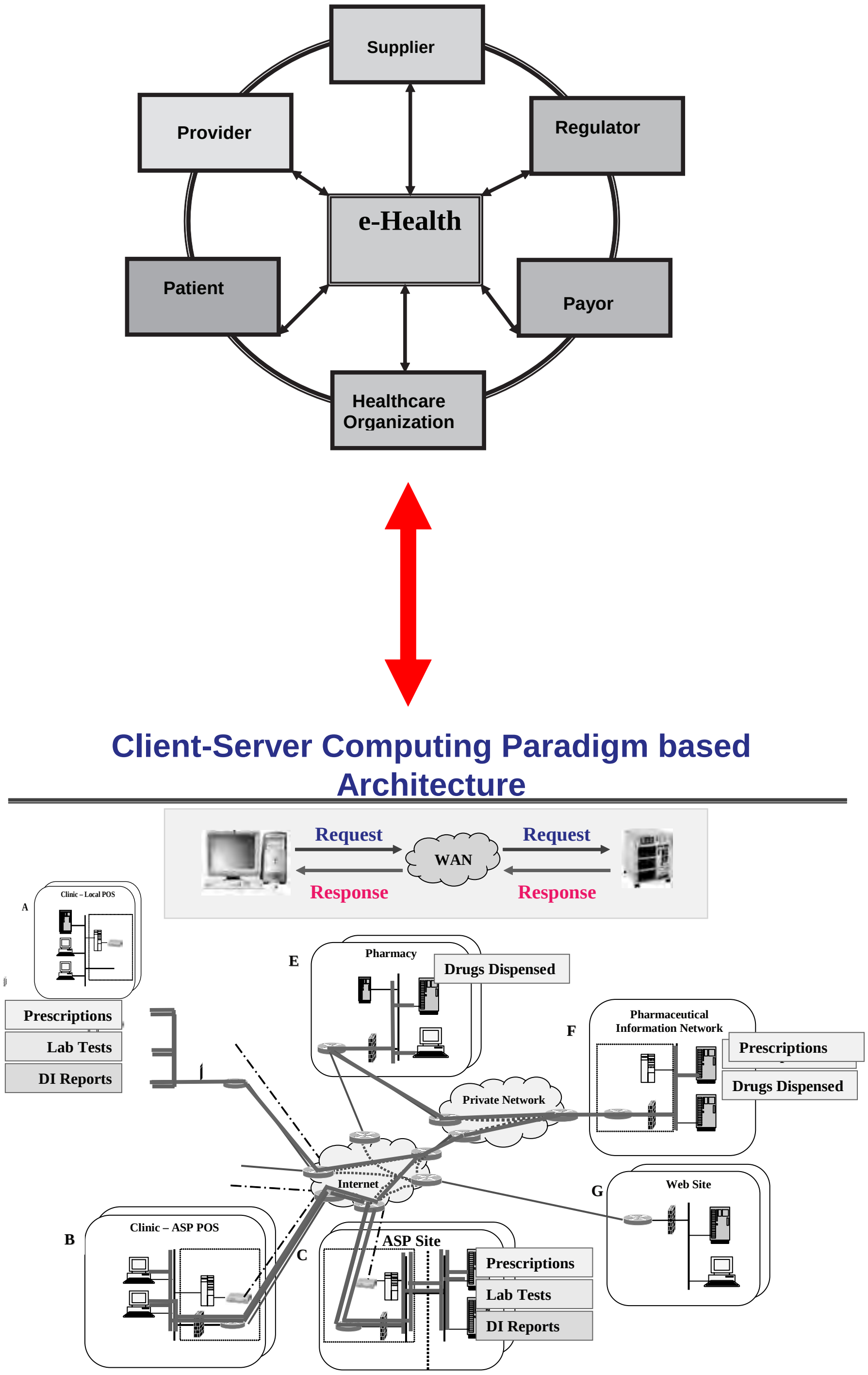
Table 2. The e-opportunities for healthcare organizations

	Components
e-operations	• Internet-based supply purchasing • Prescription writing, formulary checking, and interaction checking using hand-held devices
e-marketing	• Delivery of consumer health content and wellness management tools over the Internet • Use of consumer health profiles to suggest disease management and wellness programs
e-services	• Patient-provider communication and transaction applications • Web-based applications to support the clinical conversation between referring and consulting physicians
Crossing multiple domains	• Increasing the level of information content in the product • Increasing the information intensity along the supply chain • Increase in the dispersion of information

MODELING THE COMPETITIVE FORCES IN E-HEALTH

In order to model e-health, let us first construct a general model of the competitive forces pertaining to e-business. E-business is not simply offering traditional products and services on line. It requires broad-scale asset redeployment and process changes, which ultimately serve as the basis for a company's competitive advantage in today's Digital Economy. For this study, the e-business model could be broken into components such as; products and services, customer value, pricing component, revenue source, the cost component, and asset model as shown in Figure 3.

Figure 2. Web of e-health players adapted from Wickramasinghe et al 2004



The prime objective of business model is to make money (La Monica, 2000). The various components of business model as shown in Figure 1 work together to create profit margins for the business. First of all, the electronic business model should offer products and services online. These products and services should be differentiated with competitors by low price or unique customer value. The products are differentiated if customers perceive some value in these that other products do not have. Differentiation can be done by offering different product features, timing, location, service, product mix, linkage between functions, etc. (Afuah & Tucci, 2000). Customer value can be judged whether firm offering its customers something distinctive or at a lower cost than its competitors. The success of business model depends upon how does the firm price the value? An important part of profiting from the value that firms offer customers is to price it properly. For pricing, market shares and margins would be most critical. The good business model should strive for high market share and thus firm should devise strategies accordingly. Pricing of products depends upon the cost and asset model of the firm. The cost (fixed cost + variable cost) should be spread in a fashion that profit margins remain high. The profits in electronic business model case will not only come from sales but may come from many other sources. Therefore, revenue source is another important component for business model. The sustainability of business model can be gauged based upon non-imitable nature of products and services. How can firm continue improve market share and make more money and have competitive advantage are the kind of questions needs to answers for the sustainability of business model. For example; using simple profits equation; Profits=(P-Vc)Q-Fc , firm can assess how each of the components of business model impact profitability. If a firm offers distinctive products, it can charge premium price P for it. A good business model should keep low variable cost but should have high market share for higher profitability (Afuah et al., 2000). Taking these components of a business model into consideration, let us now map this to the healthcare domain (Figure 4).

In so doing, some of the nuances pertaining to the dynamics of healthcare become apparent; such as, the receiver of services, or the patient, is not usually the principal payer. Moreover, the model serves to underscore that for e-health initiatives to truly add value and be sustainable the dynamics

Figure 3. Generic e-business model components

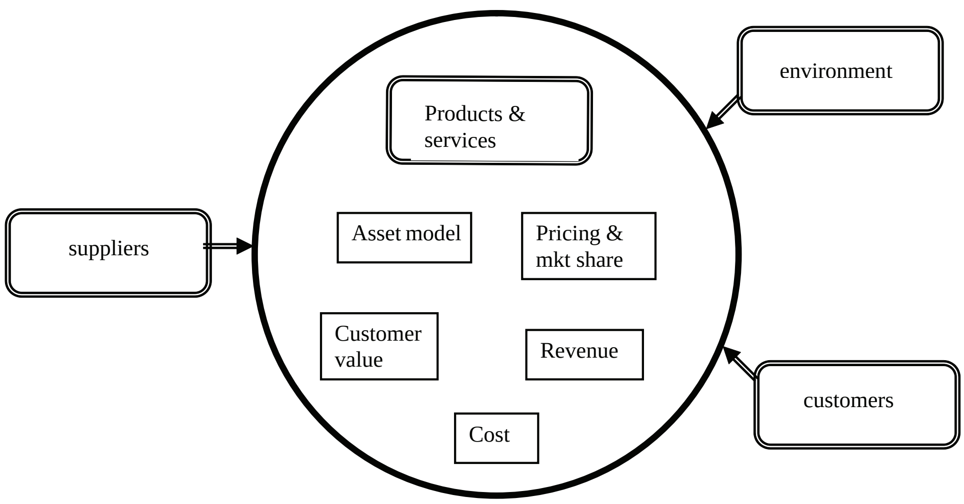
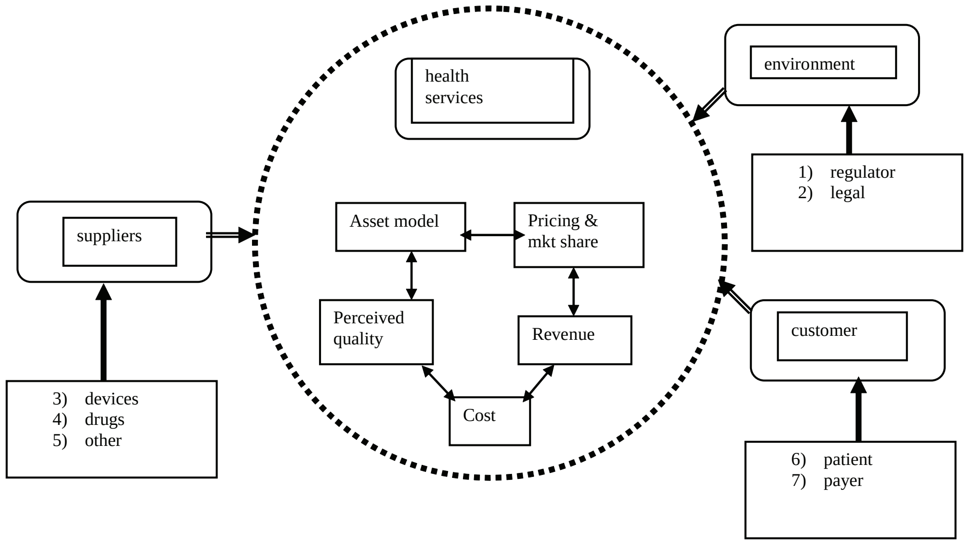


Figure 4. e-health business model components



of a generic e-business model must be satisfied. Hence, some determination needs to be made regarding Vc, Fc, P, and Q in this context.

To understand these dynamics more easily let us consider a case study example of the implementation of an electronic patient record.

Case Study

The Johns Hopkins Medicine Center for Information Services Public LAN (JPL) is a computer network designed to provide patient care providers access to clinical applications. This computer network is utilized by all types of patient care providers in both inpatient and out patient services. These providers include, but not limited to, doctors, interns, fellows, nurses, unit clerks, pharmacists, nutritionists, and admission specialists. In this article, an examination of the history of the Public LAN, the current state of the LAN, and the future of the Public LAN will be examined. Since its inception the Public LAN as been the leader in efficiency and innovation for Desktop Computing Services (DCS), a division of Information Technology @ Johns Hopkins (IT@Hopkins).

Introduction of the Public LAN

During the spring of 1996, JHMCIS and a group of doctors developed an in-house application to provide patient care. This application is called Electronic Patient Record or EPR. The application was to be used in patient areas for tracking patient record. These records can then be viewed by other clinicians throughout the hospital. A second application was introduced at the same time to provide a graphical user interface to many of the hospital's mainframe and mid-range systems. This application is Host Interface Program or HIP. The challenge at this stage was to provide a computer system that could be used by the doctors that would allow EPR and HIP to be used to provide patient care and at the same time have a desktop system that was secure.

These desktops were to be deployed in medical exam rooms and the major problem was having a desktop that could provide these applications to the clinicians without allowing the clinicians or patients the ability to access the operating system and the computer configuration. This led to the development of the Public Desktop.

The Public desktop is a Microsoft Windows based desktop that has the clinical applications installed, as well as an Internet browser and the Microsoft Office suite. The challenge was managing these systems in areas with limited access during business hours as they were in use by clinicians providing patient care. The operating system was secured and limited access was given to the users. The users were not able to install any applications or download any programs.

The Public LAN started out with 70 desktops in three clinical areas. The Harriet Lane Clinic, which is an outpatient clinic for pediatrics, the neonatal intensive care unit, and the adolescent outpatient clinic. This pilot lasted approximately six months. During the next three years, the Public LAN grew to 1100 desktops.

The Growth of the Public LAN

Today the Public LAN supports over 1800 desktops and many clinical applications. During the first three years of the Public LAN, the number of systems reached over 1100 systems. Included in this growth, not only the number of devices supported, but the number of applications that were supported on these desktops. The driving forces of these changes were outdated clinical applications that were being replaced with client server applications and the millennium with applications that were not year 2000 compliant.

During this time, the application supported grew to include BDM, a new pharmacy application, Vision—a nutrition application and ClicTate, a pediatrics version of EPR. With the intention of more clinical applications moving from the mainframe and mid-range systems to client server applications, the desktops are going to need to be able to handle these additional applications.

The process of managing these systems became a challenge as well. Since the desktops were standardized, DCS was able to implement Microsoft's System Management Server (SMS). This allowed not only the ability to manage these desktops, but also distribute software, inventory the hardware, and software of a specific system, and provide remote control capabilities. SMS was included when the pilot of the Public LAN was deployed but its true value was not realized until the rapid growth of the LAN.

The Public LAN Today

The Public LAN today is well over 1800 desktops, supporting more than 30 clinical applications. Most of these applications are still accessed via HIP, however more client server applications are also supported. The additional client server applications have lead to different configurations of the desktop's application software or "flavors" of Public workstations. Currently there are currently many different configurations for the Public Workstations. These different configurations include:

- Standard configuration.
- Training configuration.
- Wilmer Eye Clinic configuration.
- Pharmacy configuration.
- Nutrition configuration.
- Provider Order Entry configuration.
- Operating room configuration.
- DCOM image viewing configuration.
- Eclypsis Point of Care configuration.
- Procedure Reporting System configuration.
- OB/GYN Configuration.

These different configurations can be on a few as 20 desktops to as many as 600, where the standard configuration is on all of the desktops. The standard configuration is:

- Windows XP Professional.
- EPR.
- HIP.
- Internet Explorer.
- Microsoft Office Suite.
- Adobe Reader.
- Calculator.

The additional configurations are based on adding additional clinical applications to the desktops. In addition, many of the systems have multiple clinical applications installed.

The Lessons Learned

During the growth of the Public LAN, many lessons have been learned. These lessons include best practices for desktops management, application management and deployment, and reduction in the total cost of ownership of a desktop.

The current network is supported by three desktop technicians, which is an average of 600 plus desktops per technician. Desktop Computing Services needed to have a way to manage these systems not only located at the East Baltimore campus, but at other campuses within the Baltimore metropolitan area. The use of Microsoft Systems Management Software (SMS) was deployed to allow desktop management. SMS allows a technician the ability to remote control in to a desktop and perform work as if they were at the desktop. This capability also allows the support staff to view the process of the user and see the error as it happened. SMS also is used to deploy application software to the desktop.

Due to the increased number of clinical applications, the number of different application software configurations increased. In order to manage this DCS used SMS for application deployment. DCS is able to determine the application software installed on the desktop and perform upgrades to the software. The upgrade to an application is preformed by using SMS to "push" and install the software on the desktop without any user intervention. Therefore, and application could be upgraded or installed without having to visit the desktop.

With the integration of SMS to manage the desktops, this has reduced the total cost of ownership of supporting the Public LAN. This decrease is realized by having a ratio of one desktop technician per 600 desktops. DCS is able to remote control the desktop; this prevents the technician from have in walk across campus to help a user. In addition, the installation of applications and upgrades to applications is completed on many systems at once without having to visit each individual desktop. Also, DCS has secured the desktop to prevent the users from accessing the operating system and the hard drive. If the users were able to access the operating system and download and install applications, including spyware, this would greatly increase the support costs of the desktop.

The Future of the Public LAN

The future of the Public LAN at Johns Hopkins Hospital is ever evolving. The needs of the clinicians for resources to provide patient care are continually changing. With patients bringing medical records in on CD-ROM to access to network resources the Public LAN must evolve to meet these needs. In order to meet these needs the Public LAN support staff is required to find clever and innovative ways to provide these resources. New hardware is being added to the Public desktops to allow viewing of clinical data on CD-ROM, the use of USB keys for file storage has been enabled and logging in with a personal account.

The ability for a clinician to login with their personal account allows them to access network resources. These resources include access to network file servers and departmental file servers. In order for a clinician to use a personal account they are required to have a timeout of their session. The timeout of the session will log the user off after a certain amount of idle time. The reason for this is to prevent others from accessing information and to prevent non Johns Hopkins employees to access data and network resources.

The future of the Public LAN is ever evolving. The Johns Hopkins Hospital is building two new clinical towers that will be state of the art. The devices that provide patient care will also need to be state of the art and provide clinicians the ability to provide patient care in a completely paperless, film-less, and wireless network. The Public LAN will be able to provide these solutions and will realize the benefits of these efforts, as patients are cared for more efficiently and effectively.

Mapping the Case to the Model

The implementation of the EPR at Johns' Hopkins represents a relatively common e-health initiative in the current healthcare environment. The EPR enables the seamless flow of patient data and thus facilitates the delivery of efficient and effective quality healthcare to the patient. This is certainly professed as a key benefit for the embracing of EPR in most instances.

The e-health sustainability model however, suggests that one must analyze the micro- and meso-dynamics more closely to actually determine the sustainability of such an initiative. Specifically, it is necessary to capture key factors including, perceived quality, fixed and variable costs, price and market share and quantity and then look at the interaction of these factors before sustainability of the initiative can be pronounced. However, this is beyond the scope of this article but will form the focus of future research.

What can be noted at this point and will be research in more detail in future work is the size or scale of the e-health initiative. Returning to the simple profit equation Profits=(P-Vc)Q-Fc, in the case scenario above, fixed costs will be constant and Vc for any EPR will be marginal given the generic nature of the program and the applications of it by various providers hence we hypothesis that the sustainability of the EPR would increase with Q the quantity or size. Thus, the larger the EPR initiative the more likely it is to be sustainable. Quantitative data to support the relationship between scope and quantity and impact of ICTs in general in healthcare settings can be found in previous studies (Wickramasinghe & Lamb, 2002; Wickramasinghe & Silvers, 2003).

DISCUSSION

In mapping the John's Hopkins case to the model in Figure 4, we can see that the reality of an e-health initiative involves the interactions of various groups of stakeholders. Knowledge management provides an umbrella under which we may discuss a number of opportunities and raise issues relative to components of the business model. The vision of collaboration between components of the business model recognized as stakeholders is one of great opportunity. Stakeholders in this case include suppliers, the firm, the customer, and the government as a key representative of the environment. Each stakeholder brings to the table talent, resources, and differentiated perspectives that, together, create a robust whole in addressing problems and projects. For example, suppliers can be a source of knowledge that can assist the firm in delivering cost effective products and services. Customers are an additional source of knowledge in terms of personal history and preferences. The firm can manage knowledge in a form that maximizes the probability of value added products and services. The government can serve as a catalyst to create an environment conducive to knowledge exchange and management.

Unfortunately, great opportunities do not always turn into reality. Collaboration successes between suppliers, the firm, and its customers much less the government can, sadly, be few and far between. In addition to strengths and distinctions, each stakeholder also brings to the table residual weaknesses and biases that can scuttle the best of collaborative intentions. For example, internal firm bureaucracy can easily drive out the best of suppler intentions and customer good will. Problems can easily be left unaddressed and efforts can easily fail as reality drives out vision. This can be exacerbated by cultural norms and historical behaviors embedded in government policies.

A case in point is the handling of SARS. Levels of suffering and unnecessary deaths were, in part, a result of lack of collaboration between stakeholders. In this case, government agencies (specifically the hospital authorities) were negligent in sharing information and allocating resources amongst hospitals. The hospitals, however, were not guilt free and were accused of withholding information to customers including patients and their families. Further, the relationship between suppliers and hospitals was insufficient to respond to the need for supplies. Shortages were evident and supplies misapplied in circumstances that could have been adverted through collaboration. The situation was further strained as lack of information sharing across governments and excessive bureaucratic delay inhibited quick action to rapidly respond to changing circumstances. In summary, stakeholder collaboration could have, arguably, avoided hardship at individual and societal levels. Unfortunately, it didn't happen and the World Health Organization (WHO) was, rightly, exasperated.

Experiences with SARS have sensitized stakeholders at all levels with respect to effectively dealing with potential pandemics e.g., H5N1-based bird flu. Over the past months, we have already seen a much higher level of information exchange and collaboration than existed in the lead-up to SARS. Governments have more readily shared information and established channels for dealing with global adversity. Hospitals have begun preparations including emergency response practice. Suppliers have opened historically propriety processes and licenses to enable extended manufacturing capability (e.g., Roche with Tamiflu, as but one example). Customers have sought (and obtained) information relative to prevention and preparation for a variety of circumstances as well as acted as a source of information back to appropriate authorities regarding infectious incidences, e.g., bird flock deaths. Numerous conferences with multiple stakeholders present have provided forums for knowledge sharing, enhanced understanding leading towards the creation of action plans. In short, bird flu threats have galvanized stakeholders in a way that was unseen in the handling of SARS, in part, as a result of witnessing and experiencing hardship.

Knowledge management provides a focus that can enhance the probability of success in encouraging and sustaining broad-based stakeholder collaboration. Formalized knowledge management promotes the ultimate desire for the benefits of stakeholder collaboration to be sufficiently well developed and supported to offset inherent weaknesses. Knowledge management plays a key role in assuring that aspects of information creation, sharing, and dissemination compatible with multiple stakeholder objectives can be successfully achieved (Alavi & Leidner, 2001). Problems are often beyond the scope of any particular stakeholder, which encourages cooperation in order for success to be attained (Van de Ven, Angle, & Poole, 2000).

The concept of suppler, firm, customer, and government collaboration is sound but operationalization is difficult and fraught with problems. This doesn't suggest that the concept should be abandoned, just managed, and supported. Sadly, this situation is not unique (Lyytinen & Rose, 2003). The missing element is often cooperative knowledge creation and exchange. Each element of the collaboration needs a better understanding and focus on cooperation. Unfortunately, this doesn't naturally exist and easily turns antagonistic. Cooperation is difficult to achieve even when linkages are in place. It is far too easy to say that "details can be worked out." Unfortunately, the "devil" is in the detail. Towards that end, stakeholder collaboration in achieving knowledge management objectives is paramount.

CONCLUSION

The underlying goal for healthcare is to provide cost effective quality treatment (i.e., realize its value proposition in this challenging environment). In order to do this healthcare needs to maximize its information management techniques and make prudent use of ICTs (Information Communication Technologies). In such a context e-health initiatives will clearly play a dominant role in healthcare delivery. This has been underscored by leaders of US and the EU as well as leading bodies such as the World Healthcare Organization (WHO) that focus on global healthcare issues and policy. Moreover, Both European and US authorities define their initiatives primarily in terms of medical information technology centering on computerized patient record [CPR] or, in more acceptable parlance, the HER electronic health record as referred to by WHO. Hence, e-health is here to stay. What becomes critical then is the sustainability of these e-health initiatives and their ability to bring benefits to the key actor in healthcare, the patient.

This article has set out to delve into the abyss of e-health sustainability. A logical starting place to us seemed to identify the primary drivers in a generic e-business model and then map them into healthcare. Our e-health sustainability model then serves to identify the critical factors and important dynamics faced by any e-health initiative. In addition, we identified the importance of scale and scope economies in this process through the mapping of case study data. Finally, we noted that it is necessary to incorporate the techniques and strategies of knowledge management if superior collaboration between the multiple stakeholders is to ensue. Through the example of SARS we underscored how important this aspect is not only to the sustainability of e-health but in order to realize effective healthcare delivery. Clearly this is only the beginning and we now need further investigation and research, which we plan to embark upon. We close by encouraging other researchers to also delve deeper into this imperative healthcare research area.

REFERENCES

Afuah, A., & Tucci, C. L. (2000). *Internet business models and strategies—text and cases*. McGraw Hill Irwin.

Alavi, M. A., & Leidner, D. E. (2001). Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*, *25*(1), 107-136.

Feeny, D. (2001). Making business sense of the e-opportunity. *MIT Sloan Management Review*, *42*(2), 42.

Gargeya, V., & Sorrell, D. (2005). Moving toward an e-hospital. In N. Wickramasinghe, J. Gupta, & S. Sharma (Eds.), *Creating knowledge-based healthcare organizations*, (pp. 50-64). Hershey, PA: Idea Group Publishing.

Glaser, J. (2002). *The strategic application of information technology in health care organizations* (2nd ed.). San Francisco: Jossey Bass.

Hendersen, J., & Venktraman, N. (1992). Strategic alignment: A model for organizational transformation through information technology. In T. Kochan, & M. Unseem (Eds.), *Transforming organisations*. NY: Oxford University Press.

Hofer, C. (1975). Toward a contingency theory of business strategy. *Academy of Management*, *18*(4), 784-810.

La Monica, M. (2000, October). Building trust into e-business models. *InfoWorld*, *22*(28), 3.

Lyytinen, K., & Rose, G. M. (2003). The disruptive nature of information technology innovations: The case of internet computing in systems development organizations. *MIS Quarterly*, *27*(4), 557-595.

Porter, M. (1980). *Competitive strategy*. New York: Free Press.

Porter, M. (1985). Competitive advantage. New York: Free Press.

Porter, M., & Teisberg, E. (2004). Redefining competition in healthcare. *Harvard Business Review*, 65-76, June.

Sharma, S., & Wickramasinghe, N. (2005). e-health with knowledge management: areas for tomorrow. In N. Wickramasinghe, J. Gupta, & S. Sharma (Eds.), *Creating knowledge-based healthcare organizations* (pp. 110-124). Hershey, PA: Idea Group Publishing.

Sharma, S., Wickramasingeh, N., Xu, B., & Ahmed, N. (2006). Electronic healthcare: Issues and challenges. *International Journal of Electronic Healthcare*, *2*(1), 50-65.

Van De Ven, A., Angle, H., & Poole, S. (2000). Research on the management of innovation. Oxford: Oxford University Press.

Von Lubitz, D., & Wickramasinghe, N. (2006). Healthcare and technology: The Doctrine of Networkcentric Healthcare. *International Journal of Electronic Healthcare* (IJEH, in press).

Wickramasinghe, N., & Lamb, R. (2002). Enterprise-wide systems enabling physicians to manage care. *International Journal Healthcare Technology and Management*, *4*(3/4), 288-302.

Wickramasinghe, N., & Silvers, J. (2003). IS/IT: The prescription to enable medical group practices to manage managed care. *Health Care Management Science*, *6*(2), 75-86.

Wickramasinghe, N., Fadlalla, A., Geisler, E., & Schaffer, J. (2005). A framework for assessing e-health preparedness. *International Journal of Electronic Healthcare* (IJEH), *1*(3), 316-334.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, edited by J. Tan, pp. 68-81, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 1.4
The Telehealth Divide

Mary Schmeida
The Cleveland Clinic Foundation, USA

Ramona McNeal
University of Illinois at Springfield, USA

INTRODUCTION

In the United States, the public is accessing the Internet to provide information and deliver services, and to interact with citizens, business, and other government agencies (Bimber, 1999; Pardo, 2000; West, 2003, 2004). As with any change between citizen-government interactions, e-government is accompanied by speculation on its impact to both citizen and government. E-government capability of continual service delivery can make government efficient and transparent to the public (Norris, 1999; West, 2003), and more responsive to public needs through fast and convenient communication options (Thomas & Streib, 2003). It permits quicker material update than traditional distribution methods (Pardo, 2000).

However, other literature suggests e-government will not live up to these prospects. A separation exists among citizens that use and do not use the Internet. This separation is based on a number of factors, including inequalities in Internet access "digital divide" and technological skills, along with psychological and cultural barriers. Literature extensively shows the differences in United States Internet use to fall along important socioeconomic and demographic factors, such as age, race, education, and income (Mossberger, Tolbert, & Stansbury, 2003; Neu, Anderson, & Bikson, 1999; Pew Internet & American Life Project, 2003c; U.S. Department of Commerce, 2002; Wilheim, 2000). E-government may create inequalities in the delivery of government information and services.

Telehealth is a specific form of e-government aimed at improving the accessibility and quality of healthcare, and reducing service costs (Schmeida, 2004a). It relies on electronic information and telecommunication technology innovation (H.R. 2157, 2001). As nations contend with expensive healthcare, the promise of better healthcare service delivery at a reduced cost has made teleheath an increasingly attractive policy option in the United States and internationally.

Telehealth advancement greatly reflects the dramatic changes in the telecommunication industry. In the 1990s, we witnessed considerable advancement, such as the use of digital technology—interactive video and Internet. Interactive video, for example, can link doctors and medical students afar improving medical education. Rural citizens can interact with specialist(s) through interactive video rather than traveling great distances for a medical consultation. The Internet can bring health related information into the home for better healthcare decision-making.

Telehealth can be conceptualized as both an administrative reform policy and regulatory policy. As a hybrid policy type, it mostly exhibits the characteristics of administrative reform, such as e-government (McNeal, Tolbert, Mossberger, & Dotterweich, 2003; Schmeida, McNeal, & Mossberger, 2004) driven by the goals of cost reduction and increasing efficiency, paramount to telehealth adoption and implementation. Administrative reform policy does not involve the direct and coercive use of government power over citizens and are therefore associated with low levels of conflict (Ripley & Franklin, 1980). Regulatory policy, on the other hand tends to be politically salient among citizens as well as controversial among the actors within the policy community.

Traditionally, those interests who are regulated have been important players in the policy process. Important telehealth players are physicians, nurses, pharmacists, and health insurers. Since telehealth straddles both administrative and reform policy areas, it is difficult to predict the actors that will play the greatest role in assisting or impeding its implementation. Execution of regulatory policy is highly volatile and controversial with shifting of alliances and players. However, administrative policy innovations are low salience, and as some regulatory policies it involves technical issues, often driven by professional networks and elected officials.

TREND STUDIES ON INTERNET USE AS A HEALTHCARE TOOL

Although policy actors, cost containment and advances in technology are driving the adoption of telehealth, its impact is contingent on factors, such as Internet access among members of the public. The Internet provides information on various health and medical-related topics through government sponsored and private sector Web sites. Today, more Americans are conducting Internet health and medical-related searches. On an average day, about six million people get online to search for medical-related information (Pew Internet & American Life Project, 2002) for better decision-making on self-care and helping others (Pew Internet & American Life Project, 2003b). Across different health and medical issue topics, Pew (2003b) found online searches for information on a specific disease or medical problem leads the topic areas searched. In addition to using the Internet for health information searches, about 30% of e-mail users have e-mailed their doctors and other health professionals, thus bridging the gap between patient and doctor, particularly specialists over great distances. This elementary form of telehealth (information search, and doctor and patient correspondence) exemplifies the potential of telehealth. Yet, while becoming a commonly important healthcare tool, not all citizens including those most in need of online health and medical-related information (the elderly and poor) are taking advantage of the online services (Schmeida, 2004b).

Research on demographic groups using the Internet to search for health information is sparse and does not establish with any certainty what factors matter in predicting who is taking advantage of this form of e-government (Schmeida, 2004b). However, multivariate statistical research on computer and Internet access does exist and may help us better appreciate the barriers facing the utilization of telehealth. Mossberger, Tolbert, and Stansbury (2003) find both an Internet access and skills divide, indicating inequalities in home Internet access, e-mail use and computer ownership. The research suggests there are gaps based on race/ethnicity with whites more likely than African-Americans and Latinos to have Internet access. Inequalities were also found on education and income with higher income being associated with greater Internet and computer access and ownership. Age was also an important factor with young persons more likely to be connected (pp. 32-35). Compounding the access divide are differences in technological skills. Mossberger et al. (2003) find a skill divide exists closely mirroring the access divide with the poor, older, less educated, and non-whites less likely to have technical skill, that is, technical competence or information literacy. As important as having access at home are the technical skills for computer operations and information literacy to locate and effectively use computer information.

A TELEHEALTH DIVIDE MATCHING THE DIGITAL ACCESS DIVIDE

Do those who are least likely to have computer and Internet access match that of a potential telehealth divide? Based on literature showing disparities in Internet access and use (Mossberger, Tolbert, & Stansbury 2003; Pew Internet & American Life Project 2003a; U.S. Department of Commerce 2002), it is expected that disparities in Internet use for health searches would also exist. Using the 2000 Pew Internet & American Life (2003b) survey data and controlling for demographic factors with regression analysis, Schmeida (2004b) found several factors important in explaining which citizens are searching for health and medical information online.[1] Persons who are young, white, with a higher education and income, are more likely to search for information online.[2] Also, females and household healthcare givers are more likely to conduct searches than males and non-caregiver.[3] The findings suggest that older persons, who face more health-related problems, ironically are less likely to take advantage of these telehealth services. However, females are more likely to search online, as might be expected since they have been more active in health searches in the Pew Internet & American Life Project (2003b) study and are more likely to be caring for another person(s) at home.

Unexpectedly, race (Asian Americans and African-Americans) was not a significant predictor of a telehealth divide, differing from previous studies showing these minority groups are "have-nots" in Internet access at home. On the state level, McNeal et al. (2003) found racial diversity was not a significant predictor in percentage of state government Web sites providing services to state residents. While, Schmeida, McNeal, and Mossberger (2004) found state racial context mattered in influencing implementation of telehealth policy. This suggests that minority context may matter more for telehealth implementation. Latino (who were found significantly less likely to search for online health information) contextual barriers to computer literacy, such as entrenched resistance to acquiring computer skills (Stanley 2003) may be factors for further research.

CONCLUSION

The policy implementation literature directs us to examine factors, such as political actors, the need/demand to contain costs and increase healthcare efficiency, important to explaining the adoption and implementation of telehealth. To understand the impact of this policy, we cannot, however, lose sight of factors, such as public Internet access and use. Research statistically controlling for demographic factors on telehealth use is limited. However, using regression we control for these factors and with certainty find disparities between demographic groups searching for health information on the Internet. A telehealth divide exists and in general mirrors the contours of the digital divide. As with the digital access divide—those who are least likely to have a computer and Internet access at home are the poor, less educated, older, Latinos and African-Americans (Mossberger et al. (2003) face barriers in taking advantage of services made available by telehealth. That is, persons who are poor, less educated, older, and Latino are least likely to search online for health information than others (Schmeida, 2004b).

Persons most in need of online health information are not accessing it. These findings suggest a dilemma, those in greatest need of information, such as, the poor and elderly are not seeking it online. The ramifications are significant suggesting the "advantaged" will obtain superior healthcare and treatment than the "disadvantaged." As advanced technology replaces labor-intensive information specialists, for example, the U. S. Centers for Medicare & Medicaid Services (2004) Web site, parity between demographic groups is essential. As literature suggests that e-government will not live up to the prospects of efficiently improving healthcare for all people, our findings show that indeed e-government maybe creating inequalities in the delivery of government information and services.

Effective information technology use depends on user knowledge and skills. Technical competence, information literacy and basic literacy are required for the information age for effective use of telehealth. Today, public access cites, such as computer technology centers are giving individuals both access to computers and educational opportunities to overcome literacy barriers (Mossberger, et al. 2003). Bridging the gap may be progressed by community services and educational opportunities, yet evolving. The adoption of telehealth reminds us that with adoption of any new policy, it must be evaluated on more criteria then efficiency. Criteria, such as equality in costs, benefits and risks must also be considered.

Although this article considers policy diffusion among the United States, it has implications for the international community in terms of lesson drawing. As the literature reminds us, countries often look to their neighbors for policy ideas (Bennett, 1997; Wolman, 1992) for emulation. This has proven to be the scenario between the United States and many other nations, all-searching to improve the health of their citizens while facing rising heathcare costs.

REFERENCES

Bennett, C. (1997). Understanding the ripple effect: The cross-national adoption of policy instruments for bureaucratic accountability. *Governance*, *10*, 213-233.

Bimber, B. (1999). The Internet and citizen communication with government: Does the medium matter? *Political Communication, 16*(4), 409-28.

Centers for Medicare & Medicaid Services. (2004). *Medicaid information technology architecture*. Retrieved from http://cms.hhs.gov/

House of Representatives 2157. (2001). *Rural Health Care Improvement Act of 2001.* 107th Congress. 1st Session. United States of America. Retrieved from http://www.lexis-nexis.com

McNeal, R., Tolbert, C., Mossberger, K., & Dotterweich, L. (2003). Innovating in digital government in the American states. *Social Science Quarterly, 84*(1), 52-70.

Mossberger, K., Tolbert, C., & Stansbury, C. (2003). *Virtual inequality beyond the digital divide*. Washington, DC: Georgetown University Press.

Neu, C., Anderson, R., & Bikson, T. (1999). *Sending your government a message: E-mail communication between citizens and government.* Santa Monica, CA: Rand Corp.

Norris, P. (1999). *Critical citizens: Global support for democratic governance*. Oxford: Oxford University Press.

Pardo, T. (2000). *Realizing the promise of digital government: It's more than building a Web site*. Information Impacts Magazine. Retrieved from www.cisp.org/imp/october_2000

Pew Internet & American Life Project (2002). *E- patients and the online health care revolution*. Retrieved from http://www.pewinternet.org/

Pew Internet & American Life Project. (2003a). *Internet health resources*. Retrieved from http://www.pewinternet.org/

Pew Internet & American Life Project. (2003b). *The online health care revolution: How the Web helps Americans take better care of themselves*. Retrieved from http://www.pewinternet.org/

Pew Internet & American Life Project. (2003c). *The ever-shifting Internet population: A new look at Internet access and the digital divide*. Retrieved from http://www.pewinternet.org/reports/

Ripley R., & Franklin,G. (1980). *Congress, the bureaucracy, and public policy*. Homewood, Illinois: The Dorsey Press.

Schmeida, M. (2004a). Telehealth and state government policy. In J. Rabin (Ed.), *The Encyclopedia of Public Administration and Public Policy*. Marcel Dekker, Inc.

Schmeida, M. (2004b). *The telehealth divide: Mirroring the contours of the digital divide*. Telehealth innovation in the American states. Doctoral dissertation, Kent State University, Kent, Ohio.

Schmeida, M., McNeal, R., & Mossberger, K. (2004). *The scope of telehealth implementation across the states*. Paper presented at the 2004 Annual Midwest Political Science Association Conference, Chicago, Illinois.

Stanley, L. (2003). Beyond access: Psychosocial barriers to computer literacy. *The Information Society*, *19*, 407-419.

Thomas, J., & Streib, G. (2003). The new face of government: Citizen-initiated contacts in the era of e-government. *Journal of Public Administration Research and Theory*, *13*(1), 83-102.

U. S. Department of Commerce. National Telecommunications and Information Administration. (2002). *A nation online: How Americans are expanding their use of the Internet*. Retrieved from http://www.ntia.doc.gov/ntiahome/dn/anationonline2.pdf

West, D. (2003). *State and federal e-government in the United States, 2003*. Retrieved from www.insidepolitics.org/egovt03us.pdf

West, D. (2004). E-government and the transformation of service delivery and citizen attitudes. *Public Administration Review*, *64*(1), 15-27.

Wilheim, A. (2000). *Democracy in the digital age: Challenges to political life in cyberspace*. New York: Routledge.

Wolman, H. (1992). Understanding cross-national policy transfer: The case of Britain and the United States. *Governance*, *5*, 27-45.

KEY TERMS

Centers for Medicare & Medicaid Services Web site: A United States government Web site that provides information to needy groups on Medicare and Medicaid services, state children's health insurance, and clinical laboratory testing.

Digital Access Divide: Disparities between groups on demographic factors, such as gender, age, income, education, race, and ethnicity in accessing digital information technology, such as on the Internet.

Telehealth: The term is often interchanged with telemedicine. There is no consensus on its definition. However, the United States Congress defines it as the use of electronic information and telecommunications technologies to support public health and health administration, long-distance clinical healthcare, patient and professional health-related education (H.R. 2157, 2001).

Telehealth as an Administrative Reform Policy: As an administrative reform policy, telehealth is highly technical, not political charged although it may be conflictual in the policy community. Unlike regulatory and re-distributive policies, it does not involve the direct and coercive use of government power over citizens. Telehealth behaves mostly as an administrative reform, with its policy goals emphasizing greater efficiency in terms of cost savings.

Telehealth as a Regulatory Policy: Although regulatory policy may be politically salient among citizens, telehealth may be conflictual within the policy community. The regulatory aspects in this policy area may engage the participation of networks of healthcare interest groups, such as nurses, pharmacists and physicians. A moderate level of conflict over the issue may lead to some interest group activity by those who are liable to be covered by state regulations.

Telehealth Divide: Disparities between groups on demographic factors, such as gender, age, income, education, and race in using digital information technology, such as the Internet to search for public and private health and medical-related information online.

Telehealth Policy: An electronic government policy that uses the Internet to improve accessibility of public and private and non-profit healthcare services in rural and urban areas, while improving the quality of services at a lower service cost.

ENDNOTES

1. Similar question as the Pew survey is asked but demographic factors are controlled for by using negative binomial regression analysis: Have you ever looked online for any information on any of the 16 health and medical issues?" The 16 issues are: specific disease or medical problem; certain medical treatment or procedure; experimental treatments or medicines; alternative treatments or medicines; diet, nutrition, vitamins, or supplements; exercise or fitness; prescription or over the counter drugs; immunizations or vaccinations, how to quit smoking, problems with drugs or alcohol; depression, anxiety, stress or mental-health issues; environmental health hazards; sexual health; particular doctor or hospital; health insurance; and Medicare or Medicaid insurance. Our dependent variable is a count of all "yes" responses to each health and medical issue topic, ranging from 0 to 16. An individual who has searched for all 16 medical issues is coded as 16 and someone who has looked for no medical issues is coded 0 (Pew Internet & American Life Project, 2003b).
2. Age is in years, while education is a 7-point Likert scale measuring the last grade or class completed in school. Income is measured on an 8-point scale measuring the total family income from all sources before taxes in 2001 ranging from less than $10,000 to $100,000 or more (Pew Internet & American Life Project, 2003b).
3. Household healthcare giver is defined as either a primary or secondary healthcare giver to a member in their household (Pew Internet & American Life Project, 2003b).

This work was previously published in Encyclopedia of Digital Government, edited by A. Anttiroiko and M. Malkia, pp. 1524-1528, copyright 2007 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.5
Health Portals:
An Exploratory Review

Daniel Carbone
Victoria University, Australia

Stephen Burgess
Victoria University, Australia

INTRODUCTION

A lack of health services has long been the thorn in the side of many communities, especially rural and regional communities. The high costs of treating ever growing chronic and complex conditions in traditional settings, where rural allied health services providers are non-existent and doctors are already overcommitted, are prompting a shift in focus to more efficient technology driven delivery of health services. Moreover, these days it is also increasingly unlikely that health professionals will encounter patients who have not used information technology to influence their health knowledge, health behaviour, perception of symptoms, and illness behaviour.

Advances in Internet technologies offer promise towards the development of an e-health care system. This article will postulate whether portal technologies can play a role facilitating the transition to such e-health care systems.

This article aims at reviewing the literature to present to the reader the barriers and opportunities out here for effective health portals. However, the article does not intend to provide a one-fits-all technical/content solution, only to make implementers and developers aware of the potential implications.

BACKGROUND

Many rural and regional communities lack the range of allied health services that are readily available in metropolitan areas, and many rural doctors who are already overcommitted, provide services that an allied health professional could readily provide (Department of Health and Ageing, 2004). The Australian Institute of Health and Welfare data shows that death and disability from chronic disease is higher in rural and regional communities, including Indigenous people. Coronary heart disease, asthma and diabetes are the biggest killers. Participants in the Regional Australia Summit highlighted chronic disease as a major menace (Department of Health and Ageing, 2004).

This state of affairs is already prompting a change in the health care system to focus more on preventive medicine and health care away from the traditional settings (Yellowlees & Brooks, 1999). According to Yellowlees and Brooks (1999), there are three major drivers for this change:

- The *economic imperative* to restrain health care costs
- *Increasing consumerism*, and the evolution of the "informed patient"
- *Changes in communication technology*, and the evolution of the Internet

PORTALS AND HEALTH

The benefits of Web portals in aggregating information from multiple sources and making that information available to various users is well known; more importantly, they can provide the services of a guide that can help to protect the user from the chaos of the Internet and direct them towards an eventual goal (Tatnall, 2005). More generally, however, a portal should be seen as providing a gateway not just to sites on the Web, but to *all network-accessible resources,* whether involving intranets (within an organisation), extranets (for special partners of an organisation), or the Internet (Tatnall, Burgess, & Singh, 2004). In other words a portal offers centralised access to all relevant content and applications (Tatnall, 2005).

The literature on health portals tells us that the Internet offers a significant amount of health information of varying quality. Health portals, which provide entry points to quality-controlled collections of Web sites, have been hailed as a solution to this problem (Glenton, Paulsen, & Oxman, 2005). However, it has been demonstrated that the information accessible through (government run and funded) health portals is unlikely to be based on systematic reviews and is often unclear, incomplete and misleading. Portals are only as good as the Websites they lead to (Glenton et al., 2005). However, irrelevant information could easily be filtered using a number of frameworks that can be used to evaluate the quality of Web-located health information. For example, Sellito and Burgess (2005) have developed a set of affirmative response evaluation features identified across four quality categories: currency/authority, accuracy, objectivity and privacy. And they are used as the basis for determining the fundamental quality of Web-located health information (Sellitto & Burgess, 2005).

THE CONSUMER AND HEALTH INFORMATION

Increasingly, consumers are accessing health information via the Web (Thompson & Brailer, 2004). It has been estimated that 6.4 million Australian adults—almost half the adult population—accessed the Internet during 2000 (Gretchen, Berland, Elliott et al., 2001). This is not just an Australian phenomenon. In the United States, 52 million Americans access health or medical information on the Web (Fox & Fallows, 2003).

The existence of health portals has made life easier for the people that need this information. However, the quality of portal interfaces as well as the portal content has many times been in doubt (Bamidis, Kerassidis & Pappas, 2005). Using popular search engines may be aesthetically appealing and easy to use, but they often provide inaccurate information (Sutherland, Wildemuth, Campbell, & Haines, 2005). What is clear however, is that while most consumers still use word-of-mouth as a primary information source for health care decisions, the use of Internet information is increasing (Snipes, Ingram, & Jiang, 2005). In Australia, for example, more Internet users search the Web for information on depression than any other health condition (Lissman & Boehnlein, 2001). This is not surprising given the high level of disability associated with depression in the community and the fact that the Web provides a convenient, anonymous means of obtaining information about the problem (Cain, Sarasohn-Kahn, & Wayne, 2000). However, much of the depression information on the Web is of low quality and originates in the United States (Jadad & Gagliardi, 1998).

SERVICE PROVIDERS AND HEALTH INFORMATION

General Practitioners (GPs)

The gap between what GPs might do (based on evidence-based clinical practice guidelines and what they actually do is wide, variable and growing. Many factors contribute to this situation. GPs are inundated with new, often poorly evidence-based and sometimes conflicting clinical information. This is particularly serious for the generalist, with over 400,000 articles added to the biomedical literature each year. Adding further pressure to the "gap" are workloads that have increased over the past decade: GPs are seeing more patients with acute and complex conditions. Rural practitioners work even longer hours, offer more medical services and perform more clinical procedures than their urban counterparts—thus facing an even greater need for up-to-date information (Davis, Ciurea, Flanagan, & Perrier, 2004).

There are four steps in incorporating research evidence in clinical decision making: *asking* answerable questions; *accessing* the best information; *appraising* the information for validity and relevance; and *applying* the information to patient care (Craig, Irwig, & Stockler, 2001). However, a study in New Zealand suggested that to make this happen, practitioners urgently need training in searching and evaluating information on the Internet and in identifying and applying evidence-based information; as well as (health) portals to provide access to high-quality, evidence-based clinical and patient information along with access to the full text of relevant items (Cullen, 2002). Many sites have been developed to help the search for quality peer-reviewed literature. These include the Cochrane Library and the U.S. National Library of Medicine's PubMed, as well as sites offering full-text access to medical journals, such as Stanford University's HighWire Press and freemedicaljournals.com (Robinson & Day, 2004). GPs can keep up to date with reliable information from readily accessible Web sites such as PubMed and HighWire Press. PubMed is part of the National Library of Medicine in the U.S. It is a useful system for retrieving clinically relevant search results. HighWire Press has a less sophisticated search engine, but is an excellent source for obtaining the full text of journal articles (Robinson & Day, 2004). However, and although increasing, access to these resources by practitioners is still low (Young & Ward, 1999).

Nurses

E-health can deliver health care services and education, via a Web portal, to older persons with chronic conditions and their caregivers and enables the patient's home to be the point of care. This growing industry is ripe for exploration by nurses who can empower the patient and caregiver to gain self-care and coping skills. Advances in information technology now make this dream a reality (Moody, 2005). However, at the American Academy of Nurse Practitioner's Conference, it was identified that information on educational options for acute care nurse practitioner (ACNP) practice was needed (Kleinpell, Perez, & McLaughlin, 2005). Information technology skills of nurse managers and staff need to be developed in order to use information technology effectively. In order to learn in a Web-based environment, everyone needs the opportunity and access to required resources. Additionally, nurse managers' experiences are important to promote wider utilisation of Web-based learning (Korhonen & Lammintakanen, 2005).

Web portals could help nursing staff in a number of ways; for example, health assessment skills are vital to professional nursing practice. Health assessment has traditionally been taught using lecture, teacher-developed tests, practice and live demonstration, and interactive and computer-based learning materials.

Student evaluation of these types of courses revealed that online assignments enabled them to pace their learning, thereby promoting greater flexibility and independence. Students were able to master the technical skills of working online with minimal difficulty and reported that working online was no more stressful than attending class. A most helpful aspect of the online course was the instructor-developed video that was digitally streamed online (Lashley, 2005).

Hospitals

International health organisations and officials are bracing for a pandemic. For example, and although the 2003 severe acute respiratory syndrome (SARS) outbreak in Toronto did not reach such a level, it created a unique opportunity to identify the optimal use of the Internet to promote communication with the public and to preserve health services during an epidemic (Rizo, Lupea, Baybourdy, Anderson, Closson, & Jadad, 2005). What was learned was that many patients are willing and able to use the Internet as a means to maintain communication with the hospital during an outbreak of an infectious disease such as SARS. Hospitals should explore new ways to interact with the public, to provide relevant health information, and to ensure continuity of care when they are forced to restrict their services (Rizo et al., 2005).

PROVIDER EDUCATION AND HEALTH PORTALS

Claire Jackson (2005), the chair of the discipline of general practice at the University of Queensland, Australia, envisioned the primary care practitioner increasingly networked with consumers, government and professional groups, such as colleges and divisions of general practice. Primary and continuing medical education needs to play a principal role in this process. Education needs a fundamental change of focus from simply delivering content to developing the ability to manage these changes. Learning to learn and learning for life should be a major guiding influence in curriculum development (Carlile & Sefton, 1998).

Portals can certainly provide practitioners easy access to these resources; however, it has been argued that student health professionals lacked the state of readiness of for Web-based learning environments. A short survey was distributed to the Medical Faculty at Sheffield and 191 valid responses were received. Only 62% of students had access to an Internet-connected computer at home. Most students (95.8%) checked their e-mail every few days or more, with slightly less (82.8%) using the Web frequently. Relevant technologies were often never used, including Internet relay chat, message forums and video conferencing. However, 66% of students had used computer aided learning packages. Future use of online continuing professional education material is likely to be limited (Stokes, Cannavina, & Cannavina, 2004). Nevertheless, various studies have shown that appropriately designed, evidence-based online continuing medical education can produce objectively measured changes in behaviour as well as sustained gains in knowledge that are comparable or superior to those realised from effective live activities (Fordis et al., 2005).

Some very recent developments has a Pfizer-sponsored educational Web portal for GPs allows the company to track the advertisements doctors look at and the Web links they visit. Believed to be the first pharmaceutical company-sponsored portal for GPs in Australia, the My E-Portal site (www.myeportal.com.au) allows GPs to drag-and-drop links to their most frequently visited sites, and provides journal and division sites, access to continuing medical information, and links to entertainment, banking and travel sites. Pfizer can then collect information about what sites are accessed, the ads and links clicked on, and the links added to the site (Limprecht, 2005).

ELECTRONIC COMMUNICATION AND PORTALS

The Internet also offers a unique means of health promotion through the use of interactive tools like chat rooms, e-mail, hyperlinks and the like (Stout, Villegas, & Kim, 2001). Looking at all these is beyond the scope of this article, however, a brief look at e-mail communication will suffice to outline some of the basic issues facing the e-health care system of the future.

In a recent American study, a Web-based communication strategy (e-mail) was used to enhance communication between patients and GPs, where a Web mail address was promoted on the telephone (Spencer, 2005). An important observation from this study is that less than half of e-mails require the direct attention of the physician. This study is also supported by other similar research findings (Griffiths & Christensen, 2002). This of course has a number of ethical issues that need to be explored before going any further (Flicker, Haans, & Skinner, 2004).

DISCUSSION

The literature is full of evidence on portal's potential use in health, but it is all compartmentalised: there are GP studies, nurses' studies, hospital studies, patient studies, communication studies and so on. Furthermore, since the inception of the computer age, and even now with the advances in online technologies, there is ample evidence to suggest that development and implementation of these tools always lie in the realm of the technologists; where the technology is the focus of the implementation rather than the user's outcome (Tatnall, Davey, Burgess, Davison, & Fisher, 2000).

The technical issues involving portals is well documented; however, portals are but one component of the larger Health Information System. This simple fact needs to be acknowledged and efforts for wider research into the many facets of online health users and their subsystems must be taken into account, not as neat individual groups as current research seems to place them, but as dynamic partners of a Health Information System.

It is with this fear that this article has been written, in the hope that somehow developers and implementers would take heed of the barriers and opportunities for cross-field efforts to develop workable online tools that would produce, in this case, positive health outcomes.

CONCLUSION

This review recognises the potential for Web information technologies to affect some of the uses of these technologies in the development of an e-health care system for communities. However, for every potential, there are lessons that need to be embraced before rushing to developing portal technologies; for example:

1. The lack of user training is usually apparent when new technologies are introduced.
2. The need to be able to appropriately filter information to instill consistency and confidence on users of the resources.
3. Not all GPs are yet convinced the evidence-based guidelines are the clinical future for the treatment of chronic and complex conditions. This is perhaps the biggest obstacle

to an uptake of Web-based resources and treatment.

4. Health communication tools like e-mail presents an interesting challenge for clinicians, clinical treatment and ethical issues.
5. The overall message from this article is *proceed with caution*. The potential for portals is definitely there, and making users adopt them is perhaps the key to it.

REFERENCES

Bamidis, P., Kerassidis, F., & Pappas, K. (2005). Health information on the Internet: Evaluating Greek health portals and depicting users' attitudes in West Macedonia. *Studies in Health Technology and Informatics, 116*, 885-890.

Cain, M., Sarasohn-Kahn, J., & Wayne, J. (2000). *Health e-people: The online consumer experience. Five-year forecast.*

Carlile, S., & Sefton, A. J. (1998). Health care and the information age: Implications for medical education. *Medical Journal of Australia*, *168*, 340-343.

Craig, J. C., Irwig, L. M., & Stockler, M. R. (2001). Evidence-based medicine: Useful tools for decision making. *Medical Journal of Australia, 174*, 248-253.

Cullen, R. J. (2002). In search of evidence: Family practitioners' use of the Internet for clinical information. *Journal of the Medical Library Association*, *90*(4), 370-379.

Davis, D. A., Ciurea, I., Flanagan, T. M., & Perrier, L. (2004). Adopting best evidence in practice solving the information overload problem: A letter from Canada. *Medical Journal Association, 180*(6 Suppl.), S68-S71.

Department of Health and Ageing. (2004). *Better health services for rural and regional Australia*. Retrieved January 14, 2007, from the Commonwealth of Australia, http://www.health.gov.au/internet/wcms/publishing.nsf/Content/health-budget2000-fact-ruralfact3.htm

Flicker, S., Haans, D., & Skinner, H. (2004). Ethical dilemmas in research on Internet communities. *Qualitative Health Research*, *14*(1), 124-134.

Fordis, M., King, J. E., Ballantyne, C. M., Jones, P. H., Schneider, K. H., Spann, S. J., Greenberg, S. B., & Greisinger, A. J. (2005). Comparison of the instructional efficacy of Internet-based CME with live interactive CME workshops: A randomized controlled trial. *The Journal of the American Medical Association, 294*, 1043-1051.

Fox, S., & Fallows, D. (2003). *Health searches and email have become more commonplace, but there is room for improvement in searches and overall Internet access.* Washington, DC.

Glenton, C., Paulsen, E. J., & Oxman, A. D. (2005). Portals to wonderland: Health portals lead to confusing information about the effects of health care. *BioMed Central Medical Informatics and Decision Making, 5*(1), 7.

Gretchen, K., Berland, M., Elliott, M., et al. (2001). Health information on the Internet: Accessibility, quality, and readibility in English and Spanish. *The Journal of the Americal Medical Association, 285*, 2612-2621.

Griffiths, K. M., & Christensen, H. (2002). The quality and accessibility of Australian depression sites on the World Wide Web. *Medical Journal Association, 176*(10 Suppl.), S97-S104.

Jadad, A., & Gagliardi, A. (1998). Rating health information on the Internet: Navigating to knowledge or to Babel? *The Journal of the Americal Medical Association*, *279*, 611-614.

Kleinpell, R., Perez, D., & McLaughlin, R. (2005). Educational options for acute care nurse practitioner practice. *Journal of the American Academy of Nurse Practitioners*, *17*(11), 460-471.

Korhonen, T., & Lammintakanen, J. (2005). Web-based learning in professional development: Experiences of Finnish nurse managers. *Journal of Nursing Management, 13*(6), 500-507.

Lashley, M. (2005). Teaching health assessment in the virtual classroom. *The Journal of Nursing Education, 44*(8), 348-350.

Limprecht, E. (2005). *Pfizer Web site tracks GP use.* Retrieved January 14, 2007, from http://www.australiandoctor.com.au/articles/a3/0c0368a3.asp

Lissman, T., & Boehnlein, J. (2001). A critical review of Internet information about depression. *Psychiatric Services, 52*, 1046-1050.

Moody, L. E. (2005). E-health Web portals: Delivering holistic health care and making home the point of care. *Holistic Nursing Practice, 19*(4), 156-160.

Rizo, C., Lupea, D., Baybourdy, H., Anderson, M., Closson, T., & Jadad, A. (2005). What Internet services would patients like from hospitals during an epidemic? Lessons from the SARS outbreak in Toronto. *Journal of Medical Internet Research, 7*(4), e46.

Robinson, A., & Day, S. (2004). The value of PubMed and HighWire Press for the busy general practitioner. *Australian Prescriber, 27,* 16-18.

Sellitto, C., & Burgess, S. (2005). Towards a weighted average framework for evaluating the quality of Web-located health information. *Journal of Information Science, 31*(4), 260-272.

Snipes, R., Ingram, R., & Jiang, P. (2005). Information search in health care decision-making: A study of word-of-mouth and Internet information users. *Journal of Hospital Marketing & Public Relations, 15*(2), 33-53.

Spencer, D. C. (2005). Innovations in the practice of primary care: Communicating with patients through e-mail. *North Carolina Medical Journal, 66*(3).

Stokes, C. W., Cannavina, C., & Cannavina, G. (2004). *Health Informatics Journal, 10*(3), 195-204.

Stout, P. A., Villegas, J., & Kim, H. (2001). Enhancing learning through use of interactive tools on health-related Web sites. *Health Education Research, 16*(6), 721-733.

Sutherland, L., Wildemuth, B., Campbell, M., & Haines, P. (2005). Unraveling the Web: An evaluation of the content quality, usability, and readability of nutrition Web sites. *Journal of Nutrition Education and Behavior, 37*(6), 300-305.

Tatnall, A. (2005). *Portals, portals everywhere. Web portals: The new gateways to Internet information and services.* Hershey, PA: Idea Group Publishing.

Tatnall, A., Burgess, S., & Singh, M. (2004). *Community and regional portals in Australia: A role to play for small businesses? Electronic commerce in small to medium enterprises: Frameworks, issues and implications.* Hershey, PA: Idea Group Publishing.

Tatnall, A., Davey, B., Burgess, S., Davison, A., & Fisher, J. (2000). *Management information systems: Concepts, issues, tools and applications* (2nd ed.). Heidelberg, Vic.: Data Publishing.

Thompson, T. G., & Brailer, D. J. (2004). *The decade of health information technology: Delivering consumer-centric and information-rich health care framework for strategic action*.

Yellowlees, P. M., & Brooks, P. M. (1999). Health online: The future isn't what it used to be. *Medical Journal Association, 171,* 522-525.

Young, J. M., & Ward, J. E. (1999). General practitioners' use of evidence databases. *Medical Journal Association, 170,* 56-58.

KEY TERMS

Broadband Incentives: The Australian federal government provided incentives to cover the cost of voluntary connection and use of broadband in general practices to improve their poor access to the Internet.

Evidence-Based Medicine (EBM): Evidence-based medicine is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research. By individual clinical expertise we mean the proficiency and judgement that individual clinicians acquire through clinical experience and clinical practice. Increased expertise is reflected in many ways, but especially in more effective and efficient diagnosis and in the more thoughtful identification and compassionate use of individual patients' predicaments, rights, and preferences in making clinical decisions about their care. By best available external clinical evidence, we mean clinically relevant research, often from the basic sciences of medicine, but especially from patient centred clinical research into the accuracy and precision of diagnostic tests (including the clinical examination), the power of prognostic markers, and the efficacy and safety of therapeutic, rehabilitative, and preventive regimens. External clinical evidence both invalidates previously accepted diagnostic tests and treatments and replaces them with new ones that are more powerful, more accurate, more efficacious, and safer (Sackett et al., 1996).

Generalist: Refers to a general practitioner (GP) or sometimes referred to as medical doctor (MD) as opposed to specialists (cardiologist, neurologist, etc.).

General Practice (GP): Primary care is delivered by some 9,000 practices in Australia, housing some 20,000 GPs; these vary from large practices with 10-15 doctors to many single doctor practices. These are typically doctor owned and run independently as small businesses although the government has a major influence in the way services are delivered and charged.

This work was previously published in Encyclopedia of Portal Technologies and Applications, edited by A, Tatnall, pp. 431-436, copyright 2007 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.6
Semantic Web Standards and Ontologies in the Medical Sciences and Healthcare

Sherrie D. Cannoy
The University North Carolina at Greensboro, USA

Lakshmi Iyer
The University North Carolina at Greensboro, USA

ABSTRACT

This chapter will discuss Semantic Web standards and ontologies in two areas: (1) the medical sciences field and (2) the healthcare industry. Semantic Web standards are important in the medical sciences since much of the medical research that is available needs an avenue to be shared across disparate computer systems. Ontologies can provide a basis for the searching of context-based medical research information so that it can be integrated and used as a foundation for future research. The healthcare industry will be examined specifically in its use of electronic health records (EHR), which need Semantic Web standards to be communicated across different EHR systems. The increased use of EHRs across healthcare organizations will also require ontologies to support context-sensitive searching of information, as well as creating context-based rules for appointments, procedures, and tests so that the quality of healthcare is improved. Literature in these areas has been combined in this chapter to provide a general view of how Semantic Web standards and ontologies are used, and to give examples of applications in the areas of healthcare and the medical sciences.

INTRODUCTION

"One of the most challenging problems in the healthcare domain is providing interoperability among healthcare systems" (Bicer, Laleci, Dogac, & Kabak, 2005). The importance of this interoperability is to enable universal forms of knowledge representation integrate heterogeneous information, answer complex queries, and pursue data integration and knowledge sharing in healthcare (Nardon & Moura, 2004). With the recent emergence of EHRs and the need to distribute medical information across organizations, the Semantic Web can allow advances in sharing such information across disparate systems by utilizing ontologies to create a uniform language and by using standards to allow interoperability in transmission. The purpose of this article is to provide an overview of how Semantic Web standards and ontologies are utilized in the medical sciences and healthcare fields. We examine the healthcare field as the inclusion of hospitals, physicians, and others who provide or collaborate in patient healthcare. The medical sciences field provides much of the research to support the care of patients, and their need lies in being able to share and find medical research being performed by their colleagues to build upon current work. Interoperability between these different healthcare structures is difficult and there needs to be a common "data medium" to exchange such heterogeneous data (Lee, Patel, Chun, & Geller, 2004).

Decision making in the medical field is often a shared and distributed process (Artemis, 2005). It has become apparent that the sharing of information in the medical sciences field has been prevented by three main problems: (1) uncommon exchange formats; (2) lack of *syntactic* operability; and (3) lack of *semantic* interoperability (Decker et al., 2000). Semantic Web applications can be applied to these problems. Berners-Lee, Hendler, and Lassila (2001), pioneers in the field of the Semantic Web, suggest that "the semantic web will bring structure to the meaningful content of web pages". In this article published in *Scientific American*, they present a scenario in which someone can access the Web to retrieve information—to retrieve treatment, prescription, and provider information based on one query. For example, a query regarding a diagnosis of melanoma may provide results which suggest treatments, tests, and providers who accept the insurance plan with which one participates. This is the type of contextually based result that the Semantic Web can provide. The notion of ontologies can be utilized to regulate language, and standards can be used to provide a foundation for representing and transferring information. We will focus on the lack of semantic and syntactic interoperabilities in this article. The semantic interoperable concept will be utilized in the context of ontologies, and syntactic interoperabilities are referred to as standards of interoperability.

BACKGROUND

The Semantic Web is an emerging area of research and technology. Berners-Lee (1989) proposed to the Centre Europeen pour la Recherche Nuclaire (CERN) the concept of the World Wide Web. He has been a pioneer also in the concept of the Semantic Web and has expressed the interest of the healthcare field to integrate the silos of data that exist to enable better healthcare (Updegrove, 2005). He has been involved with the World Wide Web Consortium (W3C) Web site (http://www.w3.org), which offers a vast array of Semantic Web information in a variety of subject areas, including the medical sciences and healthcare. Miller (2004) states that the Semantic Web should provide common data representation to "facilitate integrating multiple sources to draw new conclusions;" and to "increase the utility of information by connecting it to its definitions and context". Kishore, Sharman, and Ramesh (2004) wrote two articles which provide detailed information about ontologies and information systems.

The concept of the Semantic Web is to extend the current World Wide Web such that context and meaning is given to information (Gruetter & Eikemeier, 2004). Instead of information being produced for machines, information will be produced for human consumption (Berners-Lee et al., 2001). There are two main aspects of Semantic Web development: (1) ontologies for consistent terminology and (2) standards for interoperability.

Ontologies

Ontologies have been defined in many ways through the areas of philosophy, sociology, and computer science. For the Semantic Web context, ontology is the vocabulary, terminology, and relationships of a topic area (Gomez-Perez, Fernandez-Lopez, & Corcho, 2004). Ontology gives the meaning and context to information found in Web resources (databases, etc.) for a specific domain of interest, using relationships between concepts (Singh, Iyer, & Salam, 2005). According to Pisnalli, Gangemi, Battaglia, and Catenacci (2004), ontologies should have:

1. *Logical consistency* and be expressed in a "logical language with an explicit formal semantics.
2. *Semantic coverage* such that it covers "all entities from its domain."
3. *Modeling precision* and represent "only the intended models for its domain of interest."
4. *Strong modularity* for the domain's "conceptual space. . .by organizing the domain theories."
5. *Scalability* so that the language is expressive of intended meanings.

The domain of an ontology should include a taxonomy of classes, objects, and their relations, as well as inference rules for associative power (Berners-Lee et al., 2001). This shared understanding of the concepts and their relationships allows a means to integrate the knowledge between disparate healthcare and medical science systems. Much of the Semantic Web research in the medical sciences area has been specific in either generating more efficient and effective information searching or to the interoperability of the EHR. Health information is inherently very tacit and intuitive, and the terminology often implies information based on physical examinations and expressions of the patient. While it uses standardized terminology, the difficulty lies in the expression of this tacit knowledge to others, especially across a network of computers. The two great needs in the medical sciences and healthcare that can be fulfilled by Semantic Web are to standardize language and to provide a consistent foundation for transferring EHR information (Decker et al., 2000).

Standards

While ontologies represent the conceptual basis for the information to be transmitted, standards allow for consistent transmission of the data between disparate systems. The data in different clinical information systems silos are in multiple formats, and relevant medical and healthcare knowledge must be accessible in a timely manner. This can be performed through interoperability standards which can enable information integration, "providing transparency for healthcare-related processes involving all entities within and between hospitals, as well as stakeholders such as pharmacies, insurance providers, healthcare providers, and clinical laboratories" (Singh et al., 2005, p. 30). The main standard for interoperability in the Semantic Web is Resource Description Framework (RDF), which is recommended by the W3C. RDF is an object-oriented based standard, which provides reusable components for data interchange over the web (Decker, Mitra, et al., 2000). It is unique in that every concept represented in RDF has a universal unique identifier (the Uniform Resource Identifier [URI]), which identifies every e-mail address, Web page, and other Web elements. This ensures no semantic ambiguity. RDF also enables knowledge representation through a series of concepts such as class, data type, and values. In order to express representations of ontologies for context, RDF allows for extensions such as the DARPA Agent Markup Language +Ontology Inference Layer (DAML+OIL) standard, which is the basis for the Web Ontology Language (OWL) standard that has recently gained popularity (Nardon & Moura, 2004).

SEMANTIC WEB APPLIED STANDARDS AND ONTOLOGIES IN THE MEDICAL SCIENCES AND HEALTHCARE

"The semantic web initiative has resulted in a common framework that allows knowledge to be shared and reused across applications" (Health Level 7, 2004) and organizations. An infrastructure of common transmission standards and terminology will enable an interconnected network of systems that can deliver patient information. There have been various calls for the decrease of medical errors via utilization of information technology, and the increase of medical information accessibility and Semantic Web technology has a critical role to play. Besides the delivery of patient information, the Semantic Web can also assist medical sciences research in providing greater accessibility and the sharing of research. In the search for information, the Semantic Web can impart a context and meaning to information so that queries are more efficient in producing results more closely related to the search terms.

Table 1 displays only a few of the main standards currently used for interoperability in the Semantic Web. The affiliated organizations are listed, showing that there are many grassroots efforts involved in generating standards. There are three main organizations that are involved in international standards for EHRs. These include the International Organization for Standardization (ISO), Committee European Normalization (CEN), and Health Level 7 (HL7)—U.S. based (HL7, 2004). Standards are also important to develop on an international basis because countries also report national health status statistics to the world community (Cassidy, 2005).

A list of ontologies in the medical domain is listed in Table 2. For clarification, a logical association to an ontology is that of the ICD-9 (ICD-10 is the new version) coding for diseases. When a patient visits the physician, the physician records a standard ICD-9 code for the diagnosis of the patient and a CPT code for the procedure that was performed on a patient. These are standardized codes that are found in manuals for medical coders; and they allow insurance companies and other medical affiliates to understand information from many different sources. For example, if a patient is seen for a mole, the mole can have many particular qualities. Is it to be removed for cosmetic purposes, or is the mole potentially cancerous? The location of the mole will be important to know, as well, because the treatment may be determined by the location. The difference in the context may determine whether the insurance company will pay for the treatment of the mole. A cancerous melanoma on the nose would have the diagnosis code of 172.3 and a benign neoplasm would be coded as 238.2. If a tissue sample were taken so that the lab could test the mole for cancerous cells, the diagnosis would be 239.9, which is unspecified until the lab results return for a firm diagnosis. The CPT procedure code for the treatment would be applied and would be determined by a number of factors including the location of the mole, amount of tissue excised, whether a modifier needs to be added to the code if the services is charged with an office visit, and the type of excision utilized. While we have CPT and ICD-9 as a vocabulary for procedure and diagnosis codes, they function only as a part of ontology's purpose. An ontology gives context to the patient's medical history and allows the diagnosis and procedure to be automatically

Table 1. Sample standards for interoperability

Name	Purpose	Associated Organization	Source
XML	eXtensible Markup Language; creation of tags		Decker et al, 2000
RDF	Standardized technology for metadata; for interpreting meaning	W3C	Nardon, 2004 Gruetter, et al, 2004
Clinical Document Architecture CDA	Leading standard for clinical and administrative data exchange among organizations	HL7	Nardon, 2004 Hooda et al 2004
Guidelines Interchange Format (GLIF)	specification for structured representation of guidelines	InterMed Collaboratory	Nardon, 2004 www.glif.org
CORBAmed	Provides interoperability among health care devices	Object Management Group	McCormack, 2000
HL7	Messaging between disparate systems	HL7	www.hl7.org

Table 2. Sample ontologies (is a terminology coding scheme and would be subsumed by an ontology)*

Name	*Purpose*	*Associated Organization*	*Source*
OIL	Oil Interchange Language; representation and inference language	European Community (IBROW and On-To-Knowledge)	Decker et al, 2000 http://www.ontoknowledge.org/oil/oilhome.shtml
Ontology Web Language (OWL)	Aim is to be the Semantic Web standard for ontology representation	W3 Consortium	Nardon, 2004
DAML	Extension of RDF which allows ontologies to be expressed; formed by DARPA Markup	DAML Researcher Group	Nardon, 2004 http://www.daml.org/
Arden Syntax	Standard for medical knowledge representation	HL7	Nardon, 2004 http://cslxinfmtcs.csmc.edu/hl7/arden/
Riboweb Ontology	Facilitate models of ribosomal components and compare research results	Helix Group at Stanford Medical Informatics	Hadzic et al, 2005 http://smi-web.stanford.edu/projects/helix/riboweb.html
Gene Ontology	To reveal information regarding the role of an organism's gene products	GO Consortium	Hadzic et al, 2005 http://www.geneontology.org/index.shtml
LinkBase	Represents medical terminology by algorithms in a formal domain ontology	L&C	Hadzic et al, 2005
GALEN	Uses GRAIL language to represent clinical terminology	OpenGALEN	Gomez-Perez, 2004
ADL	Formal language for expressing business rules	openEHR	www.openEHR.org
SNOWMED*	Reference terminology	SNOMED Int'l	Cassidy, 2005
LOINC (Logical	Database for universal names and codes for lab and clinical observations	Regenstrief Institute, Inc.	McCormack, 2000 Gillespie, 2003
UMLS—Unified Medical Language	Facilitates retrieval and integration of information from multiple sources; can be used as basic ontology for any medical	US National Library of Medicine	Nardon, 2004 Hadzic, 2005 Gomez-Perez, 2004
ICD-10*	Classification of diagnosis codes; is newer version after ICD-9	National Center for Health Statistics	Gillespie, 2003
CPT Codes*	Classification of procedure codes	American Medical Association	Gillespie, 2003

linked, possibly with appropriate medications, lab tests, and x-rays. The next section discusses ways that the Semantic Web has been applied in the medical sciences field.

SEMANTIC WEB APPLICATIONS IN MEDICAL SCIENCE

Table 3 lists only a few of the sample projects being conducted in the medical science and healthcare field. Previous research in this area has dealt with two main topics: (1) efficient and effective searches of medical science information and (2) the interoperability of EHRs. Our purpose is to provide a comprehensive review of this research to understand the current status of the Semantic Web in healthcare and medical sciences and to determine what future research may be performed.

Electronic Health Records

EHRs are comprehensive patient medical records which show a continuity of care. They contain a patient's complete medical history with information on each visit to a variety of healthcare providers, as well as medical tests and results, prescriptions, and other care histories. (Opposed to EHRs, Electronic medical records [EMRs] are typically those which reside with one physician.) Figure 1 shows the main stakeholders in the healthcare industry, and thus, the necessity for enabling these partners to communicate. Physician's, hospitals, Independent Practice Organizations (IPOs), and pharmacies interact to exchange patient information for medical purposes.

The government requires that healthcare organizations report medical data for statistical analysis and so that the overall health of the nation can be assessed. Medical information is aggregated so that patient identifiers are omitted

Table 3. Sample medical Semantic Web projects

PROJECTS			
Name	***Purpose***	***Associated Organization***	***Source***
Good European Health Record Project	To produce a comprehensive multi-media data architecture for EHRs	CHIME	Nardon, 2004 http://www.chime.ucl.ac.uk/work-areas/ehrs/GEHR/index.htm
Brazilian National Health Card	Aimed at creating infrastructure for capture of encounter information at the point of care		Nardon, 2004
Artemis	Semantic Web Service-based P2P Infrastructure for the Interoperability of Medical Information Systems	Six participating entities from	Bicer et al., 2005 http://www.srdc.metu.edu.tr/webpage/projects/artemis/
Active Semantic Electronic Patient Record	Development of populated ontologies in the healthcare (specially cardiology practice) domain; an annotation tool for annotation of patient records, and decision support algorithms that support rule and ontology based checking/validation and evaluation.	LSDIS (large Scale Distributed Information Systems and AHC (Athens Heart Center)	http://lsdis.cs.uga.edu/projects/asdoc/
MedISeek	Allows users to describe, store, and retrieve medical images; metadata model		Carro et al., 2003

Figure 1. The coordination of the healthcare industry is very diverse in its information needs

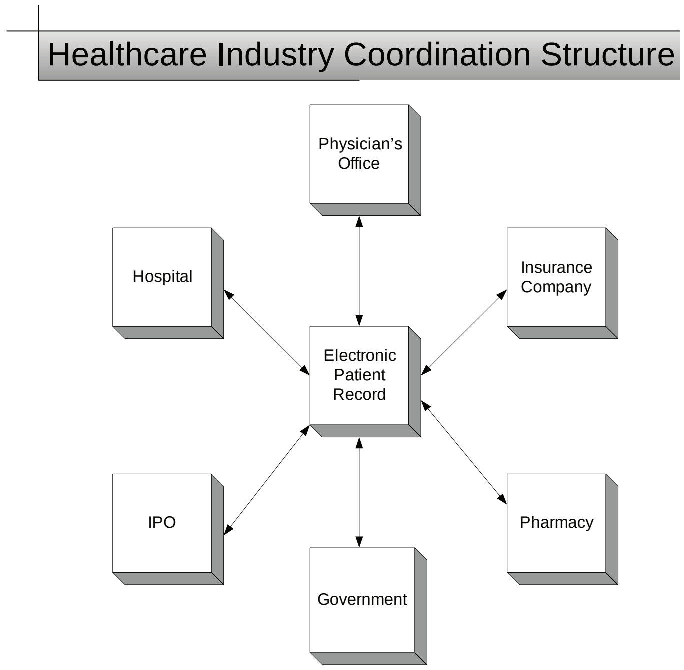
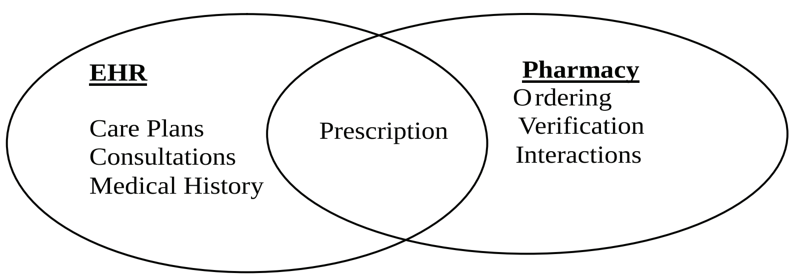


Figure 2. The sharing of information between healthcare entities can enable more efficient and effective quality of care



and reported to the government for public health purposes and to catch contagious outbreaks early as well as to determine current health issues and how they can be addressed. For example, cancer registries report specific aggregated cancer information, and healthcare organizations report instances of certain infectious diseases such as the Avian influenza (bird flu), for the welfare of the public. The importance of sharing this information is the improvement of patient safety, efficiency, self-health management (through access of medical information), and effective delivery of healthcare (HL7, 2005). Figure 2 shows how two entities may interact to share information (adapted from HL7).

Indeed, a commission on systemic interoperability has been established through the Medicare Modernization Act of 2003 and recommends product certification, interoperable standards, and standard vocabulary as a way of ensuring that healthcare data is readily accessible (Vijayan, 2005). At a North Carolina Healthcare Information Communications Alliance, one recurring theme was that of interoperable EHRs. Brailler (2005), the first National Health Information Technology Coordinator in the U.S., spoke about standards harmonization for EHRs. The discussion of developing standards for interoperability emphasized the need to "stitch together different efforts" put forth by organizations such as HL7, IEEE, ISO, and SNOMED. Undoubtedly, he recognized that "standards are about economic power" and they need to be analyzed to determine which standards are available for the commercial market. In doing so, the office of National Health Information Technology suggests that there be a compliance certification for EHR based on criteria such as security, interoperability, and clinical standards—basically a seal of approval that if a healthcare organization purchases such a product, it will be "guaranteed" to have specific interoperability certification. Brailler stated "if it's not certified, it's not an EHR." Given this, it has been suggested that the second generation of EHRs is being developed to communicate with structured datasets, middleware, and messaging between systems (Bernstein, Bruun-Rasmussen, Vingtoft, Andersen, & Nohr, 2005). Perhaps the third generation will provide full scale Semantic Web capabilities in which interoperability is seamless.

Currently, patient information is kept in silos across the aforementioned organizations; the Semantic Web will enable access to these silos through interoperability standards and consistent language. According to a white paper published by HL7 (2004), an organization which has developed HL7 standards for healthcare, improvements in the following five areas can be made through EHR standards: (1) interoperability, (2) safety/security, (3) quality/reliability, (4) efficiency/effectiveness, and (5) communication. To improve these areas, the standards proposed by HL7 include both standardized service interface models for interoperability, but also standardized concept models and terminologies. The current use of the HL7 standard is for the messaging of data to populate other disparate systems. For example, admissions data of a patient is also sent to the billing system. The problem with current messaging systems, such as HL7, is that they duplicate information across systems. Patient demographic information, for example, can be copied from one system to another, and maintenance of such data can create more messaging between systems (usually within an organization).

In Denmark, the examination of EHR use and interoperability has also been an issue of interest (Bernstein et al., 2005). The Danish Health IT Strategy project's goal is to analyze the variety of grassroots models for EHR information modeling and informatics. The National Board of Health is currently analyzing the SNOWMED ontology for use in its EHR. SNOWMED is an ontology that encapsulates classification systems such as ICD9. As a reference terminology, it is much more detailed in the medical concepts that it conveys. This level of detailed information allows the data to be used for quality assurance and resource utilization purposes and allows the EHR to relay more information than ICD9 coding for diagnoses. For example, there are around 13,000 ICD9 codes for diagnoses and SNOWMED contains 365,000 codes (Cassidy, 2005). Similar to the Denmark project, the Artemis project focuses on developing Semantic Web technology such as ontologies as a foundation to interoperability for medical records. Rather than standardizing the actual documents in the EHR, the goal is to standardize the accessibility of the records through wrappers, Web Services Description Language (WSDL) and Simple Object Access Protocol (SOAP) (Artemis, 2005). Bicer et al. (2005) discuss a project with Artemis in which OWL ontologies are used to map information messages from one entity to another.

Partners Healthcare uses RDF to enable medical history from EHRs to be accessible through computer models which select patients for clinical trials (Salamone, 2005). They utilized Semantic Web Rules Language (SWRL) to write decision support rules for this purpose. The advantage in using the Semantic Web approach is that the coding is concise, flexible, and works well with large databases. As Eric Neumann of the pharmaceutical company, Sanofi-Aventis suggests, "with the semantic web, you publish meaning, not just data" (Salamone, 2005).

Information Searching and Sharing

"Ontologies can enhance the functioning of the Web in many ways. They can be used in a simple fashion to improve the accuracy of Web searches" (Berners-Lee et al., 2001). The difficulties and complexities of searching for medical information are discussed by Pisnalli et al. (2004) in their research on medical polysemy. Because polysemy (a word having more than one meaning) can be critical to finding correct medical information, the application of ontologies can be of value in information searching. For example, the ontology of the term *inflammation* can vary depending on the context of its use. As Pisnalli et al. state, inflammation can include the size, shape, evolution, severity, and source. When one searches for the term inflammation, many results may be provided, but time is required to sort through the "hits" for relevance. The ON-9 ontology is utilized by Pisnalli et al. to map contexts for the term inflammation. As Nardon and Moura (2004) emphasize, the relationships among medical terminology is also essential to representation of the information in a logical format. Allowing for specific context to be interpreted through ontologies will enable more efficient and effective searching. Usually, this involves the creation of metadata to identify the relevant data elements and their relationships (Buttler et al., 2002).

Medical vocabularies used to represent data include the Unified Medical Language System (UMLS) from the U.S. National Library of Medicine and Arden Syntax. UMLS is perhaps the most frequently used ontology in the healthcare and medical sciences field. The purpose is to aid in integrating information from multiple biomedical information sources and enabling efficient and effective retrieval. It defines relationships between vocabularies and includes a categorization of concepts as well as the relationships among them. For example, the National Health Card System in Brazil contains an extensive knowledge base of 8 million patients in which complex queries can be run (Nardon & Moura, 2004). Through ontologies and UMLS, mapping of business rules can be applied to medical transactions to infer information and achieve semantic interoperability. For example, if a patient can undergo only a certain procedure once within a 30-day time period, a transaction for a patient setting up an appointment for that procedure can be mapped to business rules to infer that the same person cannot schedule the same procedure within that time period. UMLS would determine the ontology for the appointment and procedures and ensure that the patient is indeed the same, and RDF defines the business rules for sharing the information (Nardon & Moura, 2004).

When querying multiple medical data sources for research purposes, there are many medical science repositories in which data may not be in machine-processable format and stored in nonstandard ways. Most of the interfaces to search and retrieve medical sciences research require human interaction. Data extraction of such large data sources can be very complex and often the data is reused by researchers such as those in Genomics (Buttler et al., 2002). Large databases containing bioinformatics research can be unified through ontologies such as Riboweb, Generic

Human Disease Ontology, Gene Ontology (GO), TAMBIS, and LinkBase. These allow a standard vocabulary to exist over disparate ribosomal, disease, gene product, nucleic acid, and protein resources. As an example, the Generic Human Disease Ontology, currently being developed with information from the Mayo Clinic, allows a physician to search by symptom to determine the disease or for type of appropriate treatment, and researchers can search for possible causes of a disorder (Hadzic & Chang, 2005).

MedISeek is an interesting example of using semantic vocabularies to search for medical visual information, such as x-rays and other images (Carro et al., 2003). Biomedical Imaging Research Network (BIRN), a project of the National Institute of Health, examines human neurological disorders and their association with animal models. A significant aspect of their work is through brain imaging. Their goal is to make this information available to others through the Semantic Web via graphical search tools; standard identifiers through ontologies; and cross-referencing of imaging (Halle & Kikinis, 2004). The Semantic Web will enable BIRN, MedISeek, and other healthcare and medical science projects to filter out less appropriate data by searching for a context to the information. RDF is being utilized with MedISeek and BIRN to allow interoperability between metadata patterns.

CONCLUSION AND FUTURE TRENDS

Sharing of EHR information allows for improved quality of care for patients. Sharing medical science knowledge allows scientists to gather information and avoid redundant experiments. Searching for medical science information on the Semantic Web will be made more efficient and effective by the use of common ontologies and standards for transmissions. "Trusted databases exist, but their schemas are often poorly or not documented for outsiders, and explicit agreement about their contents is therefore rare." The opportunity to share such large amounts of information through the Semantic Web suggests that knowledge management can exist on a comprehensive level with ontology as a unifying resource (Hadzic & Chang, 2005).

While there has been some research in the area of medical sciences information searching on the Semantic Web, there have been few studies on how to better enable healthcare consumers to search for medical information on the Web. Lay terminology of consumers often increases the number of results returned when searching for medical information on the Web. Polysemy creates a multitude of results within which the consumer must further search. The goal should be to use Semantic Web technology to minimize the semantic distance between a search term and its polysemy of translations (Lorence & Spinks, 2004).

The future of the Semantic Web will involve important developments in the emergence of e-healthcare through the use of intelligent agents. Singh et al. (2005) suggest that emerging Semantic Web-based technologies offer means to allow seamless and transparent flow of semantically enriched information through ontologies, knowledge representation, and intelligent agents. Intelligent agents can enrich the information by interpretation on behalf of the user to perform an automated function. The example given at the beginning of this article in which someone queries for melanoma information and receives information regarding treatments, tests, and providers in that person's location which accept his insurance, shows how intelligent agents can be utilized to search the Semantic Web. Agents can also be utilized to verify the source of the information. When sharing of information occurs across the Web and is pulled automatically by agents, the source of the information needs to be verified. This is especially true in healthcare with Health Insurance Portability and Accountability Act (HIPAA) 1996 regulations. If the foundation of ontology and interoperable standards exists, intelligent agents will be able to search the Web for information within the context desired.

Legal issues associated with the dispersion of healthcare information need to be identified. With HIPAA (1996)), healthcare organizations are required to keep patient personally identifiable information secure and private. This means encryption, access control, audit trails, and data integrity must be insured in the transmission process (Jagannathan, 2001). Who has rights to the data and who "owns the data," particularly in EHRs? Similarly, there is an issue of trust involved with sharing medical science and healthcare data, and this is an area ripe for further research. How can authentication be provided so that others know the source of data is trusted and how can it be ensured that the data will be edited by a trusted entity? The area of e-commerce can be a foundation for future research in trust, as well.

Semantic Web technology can function as a foundation for the sharing and searching of information for the healthcare and medical sciences fields. Because of the intuitive nature of patient care, the Semantic Web will enable context and meaning to be applied to medical information, as well as the conveyance of relationships between data. With the generation of standards for transmission of data between disparate systems, the quality of healthcare through better research and the sharing of information between healthcare providers will be a critical step in the evolution of patient care. This will enable the third generation of EHRs to be seamlessly interoperable for more efficient and effective patient care. These innovations can lead to improved work satisfaction, patient satisfaction, and patient care (Eysenbach, 2003).

REFERENCES

Artemis. (2005). Retrieved November 2005, from www.srdc.metu.edu.tr/webpage/projects/artemis/home.html

Berners-Lee, T. (1989). *Proposal of Semantic Web to CERN*. Retrieved October, 2005 from http://www.w3.org/History/1989/proposal.html

Berners-Lee, T., Hendler, J., & Lassila, O. (2001, May 17). The Semantic Web. *The Scientific American*. Retrieved May 2005, from www.sciam.com

Bernstein, K., Bruun-Rasmussen, M., Vingtoft, S., Andersen, S. K., & Nohr, C. (2005). Modelling and implementing electronic health records in Denmark. *International Journal of Medical Informatics, 74*, 213-220.

Bicer, V., Laleci, G., Dogac, A., & Kabak, Y. (2005, September). Artemis message exchange framework: Semantic interoperability of exchanged messages in the healthcare domain. *SIGMOD Record, 34*(3), 71-76.

Brailler, D. (2005). Keynote address. *NCHICA 11th Annual Conference*, Greensboro, NC. Retrieved November 2005, from www.nchica.org

Buttler, D., Coleman, M., Critchlow, T., Fileto, R., Han, W., Pu, C., et al. (2002, December). Querying multiple bioinformatics information sources: Can Semantic Web research help? *SIGMOD Record, 31*(4), 59-64.

Carro, S., Scharcanski, J., & De Lima, J. (2003). MedISeek: A Web based diffusion system for medical visual information. Workshop on Web information and data management. *Proceedings of the 5th ACM international workshop on Web information and data management*, New Orleans, Louisiana, USA.

Cassidy, T. (2005, September 26). Building a bridge to EHRs: ICD-10 and SNOMED-CT. *Advance for Health Information Professionals.*

Decker, S., Melnik, S., et al. (2000, September/October). The Semantic Web: The roles of XML and RDF. *IEEE Internet Computing.*

Decker, S., Mitra, P., & Melnik, S. (2000, Novmeber/December). Framework for the Semantic Web: An RDF tutorial. *IEEE Internet Computing,* 68-73.

Eysenbach, G. (2003). The Semantic Web and healthcare consumers: A new challenge and opportunity on the horizon? *International Journal of Healthcare Technology and Management, 5*(3/4/5), 194-212.

Gillespie, G. (2003). NCVHS to extol a standard vocab. *Health Data Management, 11*(5), 50-58.

Gomez-Perez, A., Fernandez-Lopez, M., & Corcho, O. (2004). *Ontological engineering,* London: Springer.

Gruetter, R., & Eikemeier, C. (2004). *Applying the Semantic Web to clinical process.* Retrieved October 2005, from http://www.egms.de/en/meetings/gmds2004/04gmds316.shtml

Hadzic, M., & Chang, E. (2005). Ontology-based support for human disease study. *Proceedings of the 38th Hawaii International Conference on System Sciences.*

Halle, M., & Kikinis, R. (2004). Flexible frameworks for medical multimedia. *Proceedings of the 12th Annual ACM International Conference on Multimedia.* 768-775.

Health Level 7. (2004). HL7 HER system functional model: A major development towards consensus on electronic health record system functionality. White Paper.

Hooda, J. S., Dogdu, E., & Sunderraman, R. (2004). Health level-7 compliant clinical patient records system. *2004 ACM Symposium on Applied Computing,* 259-263.

Jagannathan, V. (2001, May/June). The careflow architecture: A case study in medical transcription. *IEEE Internet Computing,* 59-64.

Kishore, R., Sharman, R., & Ramesh, R. (2004). Computational ontologies and information Systems: I. Foundations. *Communications of the Association for Information Systems, 14,* 158-183.

Lee, Y., Patel, C., Chun, S. A., & Geller, J. (2004). Compositional knowledge management for medical services on Semantic Web. *The Thirteenth International World Wide Web Conference Proceedings,* New York. Retrieved August 2005, from http://www.www2004.org/proceedings/docs/2p498.pdf

Lorence, D., & Spinks, A. (2004). Semantics and the medical Web: A review of barriers and breakthroughs in effective healthcare query. *Health Information and Libraries Journal, 21,* 109-116.

McCormack, J. (2000, May). Standard Bearers cite progress. *Health Data Management, 8*(5). 98-102.

Miller, E. (2004, January). *Weaving meaning: An overview of the Semantic Web.* Paper presented at the University of Michigan, Ann Arbor, Michigan. Retrieved October 2005, from http://www.w3.org/2004/Talks/0120-semweb-umich/

Nardon, F., & Moura, L. (2004). Knowledge sharing and information integration in healthcare ontologies and deductive databases. *MEDINFO,* 62.

Pisanelli, D., Gangemi, A., Battaglia, M., & Catenacci, C. (2004). Coping with medical polysemy in the Semantic Web: The role of ontologies. *MEDINFO,* 416-419.

Salamone, S. (2005, June 8) Semantic Web interest grows. *Bio-IT World.* Retrieved November 2005, from http://www.bio-itworld/archive/microscope/document.2005-06-16.8341855754

Sharman, R., Kishore, R., & Ramesh, R. (2004). Computational ontologies and information systems: II. Formal specification. *Communications of the Association for Information Systems, 14*,184-205.

Singh, R., Iyer, L., & Salam, A. F. (2005, January-March). Semantic eBusiness. *International Journal on Semantic Web and Information Systems, 1*(1), 19-35.

Updegrove, A. (2005, June) *The Semantic Web: An interview with Tim Berners-Lee*. Retrieved November 2005, from http://www.TheConsotriumInfo.org

Vijayan, J. (2005, November 1). Federal commission looks to push e-health record adoption. *Computerworld*. Retrieved November 2005, from http://www.computerworld.com/securitytopics/security/story/0,10801,105860,00.html

This work was previously published in Semantic Web Technologies and E-Business: Toward the Integrated Virtual Organization and Business Process Automation , edited by A. Salam and J. Stevens, pp. 405-420, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 1.7
Discussing Health Issues on the Internet

Jane Moon
Monash University, Australia

INTRODUCTION

This article provides an overview of the trend in Internet usage; in particular, the trend that relates particularly to health-information-seeking behavior. It discusses a paradigm shift in patient-doctor relationships that has resulted from social changes; that is, lack of consultation time, thirst for medical knowledge, mass-media medical information and an explosion in the number of health Web sites. The Internet has become an important medium for bridging the gap in the patient-doctor relationship.

Issues of Internet quality are explored. While the Internet can help consumers by providing immediate feedback as far as treatment and medication are concerned, without proper standards and quality assurance it can give rise to diabolical consequences (Crocco, Villasis-Keever, & Jadad, 2002). Ciolek describes information on the Internet as mediocre and argues that health information on the Internet is subject to "Multi Media Mediocrity" (MMM) (Ciolek, 1997).

General Trends of Using Internet for Health Advice

The Internet has become a vital tool for individuals, families, the health profession and the health industry. One Web site reports that there are more than 10,000 health sites on the Internet, and others report more than 100,000 health-related Web sites (Eysenbach, Sa, & Diepgen, 1999). No one knows theexact number, but what is clear is that there are numerous health sites available.

Health sites vary, from academic sites to health-provider institutions and government sites. Recently, there have been an increasing number of pharmaceutical companies disseminating information or selling products and services in a variety of ways on Web sites luring consumers (Risk & Dzenowagis, 2001).

Since the emergence of the Internet in 1991, the Internet use has grown exponentially. A recent survey shows that 86% of the 168 million American adults have visited health Web sites, compared with 55% of the 60 million in Germany. Ninety percent of American primary-care physicians have used the Internet (Risk & Dzenowagis, 2001). According to Harris Interactive consulting firm, health Internet users grew steadily from 50 million in 1998 to 69 million in 1999, 97 million in 2001 and 110 million in 2002 (Harris Interactive, 2000).

Demographical Difference

No significant difference in information-seeking habits between different age groups were found (Brodie, Flournay, Altman, Blendon, Benson, & Rosenbaum, 2000) Also, there is a direct correlation between computer usage and access to health information:

Once people gain access to the Internet, its use at home to get health information is similar across income, education, race and age. Therefore, the number of persons using the Internet to access health information should rise along with computer use. (Brodie, 2000, p.262)

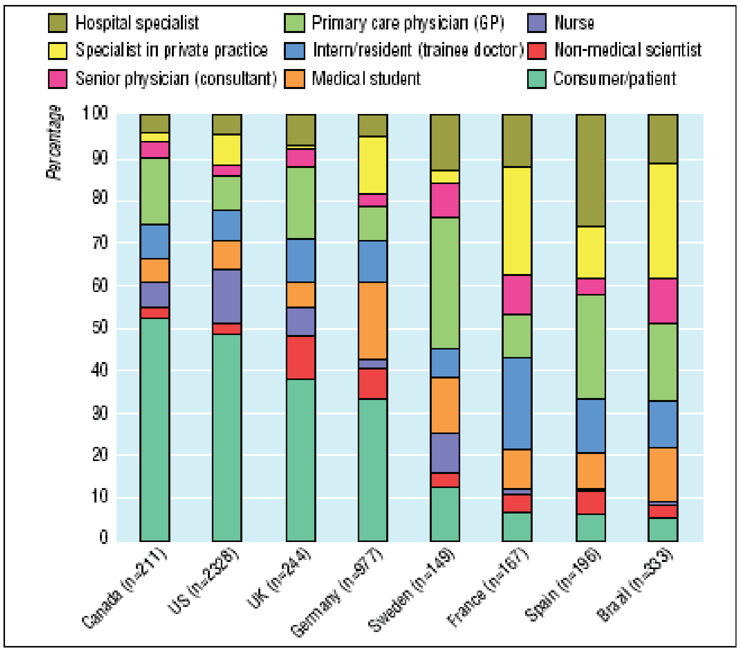
According to Brodie's report, gender, age and background do not make much difference in Internet search behavior.

A significant difference between countries is noted in research results from a questionnaire survey among the users of a dermatology atlas Web site. Pictures were used to minimize language barriers between countries (Eysenbach et al., 1999). The survey was conducted over seven months, from July 1998 to February 1999, and was answered by 6,441 users from all over the world.

Figure 1 shows the distribution of the 4,605 users who completed survey from the eight countries that showed the highest absolute numbers of users. Of those eight, Canada shows the highest percentage of users, followed by the United States (U.S.) and United Kingdom (UK), with Brazil the lowest.

There was a high proportion of general practitioners in Canada, UK, Spain and Sweden, and a high proportion of specialists in Brazil, France, Germany, Spain and U.S.

Figure 1. User profile of dermatology atlas Web site intended for health professional



A high proportion of hospital specialists were noted in Spain, Sweden, France and Brazil compared to a large number of specialists in private practice in Brazil, France and Germany. This is a reflection of the actual ratio of specialists in private practice compared with the public sector in the respective countries.

Looking at the percentage of nurses visiting the Internet, the highest percentage was noted in the U.S., Canada, UK and Sweden, as opposed to almost none in other countries, probably reflecting the differing roles and levels of responsibility nurses have in those countries.

While these data have been gathered in a specialist setting; that is, dermatology, and may be prone to self-selection bias and thus not representative of the whole, nevertheless, they tell us that there are significant differences between countries in Internet use. These differences are not only technological ones, but they also reflect differences in the health system as well as other cultural, sociological factors and economic factors—for example, capacity to afford Internet facilities (Eysenbach et al., 1999).

What causes the behavioral shift in consumers towards the Internet over traditional face-to-face contact with doctors?

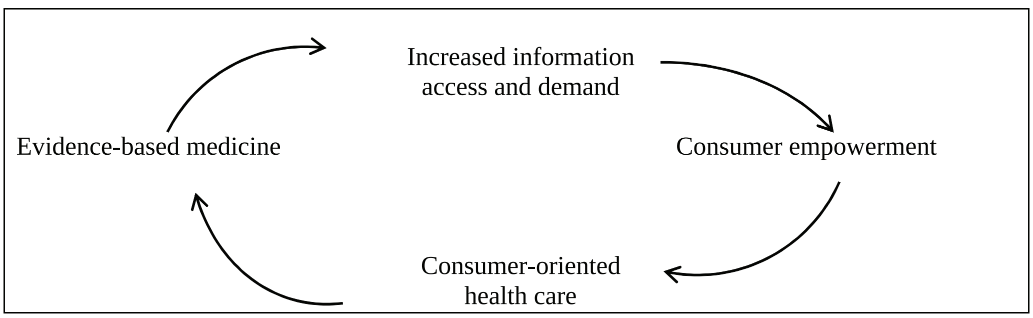
Changing Medical Practices

In the field of information technology, doctors still seem to be lagging behind other professionals. In many industrialized nations, consumers have taken a leading role in retrieving and exchanging health information (Eysenbach et al., 1999). While telemedicine (diagnostic and curative medicine) is influenced by "technological push," cybermedicine (preventive medicine and public health) is influenced by "consumer pull"; the traditional family doctor-patient relationship has to change to meet the demands of these social changes (Eysenbach et al., 1999).

Consumer Empowerment

The availability of the Internet has provided consumers with easy access to information. "The empowered, computer-literate public is exerting tremendous influence on healthcare delivery" (Ball & Lillis, 2001, p. 2). Consumers are actively seeking health information using the Internet as a major tool. Users get access to as many of the resources and databases as physicians do, although some databases are specifically available for physicians. This empowers consumers to make informed decisions and reduces their dependency on the physician. The Internet is, thus, a motor for evidence-based medicine for both physicians and consumers. It gives "increasing consumer involvement in health care decision making and increasing the pressure on caregivers to deliver high quality health services" (Eysenbach et al., 1999, p.2).

Figure 2. Positive feedback loop (Eysenbach et al., 1999)



As can be seen in Figure 2, consumers' access to health information on the Internet has increased tremendously; they have access to physicians' Web pages and databases and retrieve a wealth of knowledge on medical conditions and treatments. As computer-literate consumers become empowered, they can become involved in decision making in evidence-based medicine (Eysenbach et al., 1999).

This transformation driven by the Internet can be a challenge to physicians, as users are better informed than before (Ball & Lillis, 2001). Whether consumers are capable of making quality decisions is yet to be determined, as they can become "cyberchondriac" with more access to information on health problems. However, many consumers can challenge physicians with evidence-based information from the Internet (Eysenbach et al., 1999). The speed at which information is delivered on the Internet can cause mis-communication; for example, in Britain there was a government warning about oral contraceptives on the Internet. Some consumers found the information before physicians did (Coiera, 1996). There is a pressing need for better communication among doctors regarding information technology. Equipped with the right information, physicians can improve the quality of treatment by fostering closer partnerships with consumers.

This paradigm shift is a challenge to the doctor-patient relationship, as can be seen next.

Doctor-Patient Relationship

The term "patients" is slowly being replaced by "consumers" (Coiera, 1996). Sir William Osler (1849-1919) was a pre-eminent physician of the century whose work was based on trust and confidence, fostering a personal relationship. With the recent explosion of the Internet, the art of Osler's medical practice, which has been a fundamental to medicine, is under threat (Wheeler, 1990), as consumers challenge physicians with a "stew" of medical information they have retrieved from the Internet, and arrive at their appointments with meters of printouts challenging their knowledge (Coiera, 1996).

There is a common scenario where patients spend long hours searching for information on the Internet and bring pages of printouts to the doctor and ask questions. Such discussion can be exhausting. Some doctors have suggested, "Whatever you do, don't go on the Internet" (Ferguson, 2002, p.555). Also, the information available could be too abstruse or complex for most consumers. This health illiteracy could pose significant concerns for consumers, as they do not necessarily have sufficient medical knowledge to make informed decisions about their medical conditions (Oermann & Wilson, 2000; Wyatt, 1997).

In addition, there is an increase in litigation against doctors. The Internet transforms the traditional autocratic doctor-patient relationship into a balanced power play (Ball & Lillis, 2001).

What is very important is that, while information obtained on the Internet does support users, it cannot "replace the patient-physician relationship" (McLellan, 1998). The availability of resources to both physicians and patients could mean that consumers can get the same access to resources as the physicians. But it is often questionable whether the information is accurate or complete, which could lead to misdiagnosis and misunderstanding (Helwig, Lovelle, Guse, & Gottlieb, 1999).

National Medical Portal: Opportunities and Threats

A portal is a concept widely used within the Web. It is "a gateway or a door that provides users a single gateway to personalized information needed to make informed business decisions" (Quirk, 2001, p.2). The word "portal" has been used largely in the last two to three years and is often confused with Web sites. The difference between Web sites and portals is that the former is static and the latter is dynamic. Portals are Web-based; have dynamic links to information resources, effective information and document managers; and connect people with information (Moon & Burstein, 2004).

Medical portals provide information such as causes of diseases, medications, treatments, alternative therapies and lists of consultants. Medical portals can be either general or specific. General medical portals provide general medical information; for example, Better*Health* or Health*Insite*. Disease-specific portals provide information specific to particular disease; for example, cancer portal, breast cancer portal.

Moon and Burstein (2004) described five functional components necessary for a model of medical portal: community, personalization, quality, health services and gateway. The functional components relating to community are discussed in detail next, as it is pertinent to this article (p. 277):

- **Share medical information:** Can be done via e-mail, or by "chatting" or any other facilities such as MUD (multi-user device—through role play)
- **Promote interaction:** With doctors for receiving medical help or sharing personal experiences with other people
- **Bulletin board/newsletter:** From various stake holders can inform users of changes in the organization, such as change of member of the boards, or changes to portal sites for updates and so forth
- **Alert to new information:** With personalized settings, users can be alerted to new drugs or new research findings to a particular aspect of medical conditions
- Multimedia access via video streaming, or Web cam facilities to facilitate meetings or discussions.

Opportunities

The opportunities for users that give better empowerment are as follows:

- Better access to medical information
- New form of communication between users and care providers
- Opportunity to meet other users
- Saves time from visiting doctors
- Cheaper than visiting the doctor, as the connection fee is fairly inexpensive compared to the cost of visiting general practioners or consultants
- Not time dependent—can look for information anytime
- No geographical barrier—can find information anywhere
- New opportunity for therapy (e-therapy)
- Reduces barrier between care providers and users for second opinion
- A new medium for those who are embarrassed to see doctors for personal reasons and can find answers anonymously.

All of the above and more that can be listed add value to users and empower them to make informed decisions about their health (van Melick, Hakkenberg van Gaasbeek, & Pennings, 2001).

Threats

The Internet is anarchic in nature; that is to say that it is formless and boundless. Anyone, anywhere can write content and publish it anywhere, provided they have the software to do so. The question is not how to retrieve health information but how to retrieve relevant and accurate information (Risk & Dzenowagis, 2001). The easy access to an avalanche of health information can be an enormous threat to users if they are not properly guided.

Following is a list of some adverse effects the Internet could bring if the information is not managed with proper knowledge:

- It is not always easy to see the origin of the information, and users could be making decisions on the basis of a source that might not be quality assured

- Users are not equipped to make judgments on the basis of particular scenarios, they need to see health holistically
- Could harm users if they follow treatments that are location-specific
- Risk of making a wrong diagnosis
- Risk of taking ill advice if the discussant is not properly qualified
- Risk of Internet addiction
- Could be lured by pharmaceutical companies' advertisements
- Avalanche of information can be time-consuming and confusing.

All of the above and more add value to the users and empowers them to make informed decisions about their health (van Melick, Hakkenberg van Gaasbeek, & Pennings, 2001).

The following paragraphs discuss Internet health tools available to give guidance to consumers about the breath, depth and insightfulness of medical portals.

Internet Quality Tools

Health care information has potential benefits for many Internet users (Coiera, 1996) if they are properly educated and if the developers take quality criteria into consideration (Policy Paper, 2002). How do consumers know if a site is worthy? Many tools are available to assess the quality of content on the Internet. Wilson (2002) and Risk and Dzenowagis (2001) provide tools to classify Web sites, assisting developers to produce quality sites. They are as follows:

- **Code of conduct:** to ensure that the developers adhere to quality criteria.
- **Quality labels:** a logo or a symbol is displayed on the screen.

The most common ones are Health On the Net Foundation and Hi-Ethics code. Both of these are used on more than 3,000 Web sites.

Figure 3.

Figure 4.

Hon-code does not rate the quality of information, but it provides users the origin information and the purpose of data they are reading. It also provides ethical standards to Web site developers.

Hi-Ethics code produces quality levels for commercial sites. The uses of Hi-Ethics are: American Specialty, Health Networks, America Online HEALTHvision.

- **User guides:** provide user guidance to assess if a site meets the criteria. Web sites such as DISCERN allow users to validate the quality of written information on treatments; QUICK gives step-by-step guidance to children; NETSCORING gives guidance on all health-related topics.
- **Filters:** filter information according to a set of criteria; OMNI provides gateways to evaluate resources.
- **Third-party certification:** a seal or logo accredited by a third party, assuring consumers the site meets accepted standards for health sites. Third-party accreditors are MEDCERTAIN and TNOQMIC.

However, Wilson fails to discuss the size of the burden placed on provders, the lack of provision for educating consumers, the cost of developing and maintaining quality and the needs of developing countries. Wilson's article concentrates mostly on English-speaking countries.

FUTURE TRENDS

There is no turning back to the past. Twenty-first-century medicine has changed traditional medicine to Internet-empowered, patient-driven online support, with users controlling and managing their health (Ferguson, 2002). As more and more consumers are using the computer, it would be impossible to unwire the information. While the Internet can pose the threat of unreliable information (Oermann & Wilson, 2000), with users lured by many commercial sectors (Risk & Dzenowagis, 2001), nevertheless, it is satisfying many consumers' needs (Ferguson, 2002).

Doctors who do not recognize this trend would be losing patients to those who do understand the changes in the doctor-patient relationship. What is important is that there should be a synergy between doctors and patients (Yellowlees, 2000). The way to ensure this synergy exists, and to build trust between doctors and patients, is to make sure Web site content is reliable and that the Internet is serviceable (accessible both technically and design-wise). Doctors also need to understand consumers' needs and provide them with quality information. Given that doctors are under pressure to see more patients in a given time, leading to shorter consultation time, the Internet is an alternative for consumers to educate themselves on their health concerns.

One of the ways to ensure that consumers educate themselves is to educate the developers of portals' search engines, encouraging them to design portals that are ethical and compliant with quality standards. It is also vital to educate consumers to be cautious with the material they access on the Internet. Thus, consumers can "access health information critically" (Wilson, 2002, p.600).

CONCLUSION

The Internet presents a powerful mechanism for helping users to improve their health-care decision making by providing easy and rapid access to, and exchange and dissemination of, enormous amounts of health information. Yet, users must be aware of the potential for misinformation and recognize the critical need to assess the quality of the information provided. Content providers must be encouraged to develop and publish high-quality information, and policy makers and health-care professionals must be educated on this important health issue (Risk & Petersen, 2002).

Despite the perils of predicting the future of our health-care system, it is hard to ignore the evidence that the health system will undergo a revolution. The rapid growth of computer-based electronic communication and the fact that the new generations are comfortable with electronic transfer of health information leads to an increase of patient health management; that is, patients are taking an active role in the decision-making process (Kassirer, 1995). This trend is likely to grow and will have a substantial and hopefully positive impact (lowering the cost) on the health economy and on the well-being of consumers.

REFERENCES

Ball, M., & Lillis, J. (2001). E-health: Transforming the physician/patient relationship. *International Journal of Medical Informatics*, *61*, 1-19.

Brodie, M., Flournay, R.E., Altman, D.E., Blendon, R.J., Benson, J.M., & Rosenbaum, M.D. (2000). Health information, the Internet and the digital divide. *Health Affairs*, *19*(6), 255-265.

Ciolek, M.T. (1997). Today's WWW, tomorrow's MMM: The specter of multi-media mediocrity. *Educom Review*, *32*, May/June.

Coiera, E. (1996). The Internet's challenge to health care provision. *British Medical Journal*, *312*, 7022-7024.

Crocco, A.G., Villasis-Keever, M., & Jadad, A.R. (2002). Two wrongs don't make a right: Harm aggravated by inaccurate information on the Internet. *Pediatrics, 109*, 522-523.

Eysenbach, G., Sa, E.R., & Diepgen, T.L. (1999). Shopping around the Internet today and tomorrow: Towards the millennium of cybermedicine. *British Medical Journal, 319*, 1294.

Ferguson, T. (2002). From patients to end-users. *British Medical Journal, 324,* 555-556

Harris Interactive. (2000). *Healthcare satisfaction study.* Retrieved March 18, 2004, from *www.harrisinteractive.com/news/downloads/HarrisAriaHCSatRpt.PDF*

Helwig, A.L., Lovelle, A., Guse, C.E., & Gottlieb, M.S. (1999). An office-based Internet patient education system: A pilot study. *The Journal of Family Practice*, *48*, 123-127.

Kassirer, J. (2000). Patients, physicians, and the Internet. *Health Affairs, Nov/Dec*, 115-122.

McLellan, F. (1998). Like hunger, like thirst: Patients, journals, and the Internet. *Lancet*, *352*(2S), 39SII-43SII.

Moon, J., & Burstein, F. (2004). Intelligent portals for supporting medical information needs. In A. Tatnall (Ed.), *Web portals: The new gateways to internet information and services.* Hershey, PA: Idea Group Publishing.

Oermann, M.H., & Wilson, F.L. (2000). Quality of care information for consumers on the Internet. *Journal of Nursing Care Quality, 14*(4), 45-54.

Policy Paper, Europe. (2002). Quality criteria for health related Web sites. Commission of the European Communities. *Journal of Medical Internet Research*, *4*(3), e15. Retrieved December, 2002, from *http://jmir.com*

Quirk, M.J. (2001). *A guide to defining an enterprise portal project.* Retrieved October 19, 2002, from *www.OpenDoorTechnologies.com*

Risk, A., & Dzenowagis, J. (2001). Review of Internet health information quality initiatives. *Journal of Medical Internet Research*, *3*(4), e28.

Risk, A., & Petersen, C. (2002). Health information on the Internet: Quality issues and international initiatives. *The Journal of the American Medical Association*, *287*, 2713-2715.

van Melick, R.G.M., Hakkenberg van Gaasbeek, C.H.A., & Pennings, L.J.(2001). Gezondzoeken.nl Management model for trustworthy medical information and communication. Netherlands Organizations for Applied Scientific Research (TNO). TNO Prevention and Health.

Wheeler, H.B. (1990). Shotkick Lecture: Healing and heroism. *New England Journal of Medicine, 322*, 1540-1548.

Wilson, P. (2002). How to find the good and avoid the bad or ugly: A short guide to tools for rating quality of health information on the Internet. *British Medical Journal*, *324*, 598-602.

Yellowlees, P.M. (2000). Healthcare on the Internet: Buyers beware. *The Medical Journal of Australia*, *173*, 629-630.

KEY TERMS

Consumer Empowerment: The users are well informed of the information thus making an informed decision.

Cyberchondriac: Those users of the Internet who are obsessively persuing the Internet in search of information.

Cybermedicine: The Internet-driven practice of medicine where patients communicate with doctors via electronic medium.

Evidence-Based Medicine: Evidence-based medicine is a new paradigm that replaces old paradigm of traditional medicine based on authority. It is based on clinical research which uses current best evidence in making decisions about the care of individual patients.

Health Portals: Includes any information on the Internet relating to health, including medical information as well as products and information that are related to well being, i.e., shampoo, diets, medicine, and alternative medicine.

This work was previously published in Encyclopedia of Virtual Communities and Technologies, edited by S. Dasgupta, pp. 133-139, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.8
Networkcentric Healthcare and the Entry Point into the Network

Dag von Lubitz
Central Michigan University, USA

Nilmini Wickramasinghe
Illinois Institute of Technology, USA

INTRODUCTION

The concept of e-health gains rapid and wide-spread international acceptance as the most practical means of reducing burgeoning healthcare costs, improving healthcare delivery, and reducing medical errors. However, due to profit-maximizing forces controlling healthcare, the majority of e-based systems are characterized by non-existent or marginal compatibility leading to platform-centricity that is, a large number of individual information platforms incapable of integrated, collaborative functions. While such systems provide excellent service within limited range healthcare operations (such as hospital groups, insurance companies, or local healthcare delivery services), chaos exists at the level of nationwide or international activities. As a result, despite intense efforts, introduction of e-health doctrine has minimal impact on reduction of healthcare costs. Based on their previous work, the authors present the doctrine of *network-centric healthcare operations* that assures unimpeded flow and dissemination of fully compatible, high quality, and operation-relevant healthcare information and knowledge within the Worldwide Healthcare Information Grid (WHIG). In similarity to network-centric concepts developed and used by the armed forces of several nations, practical implementation of WHIG, consisting of interconnected entry portals, nodes, and telecommunication infrastructure, will result in enhanced administrative efficiency, better resource allocation, higher responsiveness to healthcare crises, and—most importantly—improved delivery of healthcare services worldwide.

BACKGROUND: CURRENT ISSUES OF E-HEALTHCARE

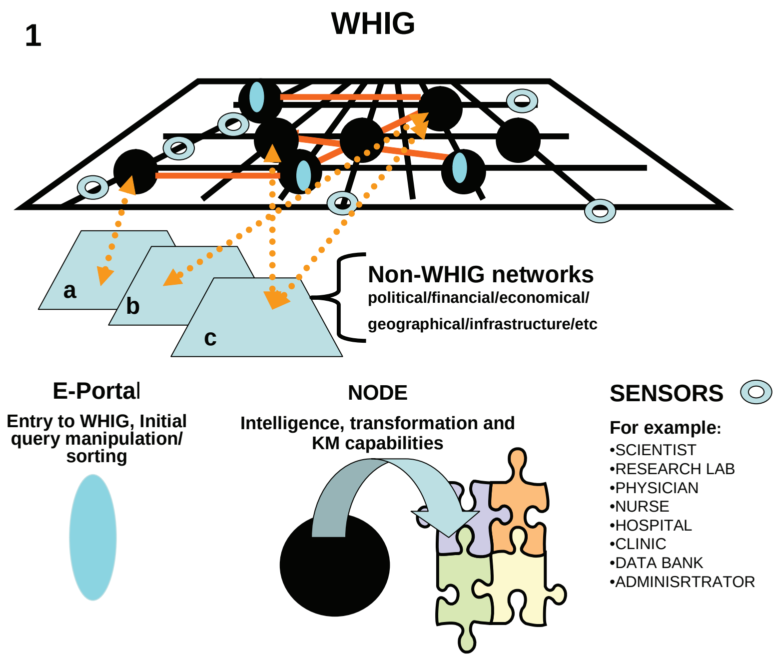
Major shifts in political and economical structure of the world that took place in the 20th century were instrumental in focusing global attention on healthcare and its importance in maintaining stability and growth of nations. At the same time, the cost and complexities of national and global healthcare operations became increasingly apparent (World Health Organization Report, 2000, 2004). In order to be efficient, healthcare providers and administrators became progressively more dependent on a broad range of information and knowledge that spans the spectrum stretching from purely clinical facts to the characteristics of local economies, politics, or geography. Consequent to the elevating demand for knowledge is the flood of a wide variety of uncoordinated data and information that emerges from multiple and equally uncoordinated sources (von Lubitz & Wickramasinghe, 2005b, 2005c). It has been hoped that vigorous use of IC^2T (Information/Computer/Communications Technology) will, in similarity to some forms of business operations, obviate the growing chaos of global healthcare. While IC^2T changed many aspects of medicine, the explosive growth of worldwide healthcare costs indicates that a mere introduction of advanced technology does not solve the problem (Fernandez, 2002: von Lubitz & Wickramasinghe, 2005). The quest for financial rewards provided by the lucrative healthcare markets of the Western world led to a plethora of dissonant healthcare platforms (e.g., electronic health records) that operate well within circumscribed (regional) networks but fail to provide a unified national or international service (Banjeri, 2004; Olutimayin, 2002; Onen, 2004). There is a striking lack of standards that would permit seamless interaction or even fusion of nonhealthcare (e.g., economy or local politics) and healthcare knowledge creation and management resources. The "inward" concentration of the Western societies on their own issues causes progressive growth of technology barriers between the West and the less developed countries, while the essentially philanthropic efforts to address massive healthcare problems of the latter continues to concentrate on "pretechnological" and often strikingly inefficient approaches (Banjeri, 2004; Olutimayin, 2002). Thus, despite the massive amount of information that is available to healthcare providers and administrators, despite availability of technologies that, theoretically at least, should act as facilitators and disseminators, the practical side of access to, and the use and administration of healthcare are characterized by increasing disparity, cost, and burgeoning chaos (Larson, 2004). Solutions to many of these acute and disturbing problems may be found in the recent approach chosen by the defence establishments of many countries to the information needs of the battlefield and to the modern, highly dynamic combat operations (von Lubitz & Wickramasinghe, 2005a).

DOCTRINE OF NETWORK-CENTRIC HEALTHCARE OPERATIONS

Our previous publications (von Lubitz & Wickramasinghe, 2005a, 2005b, 2005c) discussed the general principles and applicability of the military network-centric operations concept and its adaptation to modern worldwide healthcare activities. Network-centric healthcare operations are physically facilitated by the World Healthcare Information Grid (WHIG)—a multidimensional communications network connecting primary information collecting sources (sensors) with information processing, manipulating, and disseminating nodes. The nodes also serve as knowledge gathering, transforming, generating, and disseminating centres (Figure 1).

In similarity to the already proved attributes of network-centric military operations (Cebrowski & Garstka, 1998) of which, at the simplest level, the command centre of a joint naval task force is the simplest example and the execution of Operation Iraqi Freedom probably the most complex one, healthcare activities are characterized by multidirectional and unrestricted flow of multispectral data (von Lubitz & Wickramasinghe, 2005b, 2005c). All data, information, and node generated knowledge are characterized by fully compat-

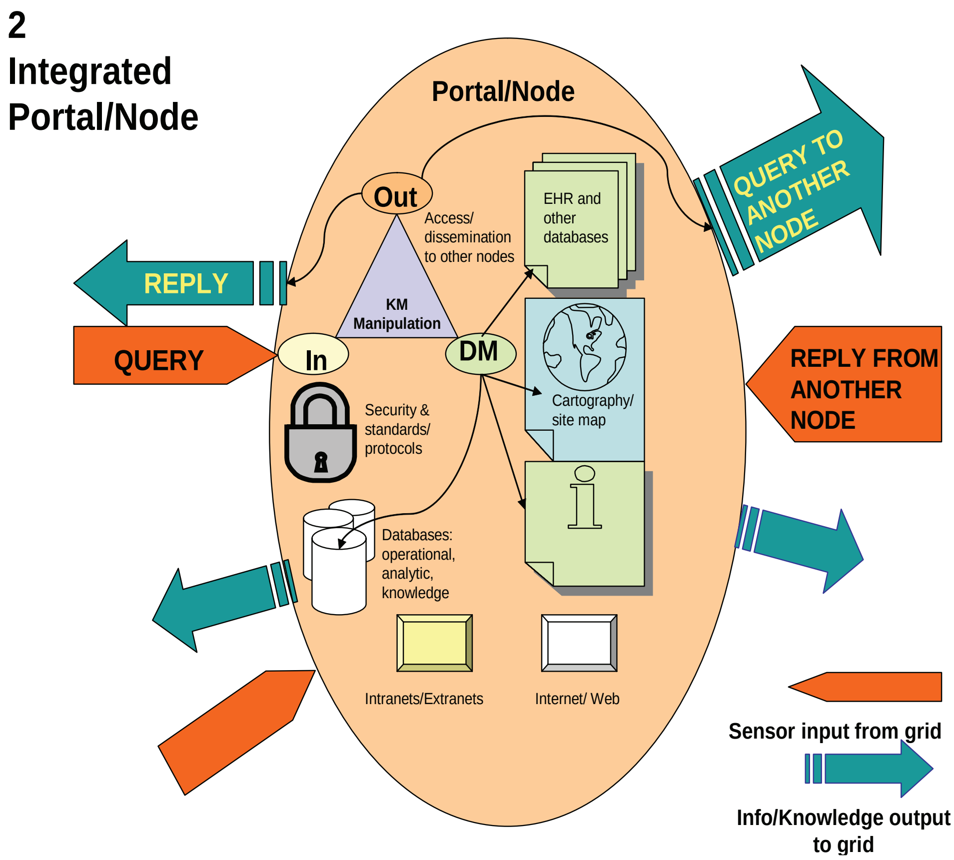
Figure 1. Schematic diagram of a WHIG segment



ible formats and standards that allow automated meshing, manipulation, and reconfiguration. Essentially, network-centric healthcare operations are based on the principles of high order network computing, where the WHIG serves as a rapid distribution system, and the nodes as the sophisticated processing centres that function not only as data/information/knowledge generating elements but also as DSS/ESS platforms providing high level, query-sensitive networkwide outputs. The nodes are also capable of extracting and analyzing data and information from healthcare-relevant sensors and electronic data sources (e.g., financial, political, military, geological, law enforcement, infrastructure level, etc.) and mesh these with the relevant biomedical elements. Incorporation of external information in healthcare operations provides readily available, rich, and necessary background that has, typically, a highly significant bearing on the success of activities that are either planned or conducted within the strict healthcare domain. The complications resulting either from the failure to include elements external to the essential healthcare activities or consequent to the exclusion caused by incompatible resource platforms have been amply demonstrated by major difficulties encountered during relief operations following tsunami-mediated destruction in December 2004.

Sensors feed raw data/information into the network through network-distributed portals. Likewise, data, information, and knowledge queries enter through portals as well. The latter provide entry level security screening and sorting/routing. Subsequent manipulation, classification, and transformation into information/pertinent knowledge is executed by interconnected nodes. Whenever required, each node can access information/knowledge existing within non-WHIG networks and databases and compare/merge the contents with the contents existing within the WHIG. While portals are associated with the nodes, implementation of ASP philosophy allows reaching the portal from anywhere within the WHIG.

Figure 2. Integrated entry portal/node



In addition to functioning as data/information/knowledge generating/manipulating/disseminating centres, the nodes also serve as the network points of entry (entry portals, Figure 2). However, contrary to the classical Web portal, where the client determines the information gathering path (O'Brien, 2004), the WHIG portal provides automated query classification, direction, and integration functions. Its operations are fuzzy logic-based, and the principal function of the WHIG portal is that of a "sorting/distribution station" which distributes the original query throughout the entire WHIG and collects and weighs the relevant outputs generated by multinodal analysis of the available resources. As the final step, the portal assigns the relevance level of the cumulative output, and provides automated pathways toward its further refinement. The WHIG portal operates thus not only as an entry point but also as either redirection station or WHIG exit site. Some of the functions of the WHIG portal are exemplified by the response to a hypothetical NGO query requiring decision support on the conduct of healthcare activities within the scope of a humanitarian relief operation in a costal region of "State X." The query will be automatically distributed within the network and the response will (equally automatically) provide multifaceted analysis of the essential medical needs of the affected population (e.g., most threatening diseases, the type and quantity of the required vaccines, need for other pharmaceuticals, tenting, water supplies, etc.). However, the response will also provide information on the local infrastructure and its nature and quality (e.g., air/sea port off loading/storage capacity, availability of beaches as the off-loading sites, capacity of local healthcare human and physical resources, quality and distribution density of roads/railways/means of transport, etc.), whether as an adverse factor, political stability/law enforcement efficiency within the region as a factor influencing distribution of aid, or movement of support teams. Clearly, even within such a simplified example, the range and complexity of factors that may significantly (and adversely) affect only one of many critical elements within a major relief operation is strikingly large. Correspondingly, the need for germane information/knowledge is equally substantial. Yet, due to the prevailing platform-centricity, despite the existence of such information, its dispersal within several, largely incompatible, systems makes it essentially inaccessible. Moreover, its retrieval demands clear awareness of the need followed by human-based/human guided search and extraction. Consequently, in situations of stress or in environments that pose acute demand for a wide range of simultaneous responses, the potential for major errors of omission and commission increases dramatically. A classical chain of such errors can be seen, for example, in the response to the events immediately preceding the destruction of World Trade Center in September 2001 (National Commission on Terrorist Attacks on the United States, 2004).

Data, information, or queries from WHIG enter through the portal where they are subjected to security/standards/protocol screening then transfer to the manipulation site (DM). The latter provides detailed sorting and redirection via intra and extra nets, and/or Internet/Web to other locations within the node, for example, patient records, information storage sites, analysis and knowledge generating sites, and so forth (unidirectional arrows). All sites within the node are capable of multidirectional communication (not indicated for the sake of clarity). Their output is transmitted to the knowledge manipulation and generation site which, in turn, generates final output stored within the node and also disseminated throughout the network (Out). If needed, the node can distribute additional WHIG-wide queries. Replies are collected, manipulated at the KM level, and incorporated into the final node output. Although neither the portal nor individual functional aspects of the node need be collocated, their operations are conducted as a single, self-contained unit; that is, none of the constituting elements can participate individually in the functions of another node. Self-containment of each node adds to its security and reduces the risk of inadvertent networkwide dissemination of integrity-compromising factors (e.g., viruses, spurious data, etc.).

FUTURE TRENDS: OPERATIONAL THEORY OF NETWORK-CENTRIC ACTIVITIES

The operational philosophy of network-centric healthcare operations is based on the principles of Boyd's (OODA) Loop (Boyd, 1987; von Lubitz & Wickramasinghe, 2005a, 2005b, 2005c) that defines the nature and the sequence of interactions with dynamic, rapidly changing environments characterized by a high degree of structural and event complexity. Accordingly to Boyd, each complex action can be subdivided into a series of consecutive cycles, loops, with the preceding cycle strongly influencing the initial stages of the following. Each revolution (cycle) of the Loop comprises four stages: observation, orientation, determination, and action. During the observation stage, all inputs describing the action environment are collected and organized into coherent entities. At the orientation stage, the organized data are converted into meaningful information that provides as complete image of the operational environment as possible based on the totality of the existing information. At this stage the weaknesses of the opposition are detected, and the centre of the future action determined. During the determination phase, the hypothesis, that is, the plan to respond to the pressure exercised by the operation environment, is formulated. The Hypothesis defines the plan of action, the required strength and nature of the response, its precise location, timing and duration, and so forth. During the Action phase, the Hypothesis is tested: the formulated plan is implemented and its results (and the consequent response of the action environment/opposition) set off the next revolution of the Loop—the new observation stage is initiated. Clearly, the nature of action determines the intervals between the stages.

Originally Boyd's Loop had been created as a tool facilitating aerial combat, where each individual stage was extremely brief (milliseconds). Nonetheless, the principles of the Loop can be applied to virtually any rapidly evolving environment. Moreover, Boyd's Loop helps to understand the critical role of the mistakes made during the initial data collection (e.g., selective or biased selection, rejection of *non-conforming* data as necessarily false, etc.) at the observation stage and their subsequent analysis (subjective analysis based on preconceived notions, influence of personal bias, inflexibility, etc.) at the orientation stage.

Errors made at these two stages influence the following two. Thus, at each subsequent cycle, error correction demands increasingly larger resources and removes them from where they should be otherwise committed—at the centre of action. Uncorrected errors compound at each new revolution of the Loop and exponentially increase the chance of failure. Probably the best example of *Loop failure* was the disastrous response of state and federal authorities to Hurricane Katrina in August 2005, while the response to Hurricane Wilma (its shortcomings notwithstanding) shows how application of Boyd's Loop-based thinking can lead to positive outcomes in situations demanding flexible, ongoing, and dynamic response to the continuously but unpredictably changing operational environment.

Clearly, to assure efficiency of action, the interval separating each individual stage of the Loop must be as short as possible, particularly when interacting with highly fluid, ultracomplex systems such as military or healthcare information. Here, the demand is not only on rapid, reliable sampling of the environment but also on a very high degree of automation at the level of multi-source data collection, analysis, manipulation, and classification into larger information/germane knowledge entities.

Contrary to the prevalent platform-centric operations, network-centricity allows vast increase in sampling speed, range, and data manipulation speed. Consequently, decision supporting outputs of the network are faster, more situation/operational environment-relevant and, most importantly, allow robustly elevated rate of stimulus-response cycle (operations "inside the Loop"). Moreover, by increasing reaction relevance and speed, network-centric operations facilitate goal-oriented manipulation of the operational environment and also increase both the level (accuracy) and predictive range of responses to environment induced pressures. Military benefits of such operations have been frequently demonstrated. However, the acceptance of Boyd's (OODA) Loop principles in the civilian world (e.g., global financial/banking operations, lean manufacturing, just-in-time supply chains, etc.) led to demonstrable gains in efficiency and productivity as well.

CONCLUSION

The preceding description is, of necessity, vastly simplified. Yet, the existence and highly efficient use of the network-centric approach to military operations has already resulted in the significant enhancement of the C^3I (Command, Control, Communications, and Intelligence) concept (Alberts, Garstka & Stein, 2000; Department of Defense, 2001). The most palpable consequences of network-centricity in warfare are increased efficiency in the use of available resources, application of resources appropriate to the operational environment, reduction of casualties, and transformation of conflict whose face changes rapidly from aggression by overwhelming force to prevention and de-escalation. Similar principles can be applied to healthcare operations, particularly in view of the already existing major technological components of the WHIG. However, in order to implement network-centricity in healthcare, a major conceptual transformation is required.

Presently, the ruling healthcare doctrine is that of e-health, which while supporting implementation of IC^2T, promotes development of individual, largely noncollaborative (particularly in the global sense) systems. While there is no doubt that the existence of such systems (for example, electronic patient records) facilitates many aspects of healthcare delivery and administration, their effect is predominantly regional. On a larger scale (national, international) most of these platforms function in isolation and major (predominantly through human interaction) effort is needed in order to extract relevant information and convert it into pertinent knowledge.

Transition to the network-centric doctrine of healthcare will greatly facilitate interoperability of multiple electronic healthcare platforms and enhance their usefulness in the broadest sense of global health. There is also no doubt that, similar to other domains in which a network-centric approach has been successfully implemented, the consequence of the proposed doctrine will be improvement of access, better delivery, increased efficiency in the use of resources, accompanied by the concomitant reduction of presently staggering expenditure.

NOTE

The authors of this article are listed alphabetically. Both contributed equally.

REFERENCES

Alberts, D. S., Garstka, J. J., & Stein, F. P. (2000). *Network centric warfare: Developing and leveraging information superiority* (CCRP Publication Series 1-284 (Department of Defense)). Retrieved January 15, 2007, from http://www.dodccrp.org/publications/pdf/Alberts_NCW.pdf

Banjeri, D. (2004). The people and health service development in India: A brief overview. *International Journal of Health Services, 34*, 123-142.

Boyd, J. R., & COL USAF. (1987). Essence of winning and losing. In *Patterns of conflict* (Unpublished Briefing). Retrieved January 15, 2007, from http://www.d-n-i.net

Cebrowski, A. K., & Garstka, J. J. (1998). Network-centric warfare: Its origin and future. *US Navy Instant Proceedings, 1*, 28-35.

Department of Defense (2001). *Report to Congress: Network centric warfare*. Washington, DC: Office of the Secretary, Department of Defense.

Fernandez, I. (2002). Barcelona 2002: Law, ethics, and human rights global battle cry: health is a right, not a commodity. *HIV/AIDS Policy & Law Review, 7*, 80-84.

Larson, E. B. (2004). Healthcare system chaos should spur innovation: Summary of a report of the Society of General Internal Medicine Task Force on the Domain of General Internal Medicine. *Annals of Internal Medicine*, *140*, 639-643.

National Commission on Terrorist Attacks on the United States (2004). *The 9/11 report with reporting analysis by The New York Times* (pp. IX-CXII, 1-636). New York: St. Martin Press.

O'Brien, J. (2004). *Management information systems*. Boston: Irwin McGraw-Hill.

Olutimayin, J. (2002). Communication in health care delivery in developing countries: Which way out? *Pacific Health Dialog, 9*, 237-241.

Onen, C. L. (2004). Medicine in resource-poor settings: Time for a paradigm shift? *Clinical Medicine, 4*, 355-360.

von Lubitz, D., & Wickramasinghe, N. (2005a). Creating germane knowledge in dynamic environments. *International Journal of Innovation Learning.*

von Lubitz, D., & Wickramasinghe, N. (2005b). Healthcare and technology: The Doctrine of Networkcentric Healthcare. *Health Affairs.*

von Lubitz, D., & Wickramasinghe, N. (2005c). Networkcentric healthcare and bioinformatics. *International Journal of Expert Systems.*

World Health Organization Report (2000). *Health systems: Improving performance* (pp. 1-215). Washington, DC: WHO.

World Health Organization Report (2004). *Changing history* (pp. 1-167). Washington, DC: WHO.

KEY TERMS

E-Health: The application of technology, primarily Internet based technology, to facilitate in the delivery of healthcare.

Germane Knowledge: The relevant and critical knowledge, or contextualized information, required to enhance a particular decision.

Information Symmetry: The gap between the available information between two entities.

Network-Centric: In contrast to a platform-centric approach, a network-centric approach is made up of interconnecting technology grids that enable and facilitate the seamless transfer of data, information and knowledge.

OODA Loop: A framework developed by John Boyd that facilitates rapid decision making in dynamic, rapidly changing environments characterized by a high degree of structural and event complexity. Each complex action can be subdivided into a series of consecutive cycles, while each revolution (cycle) of the Loop comprises of four stages: Observation, Orientation, Determination, and Action.

Platform-Centric: Based on and exploiting the exclusive properties of an employed system or specific technology platform. Useful on a small scale but does not enable seamless transferring of information and knowledge across platforms or systems.

World Healthcare Information Grid (WHIG): The technology backbone of network-centric healthcare operations, a network of interconnecting technology grids that together contain all the necessary information for effective and efficient healthcare delivery.

This work was previously published in Encyclopedia of Portal Technologies and Applications, edited by A. Tatnall, pp. 647-652, copyright 2007 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.9
Mobile E-Health:
Making the Case

Norm Archer
McMaster University, Canada

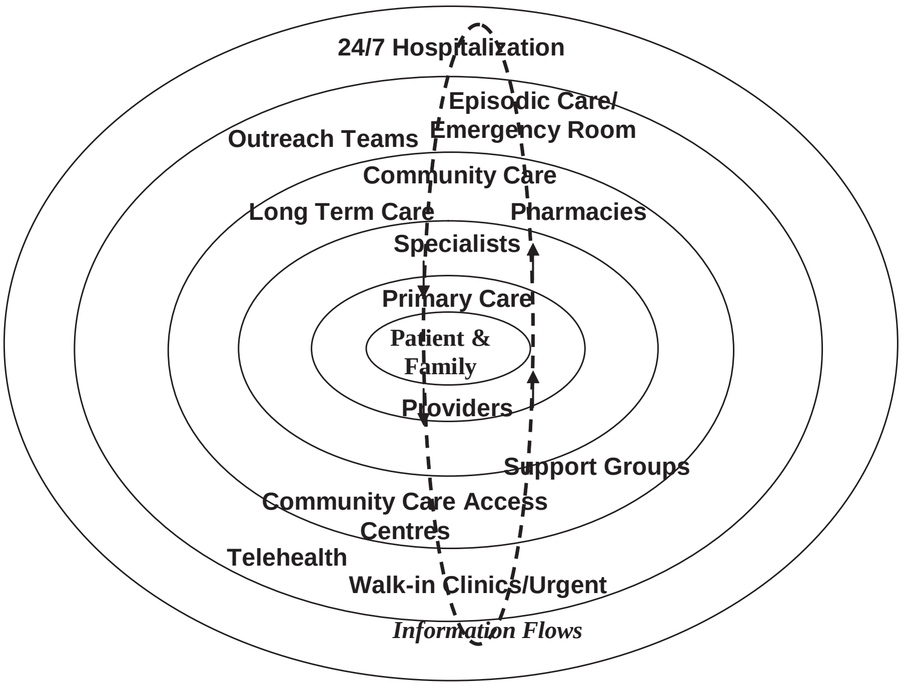
ABSTRACT

Health care is an industry with a diverse set of stakeholders: governments, private health care providers, medical practitioners (physicians, nurses, researchers, etc.), home health care providers and workers, and last but not least, clients/patients and their families. Overlapping and interacting environments include hospitals, clinics, long-term care facilities, primary care providers, homes, and so forth, involving acute, emergency, chronic, primary, and outpatient care. Patient transitions between these environments are often unnecessarily difficult due to an inability by providers to access pre-existing patient records. Mobile/wireless solutions can play an important role in supporting health care by providing applications that access health care records and reduce paperwork for clinical physicians, nurses, and other workers, community health care practitioners and their patients, or mobile chronically ill patients such as diabetics. This chapter makes the case for mobile health care and its solutions in the non-acute community health care environment, where critical issues include usability, adoption, interoperability, change management, risk mitigation, security and privacy, and return on investment. A proposed community health care application demonstrates how these issues are addressed.

INTRODUCTION

Many individuals receive their care from more than one caregiver or other provider: individual physician, group practice, hospital, long-term care facility, laboratory, pharmacy, walk-in clinic, urgent care center, work-site clinics, school clinics, and so forth. When and where choice is available, clients can select caregivers or other providers based on their proximity, bedside manner, quality and capability, cultural aptitude, or other factors. Clients/patients also must move among health care providers as their state of health changes, creating a need for their health records/histories to move with them, so each provider does not need to prepare a totally new patient history at admission (see Figure 1). Without the general existence of digital health records and some means of integration or

Figure 1. Information flows in patient-centred health care (adapted from Krull-Naraj, 2004)



interoperability, individual choice and movement leads to fragmentation of the individual's health care experience. Typically, this means that client record transfers are accomplished from one or more sources via paper, scanned digital records, and/or fax. The result is that client records may be stored on paper at a number of caregiver institutions or re-keyed into institutional databases, with no possibility of version control or compatibility. This fragmentation of records often leads to errors, duplication, lack of coordination (Brailer, 2005), conflicting approaches to a patient's health care, service and/or resource duplication and many other problems including reduced quality of care, reduced effectiveness, and increased cost to society.

Integration mechanisms have been tried in the past, although none has delivered lasting benefit. These include horizontal and vertical mergers of providers, state-sponsored networks of community care services, and so forth. However, there is theoretically no technical barrier to establishing a network of providers that would use information in an interoperable manner for integrated support of patient care, and this would not require a massive integration of physical assets and the bureaucracy required to operate it. In the U.S. alone, the potential savings from such an approach have been estimated at U.S. $77.8 billion per year (Walker, Pan, Johnston, Adler-Milstein, Bates, & Middleton, 2005). This does not include the substantial clinical and quality of life benefits from this approach. Unfortunately, it has been virtually impossible in most jurisdictions to get broad agreement on a standard and portable electronic health record (EHR) that would support this interoperability. Although the standardized EHR has a continuing focus of the health care community (Berner, Detmer, & Simborg, 2005), the process of adopting a standard would still require large investments in the database conversion process and the necessary secure communications network that would safeguard client privacy and confidentiality. However, without interoperability and health information exchange, health information will remain as it is now, in proprietary and often inaccessible silos.

The lack of a standard EHR definition, and the lack of interoperability among health care providers are both major strategic issues in information technology support for health care. No doubt at some time in the distant future both these problems will be addressed at a macro-level, but in the meantime, it is essential to make incremental improvements that will adapt to the constantly changing environment in a way that continuously improves health care support at the local level (Lenz & Kuhn, 2004). Such changes may be significant enough to be disruptive to users, so it is critical to plan and implement changes so that already overburdened health care professionals and administrative support receive relief in the most effective and efficient manner possible under the circumstances.

Health delivery practice for non-acute care in many industrialized countries is shifting toward the home. The reasons are the better possibilities for managing chronic care and controlling health delivery costs, but the appropriate infrastructure must be in place in order to maintain client quality of life through quality health services, and the need to predict and thus avoid serious complications. For this potential to be realized, new interoperable telemedicine and information technology (IT) solutions need to be implemented and integrated in the health delivery system, but these solutions need to be assessed through evidence-based medicine in order to provide solid proof for their usefulness. To ensure that quality care delivered efficiently any time and any place requires ready access to patient records and expertise from remote sources such as specialists and online databases, and mobile wireless technology can enable this support. True wireless communities where processes, technology, and people are fully aligned to mobile applications, have a great deal of potential, especially where smooth transitions by patients among the types of care they may experience, including acute, emergency, chronic, primary, and homecare, require continuing coordination among health care institutions, medical practitioners, health care workers, care givers, and the patients themselves. The real benefit of mobility support will come only when technology and process are built around a plan that embraces mobility, and where mobility is not an afterthought.

As mobile wireless unfolds, the health care world must assess both its technical and value propositions, to determine if it has a real value proposition to offer, in terms of quality of life maintenance at lower costs than existing systems. The objective of this chapter is to discuss the roles that electronic mobile solutions can play in health care, and particularly the value propositions and their evaluation that must play a role in the selection of these solutions for efficient and effective use. The concepts presented will be demonstrated by an application in a real mobile health care application.

MOBILITY

Mobility is an aspect of many environments. How mobility affects individuals tends to differ, depending on the nature of what they are doing (working, relaxing, traveling, etc.), their preferences, the form that mobility takes, and the amount of time involved. Supporting mobility through electronic solutions is having a growing impact on individuals by enabling them to carry devices that assist them to stay in constant communication with their organizations, friends, family, and advisors. This may involve using voice or data messages, paging, direct communications by telephone or teleconferencing, and database or document information access, storage, and retrieval. In working environments such as hospitals, such applications are often built upon existing e-business solutions such as corporate and operational databases, along with functional or corporate support areas and their associated networks such as LANs and Internet connectivity. The growing availability of a variety of mobile applications and technologies has encouraged the extension or replacement of existing approaches and business processes. We are only beginning to see true wireless environments where processes, technology, and people are fully aligned to a mobile environment.

Mobile Solutions

The potential selection of systems and devices for mobile support includes voice and data communications, ranging all the way from cell phones to laptop computers and PDAs (personal digital assistants). It will be assumed that end users have ready access to voice communication through cellphones, since these are rapidly becoming as ubiquitous as landline telephones. Mobile or wireless solutions discussed here may use the same networks, including and extending voice cellphone communication. Mobile solutions can be used to support hospital or clinical workers, or community health care practitioners or their patients, where the workers may be away from their home office a high percentage of the time, while traveling or meeting with clients. Others may need to travel occasionally to different sites for meetings, conferences, or training. Clients such as ambulatory care patients may also adopt mobile or monitoring devices, in order to use the services of mobile health care providers. Mobile solutions allow employment hours to be flexible and to extend beyond those hours actually spent in the office, including lunch and break time, traveling to and from work, traveling to meetings, holidays, weekends and evenings. Although this may result in an attendant increase in productivity, working with others through mobile applications may not necessarily fulfill all the needs of workers.

Pervasive Computing

Pervasive computing can be defined as personalized computing freed from the desktop, enabling information access anywhere, anytime, on demand. This provides an apt description of the objective of mobile worker support. Computing devices range from desktop (fixed), to laptops and palmtops (transportable) to handhelds and wearables (fully mobile) (Gorlenko & Merrick, 2003). Mobile devices can be differentiated according to their wireless connectivity. Content transmitted by technological solutions in the mobile wireless world can be mobile (but not wirelessly connected for synchronization with wireline content), wireless (but not mobile), or both mobile and wirelessly connected. Mobile wireless content is converging with the wireline Internet, with the result being referred to as the mobile Internet. The growing mobile wireless market demands both voice and data (text-graphics) communication services. Multimedia content is a suitable mix of the two. The content is carried through the network of a wireless network operator and a service provider. Some mobile devices are unconnected while on the move (e.g. PDAs—personal digital assistants, laptops, and palmtops) although they may be equipped for wireless connectivity in a stationary environment. Clinical, business, and technology functions are usually intertwined in a complex manner in wireless systems, it is essential to maintain patient safety and quality of life, if a health care system is to be implemented acceptably (Scalise, 2005).

Value Proposition for Mobile and Wireless Solutions in Health Care

The business model for adopting mobile solutions is the economic justification for the use of the technology, or the means by which the technology generates a value proposition. In the current wireless marketplace, with an increased system complexity that is driven by the number of players and their interactions (Olla & Atkinson, 2004), (network operators, carriers, content providers, mobile device manufacturers, etc.), the value chain is developing into linkages of partnerships for delivering value to end customers (Sabat, 2002). End customers in e-health may be individual physicians or they may be networks of health care workers who interact in such a way that mobile solutions can assist in improving their interactions cost effectively. Business partnerships involved in supporting mobile solutions continue to evolve, engage, and disengage as new technology evolves and appears, and certain business partners thrive while others fail.

It is often difficult to justify a business case for a mobile project financially. A 2002 survey by *CIO Magazine* (Worthen, 2002) indicated that the two most popular measures of ROI for wireless projects were increased productivity (54%) and improved internal customer satisfaction (40%). For mobile e-health applications, we can add an important constraint which is that the application must not decrease the quality of life of the clients, or ROI justification becomes meaningless.

One framework that has been proposed for value determination of mobile solutions, includes two dimensions: time and place—the work can be either dependent or independent of one or both these dimensions (Wiberg & Ljungberg, 2001). Mobile applications in this framework can be in one of four quadrants: anytime and anywhere; anytime and a particular place; a particular time and anyplace; and a particular time and particular place. Maglaveras et al. (2002) discuss a community-based health care support system that proved the usefulness of wireless technology in providing wireless interactivity anytime and anywhere, but also proved the necessity for restructuring educational medical knowledge for delivery to the patient. An example application of the anytime-anyplace model is in mobile support of chronically ill clients who are still able to work. Here, occurrences could be identified in all four of the quadrants, since clients have particular places where they may spend a considerable amount of time (at work and at home), and certain procedures could be specified at particular times, but there would be value in support in the "anytime and anywhere" quadrant. These concepts help in planning potential mobile support applications.

To justify mobile solutions, health care institutions that deploy wireless data solutions do so on a very selective basis, supporting only those employees who have a demonstrated need for real-time access. There are strong indications that return on investment (ROI) can be most strongly justified for specific classes of tasks in vertical markets such as health care, manufacturing, government, and transportation (Wheelwright, 2002). For example, a recent study of mobile solutions in a variety of applications in 35 major companies found hard benefits that included sales increases of 5-10%, reduced customer wait times by as much as 80%, increases in service calls of up to 32%, and service call responsiveness improvements of up to 7% (Gillott, 2002). Payback periods ranged from a few months to 30 months.

Mobility and flexibility are the biggest drivers of mobile solutions in many institutions and companies (Wheelwright, 2002). Health care companies may choose wireless solutions because they have an outbound workforce that needs to be connected within a corporate environment or when making calls on homecare clients. Cost has slowed the adoption of mobile technology in the past, although it is becoming less of a concern as prices drop and businesses recognize the benefits of offering wireless access to their workforce. Mobile portals provide convenient places where Web users can link to a set of applications that are relevant to their interests and/work (Clarke & Flaherty, 2003). Portals assist the wireless user to interact with Web-based content, and serve a valuable purpose in aggregating multiple applications and/or content providers through one Web site. They also provide a greater degree of personalization and localization than traditional Web portals.

EVALUATION OF MOBILE BUSINESS APPLICATIONS

The novelty of many of the mobile applications currently entering the marketplace, along with inexperience of business with mobile solutions, greatly increases the risks associated with adopting such solutions. For this reason, the business value proposition of proposed mobile applications must be studied with care. To that end, we have developed a process framework (Archer, 2004) that organizes the planning and evaluation process logically. This proceeds from identifying the business goals, defining potential user groups and the applications they would use, and the technical considerations that will lead to the appropriate mobility choice. When implementation issues are factored in, tempered by a variety of moderators, the application can be evaluated, along with its ROI, and compared with the existing application in terms of tangible values such as revenue, cost, and efficiency, and intangibles such as user and customer satisfaction. This process is an essential first phase in any mobile application, since it considers logically the costs and benefits of implementing the planned solution.

Key Issues for Mobile E-Health Applications

A number of key issues can affect the potential for a successful mobile e-health implementation. These include usability, adoption, interoperability, change management, risk mitigation, privacy and security, and return on investment. These are considered in more detail in the following.

- **Usability** can be defined as the quality of a system with respect to ease of learning, ease of use, and user satisfaction (Rosson & Carroll, 2002). It also deals with the potential of a system to accomplish the goals of the user. Usability is a key issue in the adoption of any information system, but it is particularly so for mobile systems, where the end-user device is often hand-held, with limited display and data entry capability (Tarasewich, 2003). Interface design and the design of the device itself have a critical impact on usability.
- **Adoption:** As in any user population, technological changes in the supporting technology for doing tasks make demands on both the quality of user interface and the functionality of mobile devices. Adoption is clearly related to usability, although it involves additional issues. The questions of interest are (Zhu, Nah, & Zhao, 2003): (1) what factors influence users' adoption of mobile computing?; (2) how does the design of mobile devices and interface affect user adoption?; and (3) to what degree do specific factors such as trust and enjoyment (in using mobile devices) play a role in adoption? Zhu et al. (2003) have proposed that perceived ease of use (input and output modalities, navigation, bandwidth), perceived usefulness (service offerings, degree of mobility, compatibility, coverage, reliability), trust (security, privacy, vendor characteristics, perceived ease of use, perceived usefulness), and enjoyment (congruence of skills and challenges, focused attention, interactivity, perceived ease of use, perceived usefulness) will affect intentions to use mobile devices, which will then influence actual usage. Evidence suggests that inadequate access to information and ineffective communication tend to be causes of error and other adverse events for in-patient care (Mendonca, Chen, Stetson, McKnight, Lei, & Cimino, 2004). Information-based handheld wireless applications at the point of care that link to clinical data can help reduce these problems.
- **Interoperability:** Interoperability with existing health care applications and/or databases is necessary, to improve or at least

avoid worsening any existing "stovepipe" characteristics that plague the health care IT field. In the absence of compatible databases or health records, separate applications may communicate through messages containing health record information using standard protocols such as HL7.[2] Supporting software could be developed with a messaging toolkit such as Chameleon,[3] or the health record information may be mapped to a particular database using Iguana.[2] Significant barriers that must also be addressed include policies on accessing and updating existing systems, due to privacy, confidentiality, legal, and regulatory concerns.

- **Change Management:** Many issues arise when technology is changed or introduced to a user population, often transforming the way users must perform their tasks. The manner with which change is introduced, irrespective of the effort invested in enabling it, will impact system adoption. There are multiple obstacles to implementing such a system. These include resistance to change by end users and IT staff, and integration with existing systems. These must be considered in advance in order to mitigate potential risks (Wang & Paper, 2005). There are two important aspects of change that must be considered. The first is in the organizational and business process structures. Not all change can be anticipated and planned in advance. Difficulties in managing change in the introduction of technology arise when the organization does not plan the management of unanticipated change. This increases the complexity of change because the organization unprepared to deal effectively unanticipated complications. For example, new technology often leads to anticipated changes in staffing levels, but there are usually psychological or social dimensions to the remaining jobs that are not anticipated.

Two types of unanticipated change include (Rivard, Aubert, Patry, Pare, & Smith, 2004):

- Emergent change, that may arise spontaneously in response to planned change, with either positive or negative impacts on the organization (for example, a requirement for new skills in analyzing data now being collected that was not previously available).
- Opportunity-based change, introduced intentionally during the change process in response to unexpected impacts. For example, the introduction of mobile technology may create new opportunities for other applications that were not previously considered possible when the mobile infrastructure was not yet in place.

The second major consideration is the technology itself. Even if the new technology offers the full functionality required, with a fully-tested interface, user compliance is far from guaranteed. Studying user interactions with the system, as well as improving it to suit their needs, can be complex, costly, and time intensive but may result in significant rewards. Simple modifications may be introduced to help users to learn and operate the system, thus increasing acceptance. Schoenberg, Safran, and Sands (2000) suggest a functionality for assessing system performance from the user perspective: acquire user information in the background and through direct survey, target population subsets of interest and avoid interrupting those who are not, invoke data acquisition methods "just-in-time" as the user interacts with the system, be as brief and concise as possible during interaction, provide incentives to compensate users for their time, and be consistent across all applications.

- **Risk Mitigation:** End-user adoption is an important issue, and this is addressed by change management, usability, and so forth. Obstacles to end-user adoption of the system must be mitigated, through training and motivation for end users and clients through improved quality of work and life. System reliability is critical in health care, and parallel testing during the test phase until there is sufficient confidence in system reliability (Mikkelsen & Aasly, 2001) is one way to develop confidence in the system.

- **Security and Privacy:** Health record privacy and confidentiality in many jurisdictions are regulated strictly by government guidelines, in terms of controlling need to know and client permission to access records. Mobile applications must adhere to these guidelines, as well as maintaining security by encoded communications and databases. An additional consideration is the reliability of the system, to ensure that information is not lost and that service interruptions are managed effectively.
- **Return On Investment:** In health service applications, ROI (return on investment) considerations are overridden by a constraint that quality of health care must be improved, or at least not reduced. In calculating ROI, cost considerations include software, hardware, installation, ongoing maintenance, and further development. Savings include some that are more easily quantifiable such as IT staff time, and employee time savings (e.g., time saved when data can be entered directly online, or reduction in errors and the associated time needed to fix them). Virtual office operations and field sales operations tend to have a higher ROI than company groups with little customer-facing work. Health services require a convincing case for the adoption of mobile applications, since there are so many other competing demands on resources. In addition, since there are typically a number of stakeholders (client, physicians and nursing staff, administrative staff, institutions—hospitals, homecare institutions, etc.), each should receive a perceived benefit in order to encourage participation.

CASE EXAMPLE: MOBILE APPLICATIONS IN HOME HEALTH CARE

Health care is the responsibility of the provincial governments in Canada, where the publicly-funded system provides universal access. In the Canadian province of Ontario, health care has been organized into a multi-tiered system. Family physicians provide primary care, typically operate as small businesses, and bill the province for services rendered. They may refer patients to specialists, who may also operate as small businesses or work as hospital employees. Acute care is supported by hospitals. Non-acute care is supported in long-term facilities or in a home environment. Laboratory testing services may be operated privately or in a public hospital. Patients are supported in the home environment by their families and friends, in addition to homecare nursing and housekeeping support that is managed and funded by the Province. To provide homecare and long-term support, the Province has chartered a number of Community Care Access Centres (CCACs) throughout the Province, that contract with private homecare providers for nursing and other patient support.

As demonstrated in Figure 1, information must flow among the various providers in order to provide continuity to patient support, and access to information that exists at many places throughout the system. Unfortunately, the number of Canadian primary care physicians that have fully digitized records is only in the neighbourhood of 5%. Although each of the types of health care institutions (hospitals, CCACs, and homecare providers) has internal digital systems and databases that support its own applications, there are no standards for health records, and little digital communication to support specific clients. Most information flows are on paper forms, usually transmitted by fax, and scanned and stored or transmitted as images. This may serve storage and legal purposes, but the information can not be retrieved in a form that can be used for management or decision support analysis. There are the usual problems in the paper-based system, resulting from delays, errors, and loss of information.

To solve some of the problems associated with current paper-based systems, a project has been proposed for Venus,[4] an Ontario county, to use wireless applications to link mobile homecare workers to centrally supported applications and databases. The Venus County Mobile E-Health Project is being planned through the collaboration of three health care organizations: Mars Home Health (provider of visiting nursing and other services), Venus County CCAC, and Jupiter Hospital. The Aquarius Research Centre plays a role in research and development, and a number of corporate vendors are also involved, including Zodiac Wireless Inc. A carefully researched process will be used to build and trial the system, with the end result being a mobile system that can be easily implemented as an ongoing commercial operation in Venus County and potentially throughout other Canadian jurisdictions.

The project is aimed at improved outpatient quality of care, while eliminating or easing routine work for health care workers. The initial focus will be on reducing or eliminating paperwork by homecare nurses accessing the remote system wirelessly onsite for client details or to enter new data. The project will be led by health care staff and professionals to maximize the chances of successful implementation. The project is incremental in approach, and undertaken in phases, with activities and outcomes carefully researched and evaluated throughout each phase. The project may be terminated or adjusted at the end of each phase. The focus will be on delivering real benefits for clients and health care providers while overcoming the barriers to information exchange among some of the entities indicated in Figure 1, without undue inconvenience to regular health care activities.

An analysis of the business case through a logical process (Archer, 2004) has predicted an approximate reduction of 50% in direct (labour and system) costs by the project. Intangible benefits include: delay until data is available online reduced from 24 to 0 hours, online availability of data increased from 10% to 100%, error rate for data entry reduced from 0.20 to 0.05 per data item, homework by homecare nurses (faxing and other work) reduced from substantial to minimal, time for client-centred care increased during nursing visits, and time required for routine work reduced.

Project Phases

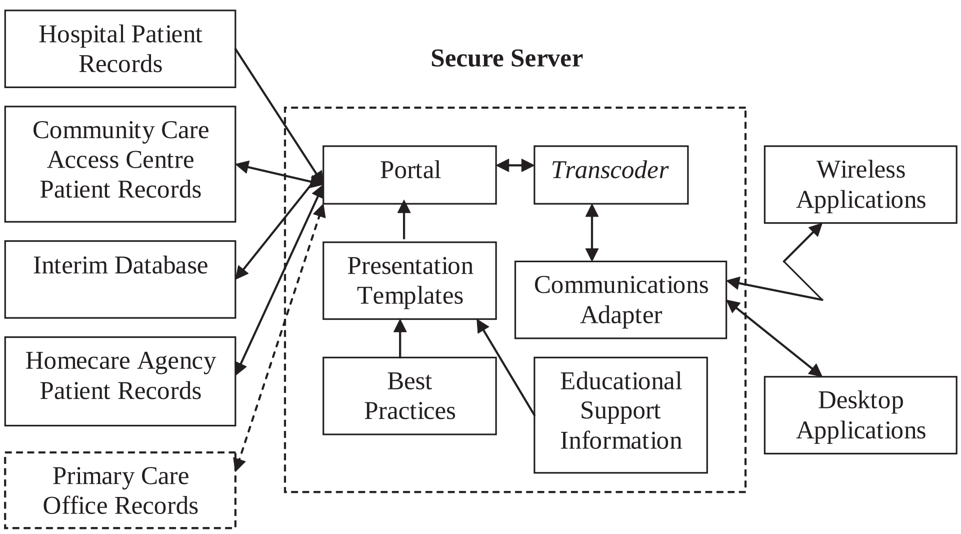
Each phase includes a research component, which will be used to plan the work, collect information, and evaluate results to ensure that the outcome of the phase is optimal from the point of view of the overall project objective. The following describes in limited detail the major phases in the project:

1. A prototype of several mobile wireless applications with a high potential impact on homecare operations, will be designed, tested, and implemented through a trial by a small group of Mars Home Health homecare nurses. Aquarius and Zodiac will be heavily involved in interface design, and a study of usability and user acceptance. The highest impact applications currently in use in paper form by homecare nurses are: supplies ordering, wound care, key path, admission, status change, and discharge. These will be implemented in the first prototype. Technical considerations include the choice of suitable end-user devices to support applications, and wireless network provision. Secure access will be available to data through both wireless mobile and office computers. Linkages

to existing databases at Mars Home Health and Venus CCAC will be designed to ensure interoperability. Figure 2 is a conceptual description of the proposed system. Interoperability challenges with existing databases require the development of an interim database for data that does not currently exist in digital form. Decision support applications for clinical and administrative use will be able to access all the digital data collected. The secure server provides application support for wireless and desktop devices. Due to the low reliability of access to wireless devices in remote areas (and sometimes in basements or other shielded areas), the devices will need to carry "fat client" applications. That is, applications will run securely on the devices rather than on the server, and data uploads and downloads will occur automatically when wireless access is available.

2. Based on a successful prototype test in phase 1, the prototype will be revised for full operations according to prototype results and learning, and rolled out as a commercial operation to the entire Mars Home Health visiting nurse population of over 120.
3. Further applications specified and required by Mars Home Health and Venus CCAC, will be designed, based on experience in Phase 1, resulting in a suite of applications based on previous and new applications. The full suite will be trialed with a small group of Mars homecare nurses, including an evaluation of the suite (usability and user acceptance) and revisions necessary for commercial operations. This will be followed by a full commercial rollout at Mars.
4. Because there will be a significant increase in data availability online, there will be an investigation into how to make use of the additional information gathered from wireless inputs available online to the health care providers, for managing and decision making. Applications will be designed and developed as appropriate, and the impact on the institutions and clients involved will be studied. An example is improved wound management data collection for monitoring and control purposes:
    - To facilitate a consistent technique in wound assessment;
    - To facilitate a consistent means of documenting wound status;

Figure 2. Data flows in mobile e-health support system



- To be a tool to infuse evidence-based practice in wound care management; and
- To provide a means of data collection that would aid in evaluation and outcome tracking.

5. The final phase will be to design and implement a wireless mobile e-health prototype to assist the three health care institutions to support newly educated Type 2 diabetics. The client process includes working through an education process in the Jupiter Hospital's Diabetes Education Centre, continuing as the clients move out of the hospital into a normal work and homecare situation over a period of several weeks. The desired end result is client self-management of diabetic condition with integrated back-up support from the health care institutions and primary care physicians.

CONCLUSIONS

Mobile health care technology has the potential for not just supporting health care in any particular health care environment (hospitals, clinics, long term care facilities, homecare), but for more easily managing transitions as patients move from or to acute, emergency, chronic, or primary care. Mobile applications can either eliminate or greatly reduce the use of paper forms, thus reducing system cost, as well as reducing errors and delays in making digital information available online. But if mobile e-health is to be introduced successfully, care must be taken to include all the stakeholders in planning and implementing mobile solutions. Critical issues that have been identified for mobile e-health applications include usability, adoption, interoperability, change management, risk mitigation, security and privacy, and return on investment.

REFERENCES

Archer, N. (2004). *The business case for employee mobility support* (Working Paper No. 9). Hamilton, ON: DeGroote School of Business, McMaster University.

Berner, E. S., Detmer, D. E., & Simborg, D. (2005). Will the wave finally break? A brief view of the adoption of electronic medical records in the United States. *Journal of the American Medical Informatics Association, 12*(1), 3-7.

Brailer, D. J. (2005, January 19). Interoperability: The key to the future health care system. *Health Affairs,* W5-19–W15-21.

Clarke III, I., & Flaherty, T. B. (2003). Mobile portals: The development of m-commerce gateways. In B. E. M. a. T. J. Strader (Ed.), *Mobile commerce: Technology, theory, and applications* (pp. 185-201). Hershey, PA: IRM Press.

Gillott, I. (2002, April). The ROI benefits of wireless. *Communications News, 39*, 6.

Gorlenko, L., & Merrick, R. (2003). No wires attached: Usability challenges in the connected mobile world. *IBM Systems Journal, 42*(4), 639-651.

Krull-Naraj, K. (2004). *Information flows in patient-centred healthcare.* Barrie, Ontario.

Lenz, R., & Kuhn, K. A. (2004). Towards a continuous evolution and adaptation of information systems in healthcare. *International Journal of Medical Informatics, 73*(1), 75-89.

Maglaveras, N., Koutkias, V., Chouvarda, I., Goulis, D. G., Avramides, A., Adamidis, D., et al. (2002). Home care delivery through the mobile telecommunications platform: The Citizen Health Citizen (CHS) perspective. *International Journal of Medical Informatics, 68*, 99-111.

Mendonca, E. A., Chen, E. S., Stetson, P. D., McKnight, L. K., Lei, J., & Cimino, J. J. (2004). Approach to mobile information and communication for health care. *International Journal of Medical Informatics, 73*, 631-638.

Mikkelsen, G., & Aasly, J. (2001). Concordance of information in parallel electronic and paper based patient records. *International Journal of Medical Informatics, 63*, 123-131.

Olla, P., & Atkinson, C. (2004). Developing a wireless reference model for interpreting complexity in wireless projects. *Industrial Management and Data Systems, 104*(3), 262-272.

Rivard, S., Aubert, B. A., Patry, M., Pare, G., & Smith, H. A. (2004). *Information technology and organizational transformation*. Amsterdam, Netherlands: Elsevier.

Rosson, M. B., & Carroll, J. M. (2002). *Usability engineering: Scenario-based development of human-computer interaction*. San Francisco, CA: Academic Press.

Sabat, H. K. (2002). The evolving mobile wireless value chain and market structure. *Telecommunications Policy, 26*, 505-535.

Scalise, D. (2005). One generation to the next. *Hospitals and Health Networks*(Winter), 10-15.

Schoenberg, R., Safran, C., & Sands, D. Z. (2000). *Studying clinician-computer interaction in Web-based systems*. Paper presented at the Annual AMIA Symposium, Los Angeles, CA, November 4-8.

Tarasewich, P. (2003). Wireless devices for mobile commerce: User interface design and usability. In B. E. M. a. T. J. Strader (Ed.), *Mobile commerce: Technology, theory, and applications* (pp. 26-50). Hershey, PA: IRM Press.

Walker, J., Pan, E., Johnston, D., Adler-Milstein, J., Bates, D. W., & Middleton, B. (2005, January 19). The value of health care information exchange and interoperability. *Health Affairs,* w5-10–w15-18.

Wang, B., & Paper, D. (2005). A case of an IT-enabled organizational change intervention: The missing pieces. *Journal of Cases on Information Technology, 7*(1), 34-52.

Wheelwright, G. (2002, Dec. 3). The business benefits of mobility. *Financial Times of London,* p. 11.

Wiberg, M., & Ljungberg, F. (2001). Exploring the vision of "Anytime, Anywhere" in the context of mobile work. In Y. Malhotra (Ed.), *Knowledge management and business model innovation* (pp. 153-165). Hershey, Pennsylvania: Idea Group Publishing.

Worthen, B. (2002, December 1). When wireless works. *CIO*, *16,* 1.

Zhu, W., Nah, F. F.-H., & Zhao, F. (2003). Factors influencing user adoption of mobile computing. In J. Mariga (Ed.), *Managing e-commerce and mobile computing technologies* (pp. 260-271). Hershey, PA: IRM Press.

ENDNOTES

[1] This research was supported by a grant from the Social Sciences and Humanities Research Council of Canada.

[2] http://www.hl7.org/library/implement/implementation.htm

[3] Trademark of iNTERFACEWARE Inc. http://www.interfaceware.com/

[4] Venus County is a pseudonym, and Venus County CCAC, Mars Home Health, and Jupiter Hospital are pseudonyms for the health care organizations involved in the project. Aquarius Research Centre and Zodiac Wireless Inc. are pseudonyms for the research and business organizations involved.

Chapter 1.10
Mobile Telemonitoring Insights

Pantelis Angelidis
Vidavo Ltd., Greece

ABSTRACT

Technology advances create new possibilities for healthcare monitoring, management, and support, focusing on prevention rather than disease management. The provision of personalized healthcare applications is also greatly supported. Developments in the wireless and mobile markets are capitalized by the medical device industry. Services are becoming personalized and location independent to fulfill the increasing patient needs for self-empowerment and quality in the healthcare delivery away from the traditional nursing areas. This overview discusses the new opportunities for the healthcare domain in the mobile times we live.

INTRODUCTION

The healthcare industry is experiencing a substantial shift to care delivery away from the traditional nursing areas due to the convergence of several technology areas. Increasingly capable health-monitoring systems are moving the point of care closer to the patient, while the patient, better informed and aware now, undertakes an active role to self-care and/or -prevention. Emerging ICTs in conjunction with the medical device industry development (intelligent devices, biosensors, novel software, etc.) demonstrate personalized healthcare delivery's potential without geographical limitations.

The concept of prevention prevails now against disease management and treatment plans. As patient-centric processes emerge, the citizens and patients undertake an active role in monitoring their health status. Meanwhile, e-wellness evolves to address the rising expectations of the e-health consumers, who are better informed, more demanding, and empowered. The empowered, worried-well consumers require quality health services on the spot. The drivers are now connectivity, speed, and personalization (McKnight, 2000).

MOBILE HEALTHCARE PROVISION

Waves of technology incorporation and scientific discoveries have driven the sector from reliance on direct communication and physician experience to a higher reliance on technology and community information. This new Web-enabled environment has taken healthcare from local areas, where telemedicine left it, literally into the patient's home and, more recently with m-Internet, to wherever the patient might be and whenever he or she needs it (Simão, 2001).

M-Internet enables information exchange and promotes the availability of services and communication modes to serve working teams with increasing mobility requirements.

Services are becoming personalized and location independent to serve increasing patient needs for self-empowerment and quality in healthcare delivery away from the traditional nursing areas.

Furthering the new approaches in the provision of healthcare services in the frame of e-health, wireless developments create new opportunities for healthcare professionals, individuals and organizations, patients, and health authorities. The scope of mobile health addresses clinical, administrative, and consumer health-information applications and, as it could contribute to the improvement of health outcomes, m-health may be utilized to measure health status and population welfare.

Many healthcare organizations are investing in IT projects that take advantage of new technologies in the mobile healthcare application space. Functionality that augments the capture of evidence-based patient plans of care is essential and must map and bridge the information flow for both inpatient and outpatient work-flow clinical-practice guidelines. As the medical community continues to embrace these new technologies, system integrators must provide functionality that reduces costs, improves the quality of care, and improves the ease with which caregivers can perform their everyday tasks (Wolf, 2001).

The most significant challenge posed by mobile technology is the seamless integration of multiple hardware and software platforms with reliable, uninterrupted wireless services in a secure manner, which will become mission critical to successful healthcare organizations, payers, and providers (Wolf, 2001).

The current state-of-the-art technology in medical sensors allows for the easy and unobtrusive electronic measurement of several health conditions. The sensors are often stand-alone devices and sometimes comprised of two or more elements connected by a cable or wireless technology. Medical sensors have the capability to measure vital signs such as blood pressure, pulse rate, respiration frequency, and so forth. Based on these medical parameters, the medical professionals can monitor the patient's health condition and act in case of an anomaly.

The application areas of the medical-device wireless telemonitoring capabilities include the following:

1. Assistance in case of accidents and emergencies
2. Increased capacity and lower costs for hospitals
3. Assistance and monitoring in a home-care setting
4. Monitoring of chronically ill patients
5. Patient involvement in setting a diagnosis
6. Medicine dosage adjustment
7. Physical-state monitoring in sports
8. Monitoring of sporadically occurring symptoms
9. Emergency alarms (Fosse & Haug, 2003)
10. Improved health management

As a result, citizens can enjoy quality healthcare provision and an elevated quality of life. As underlined by the European Council objectives set in Lisbon, "effective integration of healthcare and related support services by electronic means, including the widespread use of telecare, could improve the quality of life of citizens by enabling safer independent living and increased social inclusion."

EVOLUTION FORCES AND CHALLENGES

Empowered patients demand advanced wireless health solutions. Similar to most authors, Lerer (2000) suggests that the e-health consumer is being empowered due to an increased ability to obtain health information and to seek health-related offerings via the Internet. A Deloitte Research (2000) study suggests the e-health consumer is a mix of an empowered and an engaged consumer. Recognizing that e-health consumers' empowerment can increase efficiency and reduce health costs, Lerer argues that consumers' education and empowerment should be a key concern for all health players. E-health consumers, he suggests, are not just the ill, but the potentially ill, the worried, and those adjacent to illness, patients, their relatives, and friends. At the first level, e-health services are information-driven activities, which are mostly "event triggered." The Deloitte Research study suggests that the demographic profile of the e-health consumer population reveals a significant population group with economic clout, information sophistication, and technological familiarity, and that is generally wealthy.

An e-health consumer is an individual who is (a) fully involved in the management of health for himself or herself and his or her family, (b) proactively educated about health issues, especially in the area of prevention, and (c) concerned about the quality of care offered by physicians and institutions, with a willingness to select the highest level of services. In short, an e-health consumer manages health, in all extents possible, as the most important asset of his or her family. The main objective is to maintain the highest level of quality of life (Lerer, 2000)

The rapid proliferation of wireless personal computers, phones, appliances, and other devices will require organizations to look beyond single-platform solutions. System-integration activities have a new level of complexity and cost to support rapidly changing technology (Wolf, 2001). Mobile-health advances generate new capabilities in patient self-care and health-practice administration and reimbursement. Cost-effective solutions minimize effort in monetary and human-input terms, while the creation of new communication modes facilitates both the healthcare professionals and the patients.

When it comes to investing in new technology solutions, affordability is a major milestone to consider. Budget allocation to mobile health applications can be easily influenced both by the technology cost and the user awareness of current and future cost benefits. The complexity and fragmentation of the overall healthcare sector (i.e., centralized vs. decentralized health systems, variations in the public and private funding mix, etc.) often leads to the implementation of fragmented and disposable technological solutions. Interoperability thus is essential for large-scale applications with international scope. Conformance to global (when available) and/or U.S. and European standards enables faster and ubiquitous communications, while also ensuring the compatibility and connectivity of systems and points of care.

According to CEN/TC 251 (2001), the present lack of standardized ICT communication, which prevents appropriate access to health records, may result in important clinical risks for the patients. This is an important safety issue that has not been recognized sufficiently.

Implemented standards are often crucial for any communication, and they are especially important for open, very complex healthcare systems with many different organizations and units, with information systems from different suppliers, providing different parts of the total ICT support.

Furthermore, the wider implementation of mobile solutions requires a robust security plan to reassure the confidentiality of sensitive medical data.

M-HEALTH POTENTIAL

The next few years will witness a rapid deployment of both wireless technologies and mobile Internet-based m-health systems with pervasive computing technologies. The increasing data traffic and demands from different medical applications and roaming applications will be compatible with the data rates of 3G (third-generation) systems in specific mobility conditions. The implementation and penetration of 4G (fourth-generation) systems are expected to help close the gap in medical care. Specifically, in a society penetrated by 4G systems, home medical care and remote diagnosis will become common, checkups by specialists and the prescription of drugs will be enabled at home and in underpopulated areas based on high-resolution image-transmission technologies and remote surgery, and virtual hospitals with no resident doctors will be realized. Preventive medical care will also be emphasized: For individual health management, data will constantly be transmitted to the hospital through a built-in sensor in the individual's watch or another item worn daily, and diagnosis results will be fed back to the individual (Istepanian et al., 2004)

A fourth-generation m-health solution builds upon the mobile information portal of a 3G solution by adding the multiple devices rendering the capability of the 2G (second-generation) solutions. Now, an end user has the ability to access any application with any device ("Going Mobile," 2001). 4G solutions embrace the distributed and loosely coupled HIS applications throughout a health unit. A 4G solution can allow for the acquisition of data from various sources and for the mobile end user to view, analyze, manipulate, graph, and merge data according to his or her needs right on the mobile device.

In the home of the future, some devices will contribute physiological information about the patient (e.g., heart rate, blood pressure), while other devices in and around the home will contribute information about the patient's environment (e.g., humidity, temperature, carbon-monoxide level). In some cases, groups of devices will have enough collective awareness to function autonomously based on sensor data.

The challenge for healthcare providers and health authorities lies in the comprehension of the end users' needs for the effective integration of new technological capabilities with existing settings in order to leverage their capacities and quality of service.

CONCLUSION

Systematically sensitizing users and providing them with specific information on new mobile and wearable computing technologies will help to discover possible fields of new applications. The initiation of a dialogue between users in healthcare and developers of mobile IT solutions eventually may lead to the identification of new application fields (i.e., medical specialties) and related practices in mobile healthcare provision.

A first step to this end is the identification and definition of mobile-activities profiles, and stakeholder profiles and their level of involvement, as well as mobile application scenarios. Technologies should be designed for people rather than making people adapt to technologies in order to capitalize on the capabilities that wireless technologies create in the healthcare domain.

REFERENCES

CEN/TC 251. (2001). *Report on the Mandate M/255.*

Deloitte Research. (2000). *The emergence of the e-health consumer.*

eHealth in 2010: Meeting the Lisbon objectives with ambient intelligence technologies? (2003). Seville.

Fosse, B., & Haug, B. E. (2003). *A feasibility study and recommendation of technology and solutions for wireless monitoring of biomedical data*. Unpublished master's thesis, Agder University College, Grimstad.

Going mobile: From eHealth to mHealth. A Daou Systems White Paper. (2001).

Istepanian, R., et al. (2004). Non-telephone healthcare: The role of 4G and emerging mobile systems for future m-health systems. In *Medical and Care Compunentics.* IOS Press.

Lerer, L. (2000). *The healthcare 2020 platform: The e-health consumer.* INSEAD.

McKnight, L. (2000). *Internet business models.* Medford, MA: Tufts University.

Simão, C. M. V. Q. (2001). *A study on Internet impact in business designs for the health sector.*

Wolf, J. (2001). *MBA mobile health applications, Version 1.1.*

KEY TERMS

3G and 4G: Third- and fourth-generation wireless Internet devices. The major distinction of 4G over 3G communications is increased data transmission rates. 4G is expected to deliver more advanced versions of the same improvements promised by 3G, such as enhanced multimedia, smooth streaming video, universal access, and portability across all types of devices. 4G enhancements are expected to include worldwide roaming capability and are likely to incorporate global positioning services (GPSs). As was projected for the ultimate 3G system, 4G might actually connect the entire globe and be operable from any location on or above the surface of the earth.

Ambient Intelligence: The concept of ambient intelligence provides a vision of the information society in which the emphasis is on user friendliness, efficient and distributed services support, user empowerment, and support for human interactions. People are surrounded by intelligent, intuitive interfaces that are embedded in all kinds of objects in an environment that is capable of recognizing and responding to the presence of different individuals in a seamless, unobtrusive, and often invisible way.

E-Health Consumer: Self-reliance and empowerment are the core characteristics of the e-health consumer, who actively pursues patient-centric quality services in a frame of information-supported activities.

Empowered Patient: A patient whose self-management is based on informed decisions and who takes into account his or her quality of life, including both physical well-being and psychological state, as well as other dimensions.

E-Wellness: The utilization of Internet capabilities (information, Web-based health services, etc.) in order to maintain a condition of good physical and mental health.

Medical Sensor: A device, such as a photoelectric cell, that receives and responds to a signal or stimulus.

M-Health: Mobile health refers to ambulatory-care provision enabled by third-generation devices that allow for the collection, management, and processing of the patient's vital data. Mobile health services range from the recording of the patient's medical signs and the synchronous or asynchronous communication with health professionals via mobile communication means, to the automatic diagnosis of the data recorded to personal sensors and alarm notices in case of an emergency. Mobile health or m-health is a step beyond electronic healthcare as it enhances ubiquitous health provision regardless of the patient's or physician's geographic location.

Telemonitoring: The science and technology of automatic measurement via medical sensors and the transmission of data by radio or other means from remote sources to receiving stations for recording and analysis. Data transfer can be achieved via wireless communications means and/or via other media, such as a telephone, a computer network, or an optical link.

Vital Signs: The pulse rate, blood pressure, body temperature, and rate of respiration of a person. The vital signs are usually measured to obtain a quick evaluation of the person's general physical condition.

This work was previously published in Handbook of Research on Informatics in Healthcare and Biomedicine, edited by A. A. Lazakidou, pp. 234-239, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.11
Geographic Information Systems in Health Care Services

Brian N. Hilton
Claremont Graduate University, USA

Thomas A. Horan
Claremont Graduate University, USA

Bengisu Tulu
Claremont Graduate University, USA

ABSTRACT

Geographic information systems (GIS) have numerous applications in human health. This chapter opens with a brief discussion of the three dimensions of decision-making in organizations — operational control, management control, and strategic planning. These dimensions are then discussed in terms of three case studies: a practice-improvement case study under operational control, a service-planning case study under management control, and a research case study under strategic planning. The discussion proceeds with an analysis of GIS contributions to three health care applications: medical/disability services (operational control/practice), emergency response (management control/planning), and infectious disease/SARS (strategic planning/research). The chapter concludes with a cross-case synthesis and discussion of how GIS could be integrated into health care management through Spatial Decision Support Systems and presents three keys issues to consider regarding the management of organizations: Data Integration for Operational Control, Planning Interorganizational Systems for Management Control, and Design Research for Strategic Planning.

INTRODUCTION

Geographic information systems (GIS) have numerous applications in human health. At the most basic level, entire research and practice domains within health care are strongly grounded in the spatial dimension (Meade & Earickson, 2000).

Indeed, the pioneering work of Dr. John Snow in diagnosing the London Cholera Epidemic of 1854 not only launched the field of epidemiology, but did so in a manner closely linked with the visual display of spatial information (Tufte, 1997). The health care enterprise has become much more complex since the time of Dr. Snow and so have the technologies that are employed to conduct spatial analysis regarding heath care conditions and services (Dangermond, 2000).

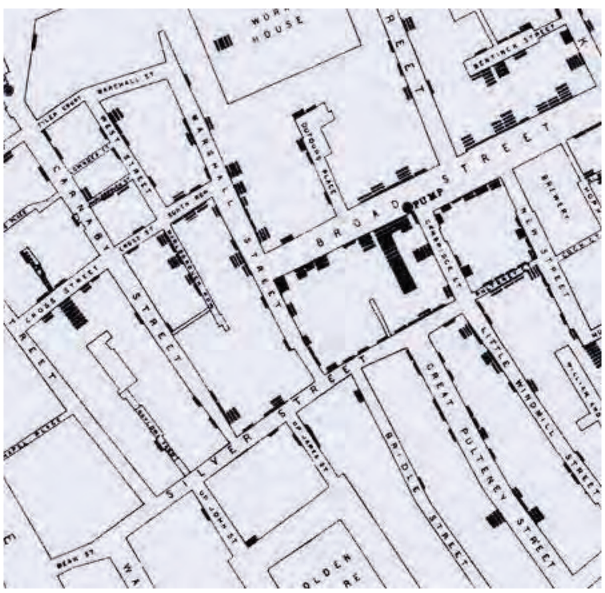
This chapter opens with a brief discussion of the three dimensions of decision-making in organizations — operational control, management control, and strategic planning. These dimensions are then discussed in terms of the case study focus of the chapter, which includes a practice-improvement case study under operational control, a service-planning case study under management control, and a research case study under strategic planning. The chapter proceeds with the analysis of GIS contributions to three health care applications: medical/disability services (operational control/practice), emergency response (management control/planning), and infectious disease/SARS (strategic planning/research). The chapter concludes with a cross-case synthesis and discussion of how GIS could be integrated into heath care management through spatial decision support systems.

BACKGROUND

One definition of a GIS is as "a group of procedures that provide data input, storage and retrieval, mapping and spatial analysis for both spatial and attribute data to support the decision-making activities of the organization" (Grimshaw, 2000, p. 33). One of the most well known models for thinking about the nature of these decision-making activities in the organization is Anthony's Model.

Anthony's Model implies a hierarchy of organizational decision-making. Here, a qualitative distinction is made between three types of decision-making: Operational Control, Management Control, and Strategic Planning (Ahituv, Neumann, & Riley, 1994). As GIS has developed,

Figure 1. John Snow's map of the Broad Street pump outbreak, 1854

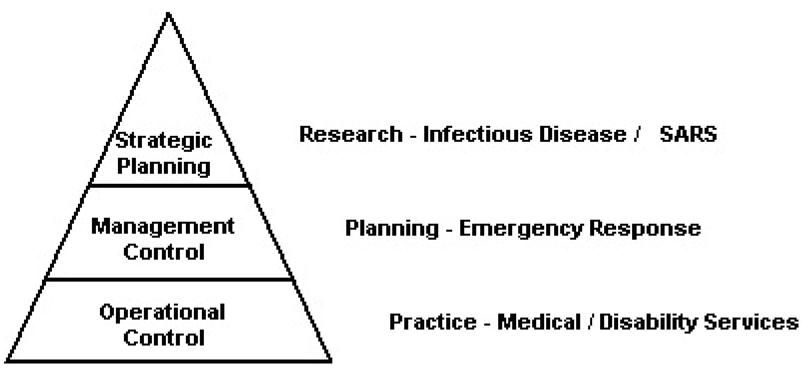


the range of applications for spatial data on human heath has grown dramatically (Cromley & McLafferty, 2002). In an effort to provide an in-depth understanding of these applications, this chapter considers three distinct application areas of GIS and Human Health: Practice, Planning, and Research. The combination of GIS and Human Health applications with the decision-making processes as defined in Anthony's Model is outlined below:

- **Operational Control** is the management of people, assets, and services using spatial information to ensure the delivery of the health care service while assuring that specific tasks are carried out effectively and efficiently. Our focus in this dimension is how spatial information can improve the *practic*e of health care.
- **Management Control** encompasses the management surrounding the health delivery system as a whole, and is specifically related to the needs and provisioning of health services, health promotion, disease prevention, and health inequalities while assuring that resources are obtained and used effectively and efficiently in the accomplishment of the organization's objectives. Our focus in this dimension is the use of spatial information to assist in the *planning* of health care services.
- **Strategic Planning** deals with the spatial distribution of diseases, their epidemiological patterns, and relation to environmental health risks and demographic characteristics while deciding on objectives of the organization, on changes in these objectives, on the resources used to attain these objectives, and on the policies that are to govern the acquisition, use, and disposition of these resources. Our focus in this dimension is how spatial-based *research* can affect the strategic design of health care delivery applications.

These combinations of GIS and Human Health applications and decision-making processes are used to present this particular series of case study summaries (*Figure 2*). The first case is an example of the *practice* of GIS regarding Disability Evaluation delivery at the Operational Control level. The second case is an example of *planning* regarding the use of a GIS for the delivery of Emergency Management Services at the Management Control level. The third case is an example of *research* regarding the conceptual design and development of a GIS as it relates to the National Electronic Disease Surveillance System at the Strategic Planning level.

Figure 2. GIS and human health decision-making and applications



RESEARCH METHODOLOGY

The research presented in this chapter draws on three case studies to illustrate various organizational scenarios in which a GIS was utilized, or could be utilized, to solve a particular Health Care Service problem.

Case Study

Case study methods can be used to explore the occurrence of a phenomenon with special attention to the context in which the case study is occurring. The most common working definition of this approach is offered by Robert Yin, who views case studies as "an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident" (1994, p. 13). Yin's defining work and related treatments further note that there are several uses of case studies:

- **Exploratory Value** – To uncover the nature of the phenomenon of interest. This can often serve as a precursor to more quantitative analysis.
- **Explanatory Value** – To help explain a phenomenon, such as when a quantitative study has revealed statistical association between variables but a deeper understanding of why they are related is missing.
- **Causal Value** – To provide a rich explanation of phenomenon of interest, including "patterns" that are not easily discernible through more abstract and/or numerical analysis.

Case studies have been used throughout the social sciences, as well as in business studies. Single-case design often uses the *extreme* or *unique* case to illustrate those phenomenon that are acutely visible such that inferences can be easily drawn that can be generalized to less extreme cases (Yin, 1994). Pare (2001) recently summarized the widespread use of case studies to examine the influence of information technology and systems in a variety of fields. Moreover, in his work with Elam, they note the promise of building a theory of IT through multiple case studies (1997). In a similar manner, this chapter uses three case studies in an exploratory fashion to enhance the understanding of IT usage, specifically GIS usage, within the context of health care.

For each case relevant literature was reviewed. For the first and second cases, the authors obtained original empirical information as part of separate studies. In the third case, a new and timely application at the strategic level is proposed.

Case Studies

Operational Control Case Study

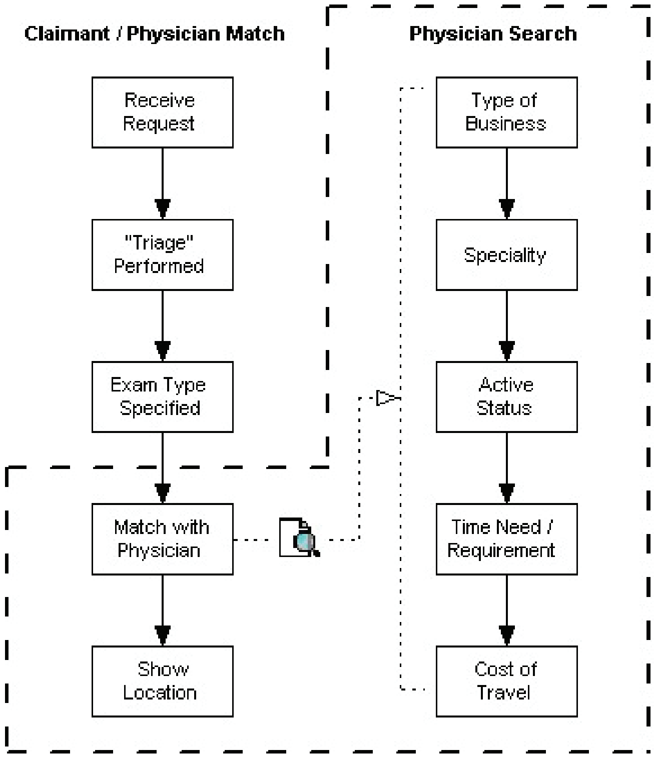
Background

One core practice area in medical services is the matching of patient/client services to providers (Cromley & McLafferty, 2002). The company in this case provides an array of disability evaluations, management, and information services nationwide. Of relevance to the subject of this chapter, this company (headquartered in Southern California) examined the use of a GIS to assist them in planning and marketing their disability evaluation services. With respect to Operational Control, this case study deals with appointment processing and the *practice* of ensuring the effective and efficient delivery of this service using spatial information.

Problem

The problem for this company was to provide an appointment for a claimant with a physician in a timely manner while meeting specific requirements and constraints. The existing workflow

Figure 3. Claim process



process was inadequate in meeting these requirements. *Figure 3* illustrates the workflow for this process, which begins when a case manager receives a request for an appointment. These requests are prioritized or "triaged" and the required exam sheets are generated using an expert knowledge base. Based on constraints such as physician specialty, availability, contract type, and location, an appointment is made for the claimant with the physician who is the "closest fit" (distance and travel-time) with minimal travel-time being the higher priority.

In this workflow process, the company was pleased with the efficiency and effectiveness of the first steps in the workflow process, which are computer-based. However, the last few steps were conducted using paper-based data sets and a number of Internet-based mapping websites (Yahoo Maps and MapQuest). As a result, these last few steps negatively impacted the amount of time a case manager interacted with a claimant, thereby increasing the company's costs in providing this service.

Solution

Using an Internet-based GIS application, a successful prototype was implemented for use within the company's Extranet. To develop this solution, a GIS planning process, also known as the GIS development cycle was employed (NYSARA & NCGIA, 1997). This process, illustrated in *Figure 4*, consists of a set of eleven steps starting with a needs assessment and ending with the on-going use and maintenance of the GIS system.

Outcome

Figures 5 through *8* provide an overview of the GIS based address-matching process that was

Figure 4. GIS development cycle

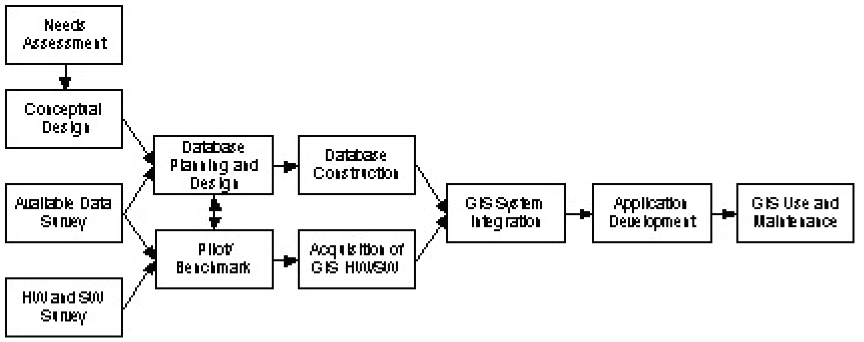
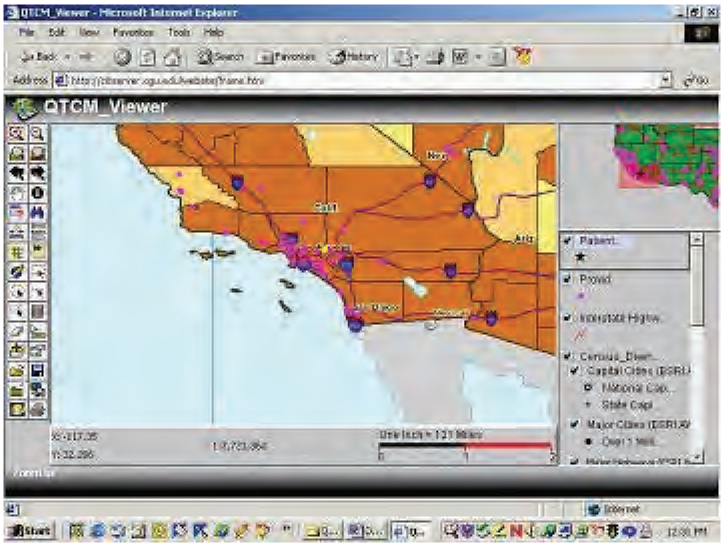


Figure 5. Step 1: Geocoding of claimant



developed. Now, when a case manager receives a request for an appointment they pinpoint the location of the claimant by "geocoding" the claimant's address (*Figure 5*). With the location of the claimant identified, the case manager then performs a physician attribute search, such as physician specialty, to identify only those physicians that meet the claimants' requirements (*Figure 6*). A physician spatial search is then performed to narrow down this list even further by locating only those physicians that are within a specified proximity to the claimant (*Figure 7*). Finally, a physician is chosen from this group that most closely matches the claimants' requirements and an appointment is made (*Figure 8*). The most important outcome of this case was a reduction in time required to set an appointment location and date. An additional beneficial outcome of this case was the development of a "live" connection between this system and the company's Oracle database. This database, which is updated on a daily basis, enables the company to exchange information between geographically distributed offices in real time. Consequently, users using the new GIS are now able to view the latest data, geocoded on a map, from any company location. Another important outcome of this case is the fact that this shared information is in a visual format.

Figure 6. Step 2: Physician attribute search

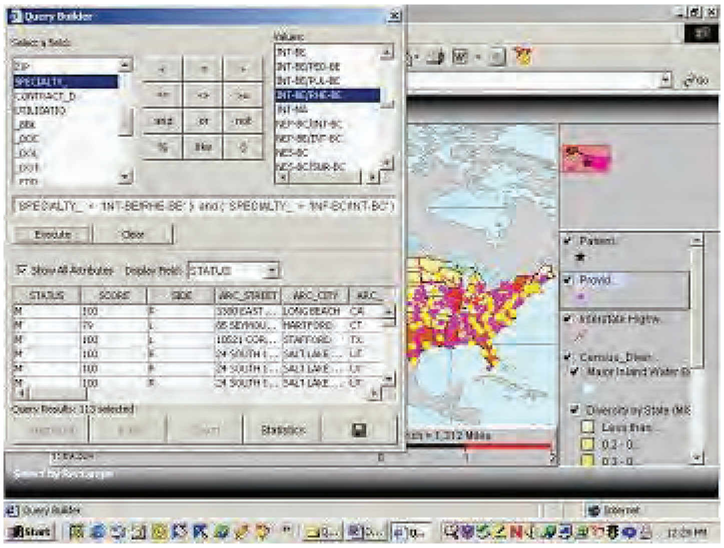
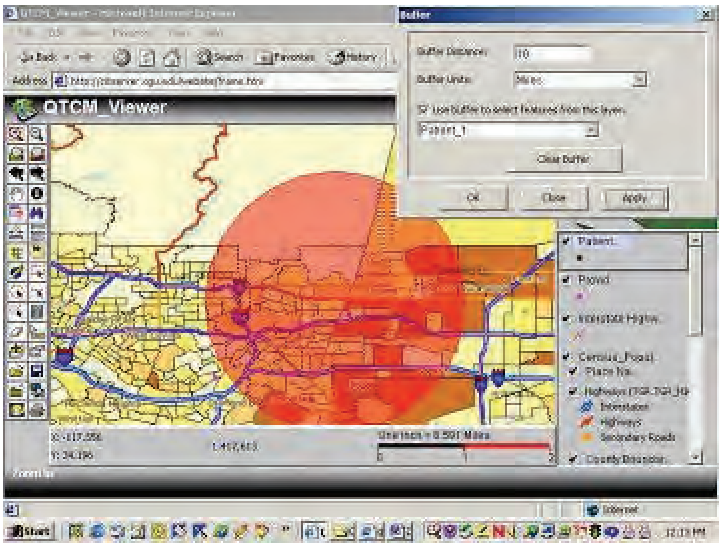


Figure 7. Step 3: Physician spatial search

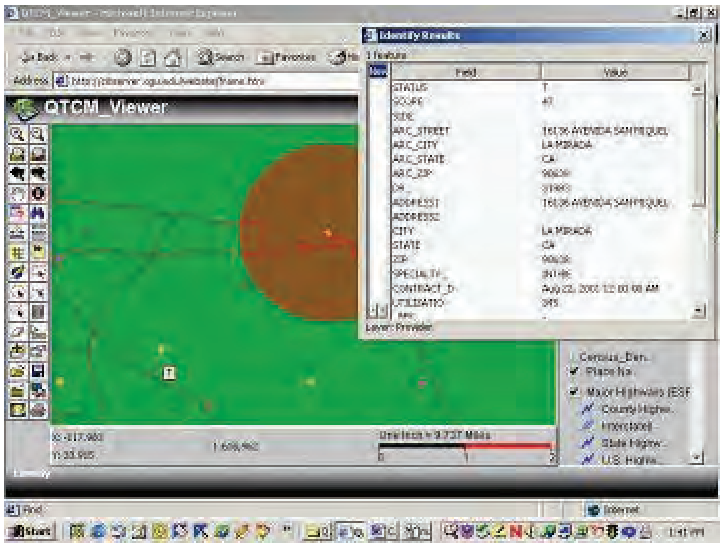


Management Control Case Study

Background

The United States Emergency 911 system, established more than 30 years ago, has become a cornerstone infrastructure for emergency management. However, the system is becoming increasingly stressed due to new wireless and digital communications technologies (Jackson, 2002; National Emergency Number Association, 2001). The original design could not anticipate the widespread use of mobile communications for emergency purposes seen today. Consequently,

Figure 8. Step 4: Final physician selection



this growth in wireless telecommunications is forcing the Emergency 911 infrastructure to change (Folts, 2002; Jackson, 2002; National Emergency Number Association, 2001). This case considers the broad development of Emergency Medical Service (EMS) systems, within the specific context of rural Minnesota. With respect to Management Control, this case encompasses the spatial properties surrounding the health delivery system as a whole, and is specifically related to the *planning* of E-911 services with the goal of assuring that wireless telecommunications services resources are obtained and used to accomplish the objectives of the State of Minnesota.

Problem

The advent of competitive sector telecommunications services in the wireless arena has played a pivotal role in the fast growth and use of the safety information network. Wireless phones have rapidly become one of our most effective tools in improving emergency response time and saving lives. A wireless 911 phone call can shave valuable minutes from the time otherwise required for a caller to find a conventional phone to access emergency medical services (Tavana, Mahmassani, & Haas, 1999). In the past 10 years, wireless phone use has grown exponentially. There are more than 120 million wireless users making approximately 155,000 emergency calls a day across the United States. The steady increase in private sector wireless subscribership and resulting mobile EMS use has created a need to better understand the implications of this rapidly growing system.

One illustration of the spatial challenges confronting EMS providers is the lack of location information regarding E-911 accessibility. The U.S. Federal Communications Commissions (FCC) has enacted mandatory requirements for wireless communications carriers to provide automatic location identification of a wireless 911 (E-911) phone call to an appropriate Public Service Answering Point (PSAP) (Federal Communications Commission, 2001). Both private carriers and public agencies are working closely together to overcome this difficult requirement. Although the technical requirements for building these systems have been thoroughly outlined, the execution of the service has materialized slowly (Christie et al., 2002; Zhoa, 2002).

One possible reason for this is the difficulty involved in committing to one of several viable technology alternatives to provide E-911. For example, one E-911 technology choice is a satellite-based system, which places a GPS-enabled chip in mobile phones along with location readers at the receiving point. With the current rate of technological change, selecting the one best solution, or combination of solutions, for long term system planning and investment is a difficult and daunting task for system administrators and designers (Proietti, 2002). The consequence of this situation is that deploying end-to-end E-911 systems will require new spatial technologies and nontraditional partnerships, particularly among wireless carriers, emergency dispatch center administrators (e.g., PSAPs), law enforcement, fire and EMS officials, automotive companies, consumers, technology vendors, and state and local political leaders (Jackson, 2002; Lambert, 2000; Potts, 2000). From a spatial perspective, the result is that location-based (E-911) services will be differentially deployed across regions, leading to a need to understand which areas are well serviced and which may require additional policy attention.

From a *planning* perspective, the need for special attention to rural areas is evident from the following statistics. According to the U.S. Department of Transportation, more than 56% of fatal automobile crashes in 2001 occurred on rural roads (National Center for Statistics and Analysis, National Highway Transportation Safety Administration, & U.S. Department of Transportation, 2002). The Minnesota Department of Transportation (MnDOT) reports that only 30% of miles driven within the state are on rural roads, yet 70% of fatal crashes occur on them (Short Elliot Hendrickson Inc. & C.J. Olson Market Research, 2000). In addition, 50% of rural traffic deaths occur before arrival at a hospital. Appropriate medical care during the "golden hour" immediately after injuries is critical to reducing the odds of lethal or disability consequences. Crash victims are often disoriented or unconscious and cannot call for help or assist in their rescue and therefore rely heavily upon coordinated actions from medical, fire, state patrol, telecommunications and other entities (Lambert, 2000).

Solution

The solution to the rural EMS program entails a combination of responses. These responses were analyzed within the context of a specific case study; an analysis of Minnesota's E-911 system (Horan & Schooley, 2003). The first activity in this case was to analyze the entire system and to construct an overall architecture of the system. This architecture, presented in *Figure 9*, illustrates Minnesota's EMS system along several key strata, technology, organizations, and policy, and identifies possible critical links (shaded gray) in the overall system. A summary of each layer follows.

- **Technology** – The top layer of the architecture illustrates some of the essential networks and communications technologies used by Minnesota EMS organizations to carry out their individual and interorganizational functions. From a GIS perspective, the GPS-equipped wireless devices and infrastructure to determine spatial location are critical elements.
- **Organizations** – The middle layer illustrates some of the public and private organizations involved in the Minnesota EMS and the general interorganizational relationships between these organizations. There is a significant geographic dimension to the organizational layer: each of the major stakeholders has distinct service boundaries (for example, there are 109 PSAPs, yet nine rural transportation operation centers).
- **Policy** – For EMS interorganizational relationships (i.e., partnerships, joint ventures, etc.) to succeed, policies need to be developed

that facilitate the interorganizational use of new and existing communications technologies. The overarching EMS technology-related policies, illustrated in the bottom layer, currently under development in the state are E-911 and 800 MHz radio. This includes the state-federal effort to develop standards and procedures for using location information received from mobile phones.

Outcomes

This case study raises several technological, organizational, and policy issues for planning EMS in rural Minnesota specifically, as well as for rural areas in general. The architecture highlights several critical areas that arose from this review and therefore have implications for future advancements. Areas, such as those denoted in gray in *Figure 9*, provide a focal point for discussing implications of this architecture for planning both EMS in general and GIS specifically.

GIS can assist greatly in understanding the extent of EMS coverage. Currently, this understanding is at a general level of detail, i.e., the level of compliance with new E-911 regulations. For example, *Figure 10* provides an example of the spatial dimension of E-911 deployment by a major provider. As displayed in this figure, the metropolitan region of Minneapolis has deployed location-identifying systems (e.g., E-911 Phase 2), while such systems have only been partially deployed in rural areas[2].

Figure 10 also provides an overview of the level of compliance with these new regulations, with lower compliance in rural areas. As shown in the figure, many regions in Minnesota are compliant with Phase 1 — the regulation requiring the provision of location-based information about mobile phones. From a service planning perspective, it will be important to monitor the spatial distribution of E-911 availability. Inadequate coverage in rural areas could give rise to the need for additional public policy regulations.

Figure 9. Interorganizational architecture for emergency management systems in Minnesota

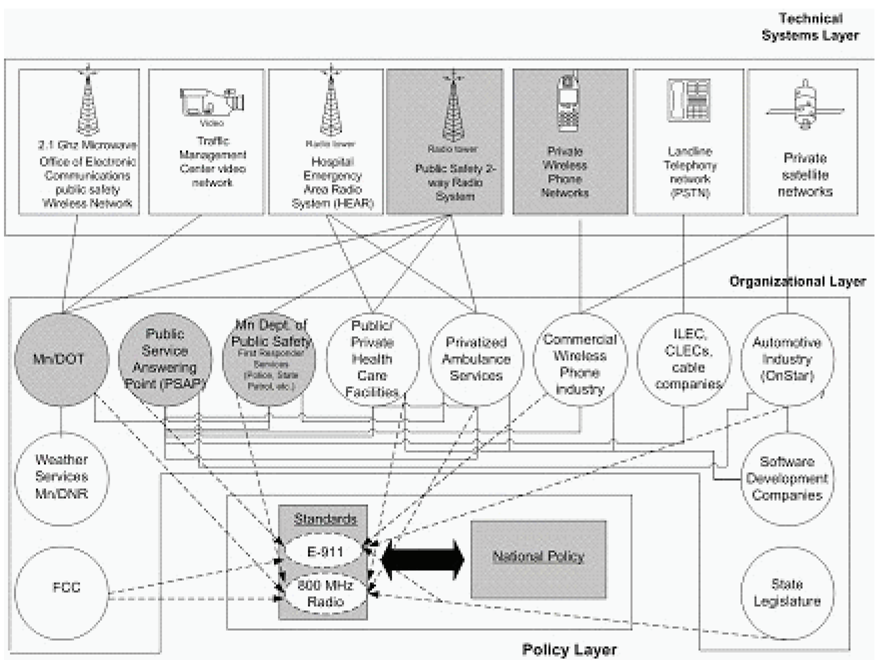
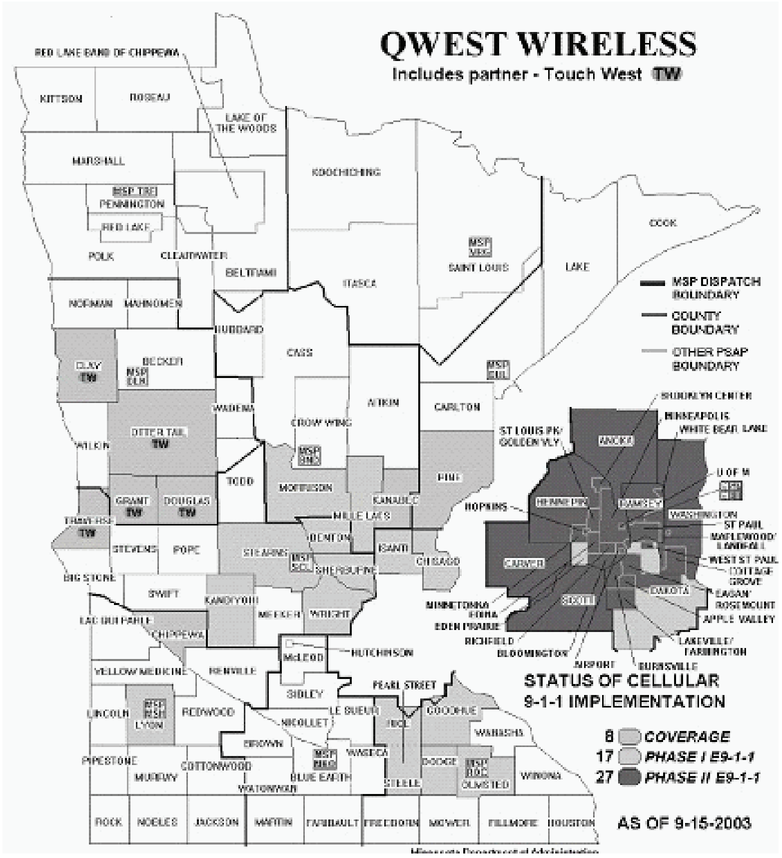


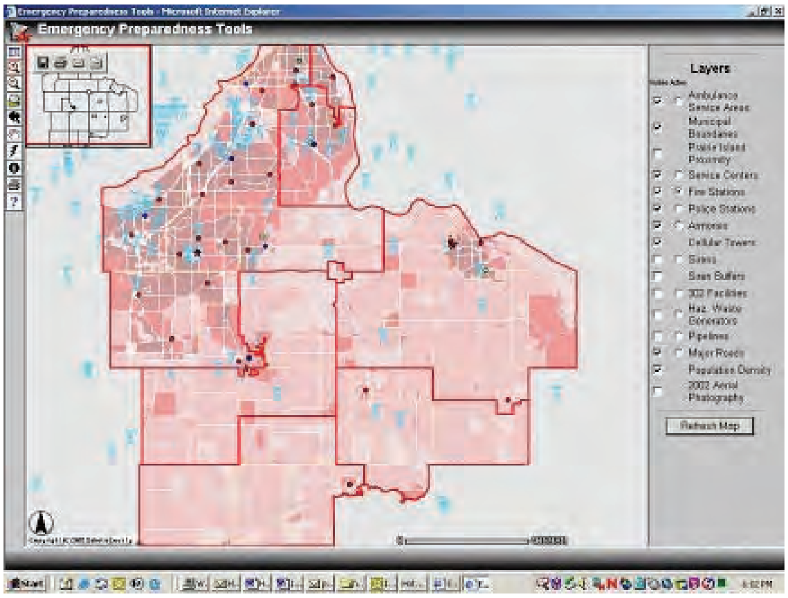
Figure 10. Spatial distribution of E-911 compliance status (Qwest)



The deployment of advanced 911 capabilities is however, only one aspect of integrated EMS services. Especially with recent concerns regarding homeland security, attention is now turning to how EMS is planned as part of an overall readiness strategy. In this context, GIS can play an important role in providing a spatial platform for EMS and related emergency services. One example of this is the GIS development work underway in Dakota County, Minnesota. This county, which includes significant rural as well as urban areas, has undertaken a comprehensive GIS-based approach to emergency preparedness with an Internet-based GIS platform that integrates both EMS related factors (e.g., Ambulance Service Areas and Cellular Towers) with other civic institutions involved in emergency preparedness (e.g., Fire Stations, Police Stations, and Armories)[3]. As illustrated in *Figure 11*, the Municipal Boundaries (outlined in bold) within Dakota County and the locations of Fire Stations, Police Stations, and Armories (dots) are identified. The locations of Ambulance Service Centers (dots) and Cellular Towers (tower symbol) are identified as well.

Platforms such as this represent a critical new dimension of interactivity whereby emergency management systems can be accessed dynamically and across institutions as well as infrastructure systems. Returning to the original architecture outlined above, this platform can be used by institutions such as the departments of transportation, public safety, and emergency services to facilitate cross-agency partnerships in service and technology deployments.

Figure 11. Internet-based GIS emergency preparedness application



In summary, this case demonstrates that the planning of emergency medical services is well suited to benefit from a dynamic GIS platform. Particularly in rural areas where resources are often scarce, such a platform can provide a common database to (1) monitor the spatial deployment of services, (2) facilitate resources sharing among institutions, and (3) provide a common understanding of "conditions" against which to plan for new technologies, systems, and policies.

Over time, it will be important to monitor the rate by which rural communities improve their "readiness." GIS can assist in this monitoring, including tracking of funding expenditures, regulatory compliance, etc. Finally, it is essential that methods of planning and analysis used to determine the form, level, and location of service and resource provision reflect the geographical components underpinning the health care system, i.e., the planning process should have an explicit geographical focus (Birkin, Clarke, Clarke, & Wilson, 1996).

Strategic Planning Case Study

Background

As noted in the introduction, spatial analysis and mapping in epidemiology have a long history (Frerichs, 2000), but until recently, their use in public health has been limited. However, recent advances in geographical information and mapping technologies and increased awareness have created new opportunities for public health administrators to enhance their planning, analysis and monitoring capabilities (World Health Organization, 1999b). Moreover, effective communicable disease control relies on effective disease surveillance where a functional national communicable diseases surveillance system is essential for action on priority communicable diseases (World Health Organization, 1999a). With respect to Strategic Planning, this case considers a *research* program regarding the spatial distribution of severe acute respiratory syndrome (SARS), its epidemiological patterns, and a theory to direct health organizations objectives and policies regarding the acquisition, use, and disposition of GIS.

The National Electronic Disease Surveillance System (NEDSS) program was initiated in the United States to provide an integrated, standards-based approach to public health surveillance and to connect surveillance systems to the burgeoning clinical information systems infrastructure (U.S. Department of Health and Human Services, 2002a). It is expected that the NEDSS will improve the nation's ability to identify and track emerging infectious diseases, monitor disease trends, respond to the threat of bio-terrorism, and other scenarios where the rapid identification of unusual clusters of acute illness in the general population is a fundamental challenge for public health surveillance (Lazarus et al., 2002).

To be effective with the rapid deployment of new health information systems, it is important to maintain effective mechanisms for rapid technology transfer to occur across government, academia, and industry (Laxminarayan & Stamm, 2003). The NEDSS program articulates an architecture that will enable public health information systems to communicate electronically, thereby decreasing the burden on respondents and promoting timeliness and accuracy (U. S. Department of Health and Human Services, 2002a). Stakeholders in the NEDSS include not only the Center for Disease Control (CDC) and other agencies within Department of Health and Human Services (DHHS), but state and local public health departments, healthcare providers, laboratories, health care standards organizations, health care product vendors, and healthcare professional services organizations.

The NEDSS Base System (NBS) is an instance of the NEDSS standards for use by all stakeholders that enable systems from different CDC program areas to be integrated. Disease-specific data and processes are incorporated and integrated within the NBS using Program Area Modules (PAM). Though this modular approach allows for the sharing of these data and processes, the NEDSS standards do not provide specific guidelines to support the development of these modules. Given that "GIS provides an excellent means of collecting and managing epidemiological surveillance and programmatic information, GIS represents an entry point for integrating disease-specific surveillance approaches" (World Health Organization, 2001, p. 7). An Information System Design Theory (ISDT) approach for the development of GIS-based PAMs would provide a useful guide for the various stakeholders in the NEDSS. This approach is detailed in the following section.

Problem

Severe acute respiratory syndrome (SARS) has emerged as a serious international occupational health disease. Since the disease was first reported it has infected numerous health care workers, some fatally (Centers for Disease Control and Prevention, 2003). Consequently, early identification of SARS cases is critical, as no specific treatment protocol exists. As a result, a GIS-based PAM for SARS is needed that could be used to manage the health care services required to combat this disease.

Unfortunately, current GIS-based attempts to mitigate the spread of this disease are lacking in respect to the NEDSS; that is, they lack an integrated, standards-based approach to public health surveillance and information dissemination[4]. The objective of this case study is to propose an ISDT such that organizations meeting the requirements of this theory will, by design, develop a GIS-based PAM compliant with NEDSS standards. Geographic information science has the potential to create rich information databases, linked to methods of spatial analysis, to determine relationships between geographical patterns of disease distribution and social and physical environmental conditions. As the core of a decision support system, geographic information science also has the potential to change the way that allocations of resources are made to facilitate preventive health services and to control the burden of disease (Rushton, Elmes, & McMaster, 2000).

Information System Design Theory Approach

While the overriding methodology in this chapter is the use of case studies, the case in this section is specifically concerned with the design of GIS as systems. As such, this analysis is informed by concepts deriving from a "design" approach to information systems. As noted in the ISDT approach introduced by Walls et al. (1992), the design process is analogous to the scientific method where hypotheses are to be tested by designing and building the artifact or product. They outlined several characteristics of design theories:

- Design theories are composite theories that encompass kernel theories from natural science, social science, and mathematics.
- While explanatory theories tell "what is," predictive theories tell "what will be," and normative theories tell "what should be," design theories tell "how to/because."
- Design theories show how explanatory, predictive, or normative theories can be put to practical use.
- Design theories are theories of procedural rationality. "The objective of the design theory is to prescribe both the properties an artifact should have if it is to achieve certain goals and the method(s) of artifact construction" (p. 41). Thus, the artifact must have all the characteristics identified in the design theory.

This case study explicitly draws upon their structure for creating a design product and process. Briefly, their first component of the design product involves a set of meta-requirements that describes the class of goals to which the theory applies. Their second component is a meta-design, which describes a class of artifacts hypothesized to meet the meta-requirements. Their third component is a set of kernel theories, theories from natural or social sciences governing design requirements. Their final component of the design product is a set of testable design product hypotheses that are used to test whether the meta-design satisfies the meta-requirements.

Beyond the design product, designers of GIS would necessarily need to be concerned with principles of the design process. According to the ISDT, the first component of the design process involves a design method which describes the procedures to be used in artifact construction (Walls, Widmeyer, & El Sawy, 1992). The second component is kernel theories, theories from natural or social sciences governing design process. The last component of the design process is the testable design process hypotheses, which are used to verify whether the design method results in an artifact that is consistent with the meta-design. While ISDT is explicitly applied in this third case study, the theme of effective design holds for all: that is, there is a design need to create GIS system designs in a manner that effectively contributes to the design goal of improving human heath.

Solution

The NEDSS has identified a number of enabling technologies for each element of the NEDSS technical architecture. One of these architectural elements is Analysis, Visualization, and Reporting (AVR). AVR capabilities support the epidemiological analysis of public health data and the communication of the analytical results of that analysis (U.S. Department of Health and Human Services, 2001). The specific requirements for this architectural element include tabular and graphical reporting, statistical analysis, and geographical information analysis and display. Also included are features such as the creation of pre-defined and ad-hoc reports, the ability to share results with colleagues and the public, and the extraction of data for use with standard analysis tools. The AVR Requirements are the foundation of the GIS Meta-Requirements as well as each of the specific GIS Meta-Design elements outlined in *Table 1.*

The Meta-Requirements and Meta-Design are derived from the OpenGIS Service Architecture (Open GIS Consortium Inc., 2002). Taken as a whole, these design elements would constitute the foundation of an ISDT for GIS-based PAMs.

Table 1. ISDT for GIS-based PAMs

AVR Requirements	GIS Meta-Requirements	GIS Meta-Design
Geographical information analysis and display Tabular reporting Graphical reporting	Geographic human interaction services	Geographic spreadsheet viewer: Client service that allows a user to interact with multiple data objects and to request calculations similar to an arithmetic spreadsheet but extended to geographic data. Geographic viewer: Client service that allows a user to view one or more feature collections or coverages. This viewer allows a user to interact with map data, e.g., displaying, overlaying and querying.
Creation of pre-defined and ad-hoc reports	Geographic model/information management services	Product access service: Service that provides access to and management of a geographic product store. A product can be a predefined feature collection and metadata with known boundaries and content, corresponding to a paper map or report. A product can alternately be a previously defined set of coverages with associated metadata.
Extraction of data for use with standard analysis tools	Geographic processing services – spatial	Subsetting service: Service that extracts data from an input in a continuous spatial region either by geographic location or by grid coordinates. Sampling service: Service that extracts data from an input using a consistent sampling scheme either by geographic location or by grid coordinates.
Extraction of data for use with standard analysis tools	Geographic processing services – thematic	Subsetting service. Service that extracts data from an input based on parameter values. Geographic information extraction services: Services supporting the extraction of feature and terrain information from remotely sensed and scanned images. Image processing service: Service to change the values of thematic attributes of an image using a mathematical function. Example functions include: convolution, data compression, feature extraction, frequency filters, geometric operations, non-linear filters, and spatial filters.
Extraction of data for use with standard analysis tools	Geographic processing services – temporal	Subsetting service. Service that extracts data from an input in a continuous interval based on temporal position values. Sampling service. Service that extracts data from an input using a consistent sampling scheme based on temporal position values.
Statistical analysis	Geographic processing services – metadata	Statistical calculation service: Service to calculate the statistics of a data set, e.g., mean, median, mode, and standard deviation; histogram statistics and histogram calculation; minimum and maximum of an image; multi-band cross correlation matrix; spectral statistics; spatial statistics; other statistical calculations.
Sharing of results with colleagues and the public	Geographic communication services	Transfer service: Service that provides implementation of one or more transfer protocols, which allows data transfer between distributed information systems over off-line or online communication media.

Outcome

As seen in this case, “GIS offers new and expanding opportunities for epidemiology as they allow the informed user to choose among options when geographic distributions are part of the problem, and when used for analysis and decision-making, they become a tool with a rich potential for public health and epidemiology" (Clarke, McLafferty, & Tempalski, 1996, p. 1). GIS-based PAMs developed using the proposed ISDT would allow the sharing of common data and processes while incorporating disease-specific data and processes. As seen in *Figure 12*, a SARS-specific PAM developed in this manner would become one of many PAMs and part of the larger NBS.

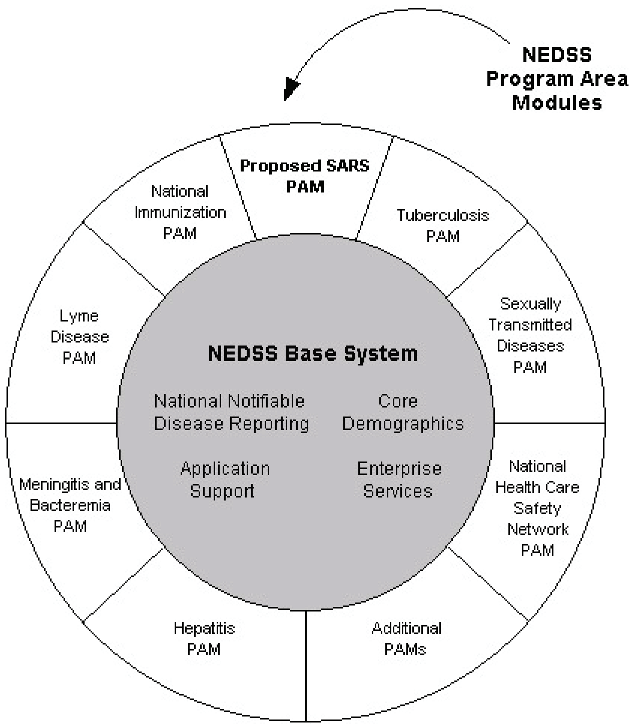
CASE STUDY SUMMARY ANALYSIS AND MANAGEMENT IMPLICATIONS

A summary of the three case studies is presented in *Table 2*. Each of the three cases presented is based on the health care service provided, specific problem type, solution generated, and final outcome.

As discussed in the introduction, the visual display of spatial information in healthcare, starting with epidemiology, has a long history. The literature reveals, as do these case studies, that the role of the private sector becomes increasingly more important as you move from the strategic planning level down to the management and operational levels. For those charged with the management of organizations, the following issues should be considered:

- **Data Integration for Operational Control:** Geospatial location is considered a valuable organizing principle for architecting and constructing enterprise data stores where interoperable geospatial technologies play a foundational role in exploiting these data

Figure 12. Conceptual picture of the NEDSS base system and program area modules



Adapted from U. S. Department of Health and Human Services (2002a) stores for enterprise missions (Open GIS Consortium Inc., 2003). As seen in the first case, this can occur at even the most basic level where demand from users at the operational level prompted the linkage of spatial and non-spatial data. As a consequence, the linkage of spatial and non-spatial databases will be among the challenges that organizations face as the spread of GIS applications and their extended functionalities drive organizations to integrate their existing structures with the increasingly important spatial dimension.

- **Planning Interorganizational Systems for Management Control:** Managing ad hoc inter-organizational networks are the nature of GIS data in health care environment. Emergency management applications are being developed in numerous areas around the country, ranging from instances where GIS-based emergency planning has been carried out by local governments, to cases where the Federal Emergency Management Agency has supported the creation and implementation of a GIS for this purpose (O'Looney, 2000). However, these diverse organizations collect spatial data for different purposes and goals. For successful planning, spatial data must be organized in such a manner as to appear as residing in one central database. Thus, the issue becomes how to successfully manage the relationship between these loosely coupled organizations that have dependence on tightly coupled systems without compromising their organizational purposes and goals.
- **Design Research for Strategic Planning:** The global implication of geographic data, as seen recently with SARS, indicates that it is no longer confined to the local or national level. Unfortunately, global geographic information is currently little more than the sum of the highly varied national parts and is not readily available (Longley, Goodchild, Maguire, & Rhind, 2001). The Global Spatial Data Infrastructure Association was formed to address this issue and is dedicated to international cooperation and collaboration in support of local, national, and international spatial data infrastructure

Table 2. Summary of case studies

Case Study	Health Care Service Provided	Problem	Solution	O utcome
Operational Control (Practice)	Disability Evaluation	Accuracy and time to schedule appointment location and date	Internet-based GIS application for use within company Extranet	Improved decision making and reduction in time required for scheduling physician appointments
Management Planning (Control)	Emergency Medical Service Delivery	Need for Ready EMS Delivery	Use GIS to provide efficient planning and management of E-911 health emergencies	Reduced EMS delay time and more rapid onset of trauma service delivery
Strategic Planning (Research)	Disease Surveillance System	Lack of Spatial Data Integration and Dissemination	Devise Inter-organizational Spatial Database for early disease detection	Program Area Module Design Theory for GIS

developments to allow nations to better address social, economic, and environmental issues of pressing importance (The Global Spatial Data Infrastructure Association, 2003). Their vision for a Global Spatial Data Infrastructure would support ready access to global geographic information. This vision would be achieved through the coordinated actions of nations and organizations through the implementation of complementary policies and common standards for the development and availability of interoperable digital geographic data and technologies to support decision-making at all scales for multiple purposes. The research challenge then is learning how to create a robust information system design process that considers spatial dimensions from global level international standard bodies to local level implementing organizations.

DISCUSSION

IS, and spatial analysis more broadly, can inform a number of pressing health care service issues, both domestically and internationally. From a domestic perspective this includes issues such as treating highly distributed populations and efficiently targeting critical services. From an international perspective, this includes developing a rapid means to detect and treat human health outbreaks in developed and developing regions as well as the creation of spatial datasets to drive international health enterprises.

Toward a Spatial Decision Support System (SDSS) for Human Health

While the geographic dimension of human health has long been recognized, the integration of GIS into the health care industry has not received the level of attention as that of more cutting-edge medical technologies. Nonetheless, as revealed in these cases as well as others that are increasingly appearing in the literature (Lang, 2000), GIS can provide a useful tool at the operational, management, and strategic levels of health care resources. For this reason, researchers such as Cromley & McLafferty (2002) have begun advocating for Spatial Decision Support Systems (SDDS) in support of health care services. Furthermore, a fairly long history of decision support systems in health care has been recognized (Shortliffe & Perrault, 2001). The new view is to integrate the spatial element into these systems. It is hoped that this chapter provides yet another indicator of the value in using spatial analysis for health care delivery.

ACKNOWLEDGMENTS

This chapter draws upon a series of research projects undertaken by the authors, as well as additional resources as appropriate. Findings related to the use of GIS and Disability Evaluations are based in part on the research conducted in collaboration with QTC, Inc. and have been reported in Hilton, Horan, & Tulu (2003). Findings related to the use of Emergency Management Systems and GIS are based in part on the research conducted in collaboration with the Humphrey Institute of Public Affairs, University of Minnesota and funded by the ITS Institute and have been reported in Horan, Kaplancali, & Schooley (2003) and Horan & Schooley (2003). The authors gratefully acknowledge the contributions of the following individuals to the chapter: Lee Munnich, Benjamin Schooley, and Dr. Subbu Murthy.

REFERENCES

Ahituv, N., Neumann, S., & Riley, H. N. (1994). *Principles of information systems for management* (4th ed.). Business and Educational Technologies.

Birkin, M., Clarke, G., Clarke, M., & Wilson, A. (1996). *Intelligent GIS: Location decisions and strategic planning.* New York: John Wiley & Sons.

Centers for Disease Control and Prevention. (2003). *Public health GIS news and information* (No. 52). Washington, DC.

Christie, J., Fuller, R., Nichols, J., Chen, A., Hayward, R., Gromov, K., et al. (2002). Development and deployment of GPS wireless devices for E911 and location based services. Paper presented at the *IEEE Position Location and Navigation Symposium*.

Clarke, K., McLafferty, S., & Tempalski, B. (1996). On epidemiology and geographic information systems: A review and discussion of future directions. *Emerging Infectious Diseases, 2*(2).

Cromley, E., & McLafferty, S. (2002). *GIS and public health.* New York: Guildford Press.

Dangermond, J. (2000). Preface. *In L. Lang, GIS in health organizations.* Redlands, CA: ESRI Press.

Federal Communications Commission. (2001). *Fifth report and order; First report and order: Implementation of 911 Act* (Electronic version No. WT-Docket No. 00-110 and CC Docket No. 92-105). Federal Communications Commission.

Folts, H. (2002, July). Standards initiatives for emergency telecommunications service (ETS). *IEEE Communications, 40,* 102-107.

Frerichs, R. (2000). History, maps and the Internet: UCLA's John Snow site. *The Society of Cartographers Bulletin, 34* (2), 3-8.

Paré, G., & Elam, J.J. (1997). Using case study research to build theories of IT Implementation. In J. I. DeGross (Ed.), *Information systems and qualitative research* (pp. 542-568). Chapman & Hall.

Grimshaw, D. (2000). *Bringing geographical information systems into business.* New York: John Wiley & Sons.

Hilton, B., Horan, T. A., & Tulu, B. (2003). GIS-enabled health services call center system. Paper presented at the *14th Annual Information Resources Management Association International Conference*, Philadelphia, Pennsylvania, USA.

Horan, T. A., & Schooley, B. (2003). Stranded but connected: Case study of rural deployment of emergency management services. *Journal of the Transportation Research Board, forthcoming.*

Horan, T. A., Kaplancali, U., & Schooley, B. (2003). Devising a web-based ontology for emerging wireless systems. Paper presented at the *Americas Conference on Information Systems*, Tampa, Florida, USA.

Jackson, A. (2002). *Recommendations for ITS technology in emergency medical services* (No. PATH Record Number 26202). Washington DC: Medical Subcommittee of the ITS America Public Safety Advisory Group.

Lambert, T. (2000). The role of ITS in the emergency management and emergency services community. In U. of M. Center for Advanced Transportation Technology, ITE, USDOT, ITSA (ed.), *Intelligent Transportation Primer.* Washington, D.C.: Institute of Transportation Engineers.

Lang, L. (2000). *GIS in health organizations.* Redlands, CA: ESRI Press.

Laxminarayan, S., & Stamm, B. H. (2003, June 6-7). Biomedical engineering evolutions in combating bioterorism. Paper presented at the *5th International Workshop on Enterprise Networking and Computing in Healthcare Industry*, Santa Monica, California, USA.

Lazarus, R., Kleinman, K., Dashevsky, I., Adams, C., Kludt, P., DeMaria, A., et al. (2002). Use of automated ambulatory-care encounter records for detection of acute illness clusters, including potential bioterrorism events. *Emerging Infectious Diseases, 8*(8), 753-760.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2001). *Geographic information systems and science*. New York: John Wiley & Sons.

Meade, M. S., & Earickson, R. J. (2000). *Medical geography* (2nd ed.). New York: Guildford Press.

National Center for Statistics and Analysis, National Highway Transportation Safety Administration, & U.S. Department of Transportation. (2002). *National center for statistics and analysis fatality analysis reporting system (FARS) web-based encyclopedia*. Retrieved July 21, 2002 from: *http://www-fars.nhtsa.dot.gov.*

National Emergency Number Association. (2001). *Report card to the nation: The effectiveness, accessibility and future of America's 9-1-1 service*. Columbus, Ohio: The RCN Commission and NENA.

NYSARA & NCGIA. (1997). *Geographic information system development guides.* Retrieved from: *http://www.sara.nysed.gov/pubs/gis/gisindex.htm*.

O'Looney, J. (2000). *Beyond maps: GIS and decision making in local government.* Redlands, CA: ESRI Press.

Open GIS Consortium Inc. (2002). *The OpenGIS abstract specification, topic 12: OpenGIS service architecture, version 4.3.* Open GIS Consortium, Inc.

Open GIS Consortium Inc. (2003). *OpenGIS reference model.* Open GIS Consortium, Inc.

Pare, G. (2001). *Using a positivist case study methodology to build and test theories in information systems: Illustrations from four exemplary studies.* Working Paper.

Potts, J. (2000, Summer). Wireless phone calls to 911: Steps toward a more effective system. *Currents.*

Proietti, M. (2002). Carrier choices in location: The system integrator's view. *GPS World, 13* (3), 23-28.

Rushton, G., Elmes, G., & McMaster, R. (2000). Considerations for improving geographic information system research in public health. *URISA Journal*, *12*(2), 31-49.

Short Elliot Hendrickson Inc., & C.J. Olson Market Research. (2000). *Advanced rural transportation information and coordination (ARTIC) operational test evaluation report*: Minnesota Department of Transportation (Mn/DOT).

Shortliffe, E. H., & Perrault, L. E. (2001). *Medical informatics: Computer applications in health care and biomedicine* (2nd ed.). New York: Springer-Verlag.

Tavana, H., Mahmassani, H., & Haas, C. (1999). Effectiveness of wireless phones in incident detection: Probabilistic analysis. *Transportation Research Record*, *1683*, 31-37.

The Global Spatial Data Infrastructure Association. (2003). *Global spatial data infrastructure: Strategic development plan*. Retrieved from: *http://www.gsdi.org/.*

Tufte, E. R. (1997). *Visual explanations: Images and quantities, evidence and narrative.* Cheshire, CT: Graphics Press.

U. S. Department of Health and Human Services. (2001). *NEDSS systems architecture version 2.0*. Washington, D.C.: Centers for Disease Control and Prevention.

U. S. Department of Health and Human Services. (2002). *NEDSS and NEDSS PAMS business discovery statement version 1.2*. Washington, DC.: Centers for Disease Control and Prevention.

Walls, J. G., Widemeyer, G. R., & El Sawy, O. A. (1992). Building an information system design theory for vigilant EIS. *Information Systems Research, 3*(1), 36-58.

World Health Organization. (1999a). *WHO Recommended surveillance standards* (2nd ed.). Department of Communicable Disease Surveillance and Response.

World Health Organization. (1999b). *WHO Recommended surveillance standards - annex 3: Role and use of geographic information systems (GIS) and mapping for epidemiological surveillance*. Department of Communicable Disease Surveillance and Response.

World Health Organization. (2001). *Protocol for the assessment of national communicable disease surveillance and response systems.* Department of Communicable Disease Surveillance and Response.

Yin, R. K. (1994). *Case study research, design and methods* (2nd ed.). Newbury Park: Sage Publications.

Zhoa, Y. (2002, July). Standardization of mobile phone positioning for 3G systems. *IEEE Communications, 40*, 108-116.

ENDNOTES

1 "Instead of plotting a time-series, which would simply report each day's bad news, Snow constructed a graphical display that provided direct and powerful testimony about a possible cause-effect relationship. Recasting the original data from their one-dimensional temporal ordering into a two-dimensional spatial comparison, Snow marked deaths from cholera on this map, along with locations of the area's 13 community water pump-wells. The notorious well is located amid an intense cluster of deaths, near the D in BROAD STREET. This map reveals a strong association between cholera and proximity to the Broad Street pump, in a context of simultaneous comparison with other local water sources and the surrounding neighborhoods without cholera" (Tufte, 1997. p. 30).

2 See *http://www.911.state.mn.us/911_enhanced.html*.

3 See *http://www.co.dakota.mn.us/gis/*.

4 SARS related GIS websites: *http://www.esrichina-hk.com/SARS/Eng/sars_eng_main.htm*, *http://www.sunday.com/Sunday/en/index.html*, *http://www.corda.com/examples/go/map/sars.cfm*, *http://www.cdc.gov/mmwr/preview/mmwrhtml/figures/m217a4f2.gif*, *http://www.mapasia.com/sars/default2.htm*, *http://www.info.gov.hk/dh/diseases/ap/eng/bldglist.htm*, *http://www.who.int/csr/sars/map2003_04_13.gif*, *http://spatialnews.geocomm.com/features/sars/*, *http://www.hku.hk/geog/hkgisa/sars.htm*.

This work was previously published in Geographic Information Systems in Business, edited by J. Pick, pp. 212-235, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 1.12
Information Technology (IT) and the Healthcare Industry:
A SWOT Analysis

Marilyn M. Helms
Dalton State College, USA

Rita Moore
Dalton State College, USA

Mohammad Ahmadi
University of Tennessee at Chattanooga, USA

ABSTRACT

The healthcare industry is under pressure to improve patient safety, operate more efficiently, reduce medical errors, and provide secure access to timely information while controlling costs, protecting patient privacy, and complying with legal guidelines. Analysts, practitioners, patients and others have concerns for the industry. Using the popular strategic analysis tool of strengths, weaknesses, opportunities, and threats analysis (SWOT), facing the healthcare industry and its adoption of information technologies (IT) are presented. Internal strengths supporting further industry investment in IT include improved patient safety, greater operational efficiency, and current investments in IT infrastructure. Internal weaknesses, however, include a lack of information system integration, user resistance to new technologies and processes, and slow adoption of IT. External opportunities including increased use of the Internet, a favorable national environment, and a growing call for industry standards are pressured by threats of legal compliance, loss of patient trust, and high cost of IT.

INTRODUCTION

The healthcare industry faces many well-recognized challenges: high cost of operations, inefficiency, inadequate safety, insufficient access to information, and poor financial performance. For years, many have called for a fundamental change in the way healthcare is delivered. And while there is yet no clear picture of what this change will be, many believe a paradigm shift in healthcare is imminent and that information technology (IT) is the catalyst.

Increasingly, IT is seen as a way to promote the quality, safety, and efficiency of healthcare by bringing decision support to the point of care, providing vital links and closing open loop systems, and allowing routine quality measurement to become reality. IT can not only reduce operating costs, but IT can also ensure a reduction in the number of medical errors. IT in the healthcare industry provides new opportunities to boost patient confidence and reinforce patient trust in caregivers and healthcare facilities. With health insurers feeling pressure from all directions (new regulations, consumers, rising medical costs), IT is an even more important asset for carriers (Balas, 2000).

When compared to other information intensive industries, healthcare organizations currently invest far less in IT. For many years, the healthcare industry has experienced only single digit growth in terms of IT investment (Gillette, 2004). As a result, current healthcare systems are relatively unsophisticated compared to those in industries such as banking or aviation. With the many issues and variables surrounding healthcare's IT investment, a framework for better understanding of the current situation is needed before more improvements and enhancements can result. This article draws upon a comprehensive framework from the strategic planning literature to compile and summarize the major issues facing IT and the healthcare industry.

METHODOLOGY

By categorizing issues into strengths, weaknesses, opportunities, and threats, SWOT analysis is one of the top tools and techniques used in strategic planning (see Glaister & Falshaw, 1999). SWOT assists in the identification of environmental relationships as well as the development of suitable paths for countries, organizations, or other entities to follow (Proctor, 1992). Valentin (2001) suggests SWOT analysis is the traditional means for searching for insights into ways of crafting and maintaining a fit between a business and its environment. Other researchers (see Ansoff, 1965; Porter, 1991; and Mintzberg, Ahlstrand, & Lampel, 1998) agree SWOT provides the foundation to gather and organize information to realize the desired alignment of variables or issues. By listing favorable and unfavorable internal and external issues in the four quadrants of a SWOT analysis, planners can better understand how strengths can be leveraged, realize new opportunities, and understand how weaknesses can slow progress or magnify threats. In addition, it is possible to postulate ways to overcome threats and weaknesses (e.g., Hofer & Schendel, 1978; Schnaars, 1998; Thompson & Strickland, 1998; McDonald, 1999; and Kotler, 2000).

SWOT has been used extensively to aid in understanding a variety of decisions and issues including: manufacturing location decisions (Helms, 1999); penetration strategy design for export promotions and joint ventures (Zhang & Kelvin, 1999); regional economic development (Roberts & Stimson, 1998); entrepreneurship (Helms, 2003); performance and behavior of micro-firms (Smith, 1999), and strategic planning (Khan & Al-Buarki, 1992). Hitt, Ireland, Camp, and Sexton (2001) suggest that identifying and exploiting opportunities is part of strategic planning. Thus, SWOT analysis is a useful way to profile the general environmental position of a new trend, technology, or a dynamic industry.

By using SWOT analysis, it is possible to apply strategic thinking toward the implementation of IT in healthcare. By examining the internal and external factors interacting both for and against IT in healthcare, healthcare providers and supply chain organizations can formulate a strategic IT plan for developing their information resources over the next several years. By uncovering and reviewing the issues, policy makers can enact changes to make the process of IT implementation easier while simultaneously working to change the culture to foster IT benefits for institutions and the patients and other stakeholders they serve.

In SWOT analysis, strengths act as leverage points for new strategic initiatives, while weaknesses are limiting factors. Specifically applied to IT in healthcare, strengths should indicate areas where either IT or healthcare is particularly strong, (i.e., the technical skills of IT professionals or the quality of existing healthcare information systems). Weaknesses should display areas where either IT or healthcare requires improvement, and may range from personnel issues within IT to limited healthcare applications beyond routine transaction processing. The threats and opportunities identified during the external analysis should be both factual and attitudinal issues that must be addressed in any strategic plan being formulated, and should include both healthcare and IT issues (Martin, Brown, DeHayes, Hoffer, & Perkins, 2005). The following section presents the internal strengths and weaknesses currently confronting the implementation and proliferation of IT in healthcare.

INTERNAL STRENGTHS

Improved Patient Safety

Patient safety, as expressed in the Hippocratic Oath (Classical Version)—"I will keep them from harm and injustice"—is an underlying principle of professional healthcare throughout the world. Improving patient safety is a primary objective at all levels of the healthcare industry. The strategic initiative to increase the role of IT in healthcare can advance the cause of greater patient safety by enhancing the quality of that care. With comprehensive data available in a timely manner, healthcare providers can make better decisions about their patients' care, thereby reducing errors due to incomplete or insufficient information at the point of decision (Goldberg, Kuhn, & Thomas, 2002). Lenz (2007) agrees IT has a huge potential to improve the quality of healthcare and that this aspect has not been fully explored by current IT solutions. Advanced process management technology is seen as a way to improve IT support for healthcare processes by improving the quality of those processes.

The Joint Commission on Accreditation of Healthcare Organizations (JCAHO) established the Healthcare Information Technology Advisory Panel in 2005 to focus attention on the improvement of patient safety and clinical processes as new healthcare information systems are implemented. Members of the panel include researchers, physicians, nurses, chief information officers, educators and leaders of healthcare organizations, as well as representatives from the Office of the National Coordinator for Health Information Technology, the American Health Information Management Association, the Agency for Healthcare Research and Quality, the Veterans Health Administration, and the Healthcare Information and Management Systems Society. The panel was formed to recommend ways JCAHO's accreditation process and the widespread use of technology can be used to help re-engineer patient care delivery and result in major improvements in safety, quality and efficiency. The panel was also charged with the task of examining such topics as the effect of electronic health records on performance benchmarking and public reporting capabilities. Based on the panel's recommendations, JCAHO will evaluate its strategic plan and future direction relative to healthcare information technology (Anonymous, 2005b).

Two examples of existing information technologies and computer-based information systems contributing to improvements in patient safety through better quality of care and reduction of errors are smart cards (and/or compact discs) and computerized physician order entry (CPOE) systems. Smart cards containing a patient's entire medical history can be designed to be accessible only by devices in a hospital, doctor's office, or other medical facility. They not only eliminate the problems of lost and comprised hard copies of patient records, but also enable more secure electronic transfers of patient information to other healthcare providers and insurers (Anonymous, 1997). The technology for Java-based cards can securely support applications for multiple health-care facilities and be combined with biometric measures for identification purposes (Sensmeier, 2004). With complete information available, physicians are able to make better decisions for the care of the patient, and order appropriate tests and treatments (Goldberg et al., 2002).

Several studies indicate medication errors are the most likely type treatment error to occur because drug therapy is one of the most widely used interventions in healthcare (Kohn, 2001). Computerized physician order entry (CPOE) systems eliminate transcription errors and can warn of allergies and drug interactions. Such systems reduce errors by more accurately dispensing the correct dosage of the correct medication for the correct patient (Bates, Teich, Lee, Seger, Kuperman et al., 1999;Kuperman, Teich, Gandhi, & Bates, 2001; Mekhjian, Kumar, Kuehn, Bentley, Teater et al., 2002; Order Entry Rules, 2002; Scalise, 2002; Shane, 2002).

Computerized Physician Order Entry (CPOE) systems reduce medication errors by 80%, and errors with serious potential patient harm by 55% (Bates et al., 1999).

In pharmaceuticals, e-prescribing, or the electronic transmission of prescription information from the prescribing physician to a pharmacist can reduce medical errors. Since Americans receive more than three billion prescriptions yearly and pharmacists have to call these physicians 150 million times a year because they cannot read or understand the prescriptions, e-prescribing could reduce injuries from medication errors (Brodkin, 2007).

Technology-enabled improvements could also aid disease prevention and management. Other benefits could include lowering age-adjusted mortality by 18% and reducing annual employee sick days. Lieber (2007) stresses shared experiences regarding pandemic diseases can provide the best solutions and this is aided by IT solutions and global information exchange.

When medical records are available electronically, patients too can have access to their personal health records. Five large U.S. employers have funded an institute where their current and retired employees and their families can have access to and maintain their lifelong personal health records (Five Large Companies, 2007). With access to longitudinal and comprehensive records, patient safety can continue to improve.

A recent example of IT and improved patient safety is the North Mississippi Medical Center (NMMC) in Tupelo, Mississippi. Serving 24 rural counties, NMMC is the largest rural hospital in the country and the 2006 winner of the Malcolm Baldrige National Quality Award in the Healthcare Category. NMMC's recognition was due largely to their success in utilizing IT. Patients' electronic medical records can be accessed by nurses, by partner community hospitals, by physicians in their offices, and even by specialists and primary care providers in remote sites, reducing medical errors and duplication of effort. These enhancements have earned NMMC the distinction of one of the most wired facilities in the country. The organization has a shared radiography information system for all its hospitals and clinics which reduces report preparation time. For example, patients can have a radiology procedure, see their doctor, and obtain their results the same day (Baldrige Award Recipient, 2006; Anderson, 2007).

Greater Efficiency of Operation

Information technology, or the digital world of bits and bytes, delivers information faster, smarter, and cheaper (Conger & Chiavetta, 2006). In healthcare, IT has improved operational efficiency and increased productivity by reducing paperwork, automating routine processes, and eliminating waste and duplication. Lieber (2007) reports the use of electronic health records could save as much as $8 billion yearly in California alone through improvements in delivery efficiency.

Picture archival and communication systems (PACS) not only save providers' costs for file room, storage space and film supplies, but also decrease time spent reporting, filing and retrieving records. Web access enables physicians to view radiological images from their offices, homes, or other remote facilities. IT provides emergency rooms with tools for electronic prescriptions, order entry, provider documentation, and aftercare instructions for patients and their family. Updating electronic instructions is quick and easy. Purchasing departments are aided by the ability to buy products for specialty areas, such as anesthesia, infection control, substance-abuse programs, and home healthcare. Increased productivity and positive return on investment are seen in many areas of IT and are continuously improving (Parker, 2004b).

Three other important technological advances for improved productivity are voice-technology systems, two-way communications systems, and radio-frequency identification (RFID). Voice-technology systems can significantly reduce the time nurses and admissions personnel spend on pre-authorizations and pre-certifications required by third-party payer plans. Within six months of implementing a phone-based voice-technology system, Erlanger Hospital in Chattanooga, Tennessee, reported a greater than 50% drop in phone transaction times. Moreover, in May, 2006, after four years in operation, Erlanger reported a total payback of $920,201, decreased percentages in days denied, and the reassignment of three full-time equivalent employees to other departments within the hospital (Bowen & Bassler, 2006). Automated two-way communications systems for scheduling can greatly improve workflow by managing automatic appointment reminders, waiting lists, and cancellation notices. These systems call patients, and in a pleasant voice, remind them of a doctor's appointment, ask them to confirm their intention to keep the appointment, and report the information to the provider's office (Sternberg, 2005). RFID technology, particularly in the areas of human and material resources, offers healthcare facilities a way to measure and control their resources as well as the relevant workflow processes (Janz, Pitts, & Otondo, 2005).

Some healthcare providers suggest Emergency Department Information Systems (EDIS) can improve operations efficiency in this extremely time-critical area by facilitating the flow of patients through the emergency department, eliminating redundant patient records, promoting information sharing, and providing quicker access to laboratory test results and radiology films (Parker, 2004a). Because electronic medical records allow tracking of patients' conditions and medications, emergency room providers and hospitals have immediate access to detailed information; both patients and providers have a better sense of what occurred and when. Both groups also report increased satisfaction with the process. Interoperability of systems also makes patient information available across budgetary and functional units, thereby providing greater continuity of patient care (Cohen, 2005).

Lopes (2007) agrees it is more efficient when an internal medicine physician can consult with a cardiologist electronically while viewing a patient's medical work-up and history. Such systems create efficiencies, have a positive return on investment, and there are no misfiled lab results.

IT tools and software are being developed to meet the growing needs of the healthcare community. As an example, VHA, Inc. recently introduced an updated version of a Comparative Clinical Measurement tool to give member hospitals flexibility in both collecting and reporting clinical improvement data. VHA, Inc. has worked with the Joint Commission on Accreditation of Healthcare Organizations to include their measures into the new tool (VHA, 2007).

Current Investment in IT

Is there a hospital in the United States that has not already made an investment in their IT infrastructure? Probably not. In the past ten years, advances in health information technologies have occurred at an unprecedented rate and healthcare organizations have responded by increasing their IT investments "threefold" (Burke & Menachemi, 2004). Today, albeit at varying levels of sophistication, all hospitals use IT to run their core administrative and clinical application systems, that is, patient accounting, insurance billing, human resources, staff and facilities scheduling, pharmacy, laboratory results reporting, and radiology (Cohen, 2005). Most healthcare organizations in the U.S. are spending between 2.1% and 10% of their capital operating budget on IT (Conn, 2007c). Within this spending on IT, healthcare providers cite electronic health record development as a top priority followed by development and implementation of clinical IT systems to improve patient care capability (Conn, 2007b).

According to the U.S. Department of Health and Human Services (HHS), approximately 13% of the nation's 4,000+ hospitals use electronic medical records and 14% to 28% of the 853,000 U.S. physicians are wired (Swartz, 2005). A recent study reported, on average, hospitals have acquired 10.6 clinical application systems, 13.5 administrative application systems, and 50.0 strategic application systems (Burke & Menachemi, 2004). Some healthcare organizations including the Cincinnati Children's Hospital, Baylor Healthcare System in Dallas and The Heart Center of Indiana are going beyond their core systems to develop, acquire, and integrate applications for decision support, benchmarking, facilities management, and workflow processes (Cohen, 2005; Kay & Clarke, 2005).

If predictions are correct, by 2015 the electronic medical records market is expected to grow to more than $4 billion, up from $1 billion in 2005. The Kalorama Information Research firm, after studying the healthcare, diagnostics, pharmaceuticals, and medical devices markets, predicts the surge will be led by the increase in IT budgets of hospitals, physicians' offices and other U.S. healthcare organizations (Study: U.S., 2007).

Healthcare organizations, having already made investments in their computing and communications hardware and software, application software, and personnel, can leverage their existing IT investments as they expand their IT infrastructure to meet growing demands to achieve more efficient operations and more effective levels of healthcare. Whether or not the IT investments meet the ROI goals of financial departments, hospitals are going to implement IT, according to a survey by the Healthcare Information and Management Systems Society. Some 88% of hospitals have adopted electronic medical records and 24% already have them in place. Some 36% are implementing them and 28% have plans to. Only 12% lack IT plans (Greene, 2007).

INTERNAL WEAKNESSES

Lack of System Integration

Integrated systems offer seamless data and process integration over diverse information systems (Landry, Mahesh, & Hartman, 2005). Since a patient's treatment involves receiving services from multiple budgetary units in a hospital, information system integration should exist between the computer-based applications within a single hospital. When healthcare organizations coordinate and integrate their internal data, they can improve operations and decision making; however, most healthcare organizations are not yet at this level of system integration. Clinical, administrative, and financial systems are not linked, and as a result, many healthcare institutions are not yet maximizing their IT potential (Cohen, 2005).

Moreover, system integration need not be confined to applications within a single facility. There are many types of healthcare providers and healthcare-related agencies in the complex healthcare network. Since a patient's treatment usually involves receiving services from multiple providers and interacting with various other healthcare-related entities, information system integration should also exist between the computer-based applications of those separate agencies. Immediate benefits of information sharing between different agencies include the elimination of duplicate work in gathering and inputting the data, the immediate availability of the information, a lower probability of error, and greater convenience for the patient. Called a "vision of unsurpassed information technology integration," The Heart Center of Indiana, a joint venture between St. Vincent Health, The Care Group, and CorVasc, reports improved quality of care at lower costs through its IT partnership (Kay & Clarke, 2005).

System integration between agencies could also take increased efficiency to the industry or national level. A report issued by the Foundation of Research and Education of the American Health Information Management Association supports a fully integrated fraud management system which it believes could help address the growing problem of healthcare fraud (Swartz, 2006a).

User Resistance

User resistance, more commonly termed *user acceptance* in the information systems literature, is nothing new to IT. The original Technology Acceptance Model (TAM) put forth by Davis (1989) states a user's level of system acceptance is explained by two factors: the system's perceived usefulness and its perceived ease of use. Perceived usefulness is defined as the degree to which a person believes that using a particular system would enhance job performance, while perceived ease of use is defined as the degree to which a person believes that using a particular system would be free of effort. Subsequent research across a variety of research settings confirms perceived usefulness as the strongest predictor of user acceptance (Adams, Nelson, & Todd, 1992; Taylor & Todd, 1995; Venkatesh & Davis, 1996; Mahmood, Hall, & Swanberg, 2001). Some believe that IT implementations in the healthcare environment, however, encounter more resistance than in any other environment (Adams, Berner, & Wyatt, 2004).

The healthcare-related literature suggests physician resistance is a key weakness existing in the doctor's office and at the hospital. Consistent with Davis' (1989) principle of perceived usefulness, physician acceptance of new IT systems at the hospital is linked to the system's impact on patient safety (Rhoads, 2004), while physician acceptance of new IT at their office largely depends on cost (Chin, 2005). Healthcare literature suggests nurse acceptance of new IT has steadily improved as applications demonstrate increased support of the practice of nursing and improvement of patient safety resulting from the reduction of human error (Sensmeier, 2005; Simpson, 2005).

A study of 12 critical access hospitals found barriers to health information technology included funding, staff resistance to change, staff adaptation to IT and workflow changes. Other user resistance was noted by the time constraints on small staff, facility and building barriers, and lack of appropriate IT support. While all agree that IT will improve safety and reduce errors, barriers to implementation are numerous and must be addressed (Hartzema, Winterstein, Johns, de Leon, Bailey, McDonald, & Pannell, 2007).

Slow IT Adoption

Traditionally, healthcare has been slow to adopt IT and has lagged significantly behind other industries in the use of IT (Ortiz & Clancy, 2003; Adams et al., 2004). A 2005 report from the National Academy of Engineering and the Institute of Medicine agrees healthcare's failure to adopt new strategies and technologies has contributed to the list of problems now associated with the industry: thousands of preventable deaths a year, outdated procedures, billions of dollars wasted annually through inefficiency, and costs rising at roughly three times the rate of inflation. Lack of competition, resistance to change, and capital costs are among the major causes for healthcare's slowness to adopt IT (Hough, Chen, & Lin, 2005).

There are signs of progress, however, which offer promise of accelerated change. Many hospitals and physician groups are now digitizing their medical records and clinical data (Hough et al., 2005). As noted earlier, some hospitals like Cincinnati Children's Hospital, Baylor Healthcare System in Dallas, and The Heart Center of Indiana have adopted IT at advanced levels (Cohen, 2005; Kay & Clarke, 2005). These hospitals are models for the industry, forging a path for other healthcare organizations to follow, and emerging as healthcare leaders in IT whose techniques can be benchmarked, emulated and implemented. As the healthcare technologies are developed to greater sophistication and functionality, it will be possible for other healthcare organizations to "leapfrog" over the slow, expensive evolutionary learning process experienced by the leaders (Conger & Chiavetta, 2006).

The following section outlines external opportunities and threats facing IT and healthcare. Specific opportunities are the Internet, the national environment, and industry standards. Key threats include legal compliance, loss of patient trust, and the costs of IT systems, training, implementation, and support.

EXTERNAL OPPORTUNITIES

The Internet

Across the industry, healthcare facilities and providers are in various stages of incorporating the Internet into their operations to allow new ways to communicate with the general public, specific patients, patient groups, physicians, other providers, and employees. Notable Web-based services include public Web sites, various telemedicine applications for targeted patient audiences, physician portals, physician education sites, and facility intranets which serve an organization's internal audiences. Generally, there is an increased focus throughout the healthcare industry to improve all Web-based applications (Sternberg, 2004).

Through their public Web sites, hospitals and other healthcare agents provide medical information to the general public (Natesan, 2005). E-Health Web portals offer healthcare services and education to people with chronic conditions and to their caregivers (Moody, 2005). Through various telemedicine initiatives, the healthcare industry has reached significant numbers of people living in rural areas, providing access to expert advice and reducing their health risks (Harris, Donaldson, & Campbell, 2001). Web-based patient support systems educate patients and allow them better participation in their own care. Patients can research detailed information for their particular conditions, medications, and treatments to understand what is happening and to reduce their anxiety. Online surgery videos and graphics can be presented in user-friendly formats to assist patients in procuring information. The Internet has also had a major impact in the delivery of information and education to healthcare professionals (Kiser, 2001). Numerous organizations have Web sites for disseminating new medical information to physicians. Various Web-based physician education services have been established. Some hospitals offer physician portals allowing physicians to access patients' medical records, lab results, and radiological images and reports from their offices, homes, or other remote locations (Cohen, 2005).

The Internet is also redefining communication channels between doctors and patients, as well as between healthcare providers and other healthcare-related agencies. DeShazo, Fessenden, and Schock (2005) suggest the top two emerging trends in healthcare are (1) online patient/physician communication and (2) secure connectivity and messaging among hospitals, labs, pharmacies, and physicians. Advances in home technology coupled with the aging of the baby-boom generation have created the demand for better communications with patients about their on-going care and monitoring. Improving the communication between the patient's at-home technology and the provider's technology is also a growth opportunity. Based on the adequacy of information transmitted to the healthcare provider, the physician saves appointment times and patients are freed from excessive office visits, thereby lowering transaction costs (Flower, 2005).

The Internet and other advances in IT have enabled new models for electronic delivery of a variety of healthcare services. Kalyanpur, Latif, Saini, and Sarnikar (2007) describe the market forces and technological factors that have led to the development of Internet-based radiological services and agree the Internet has provided the platform for cost-effective and flexible radiological services. Wells (2007) agrees the practice of evidence-based medicine requires access to the Internet, mobile devices, and clinical decision-support tools to assist practitioners in improving preventable medical errors.

Favorable External Environment

There is growing support worldwide for the utilization of more IT in healthcare (Caro, 2005). Reports from Australia, Great Britain, India, Italy, and Norway, for example, document local, regional and national healthcare projects and initiatives utilizing IT (Sharma, 2004; Grain, 2005; Marino & Tamburis, 2005; Bergmo & Johannessen, 2006; Fitch & Adams, 2006).

In the U.S., more funds are being made available in the form of grants and demonstration projects by the Federal government to encourage greater adoption of IT in healthcare. In 2004, officials of the U.S. Department of Health and Human Services (HHS) disclosed a ten-year healthcare information infrastructure plan, the "Decade of Health Information Technology," to transform the industry from a paper-based system to an electronic one. More than 100 hospitals, healthcare providers, and communities in 38 states were awarded $96 million over three years to develop and use IT for healthcare. Awards were focused on communities and small and rural hospitals. Five states, Colorado, Indiana, Rhode Island, Tennessee, and Utah, were awarded $25 million over five years to develop secure statewide networks for accessing patient medical information. The National Opinion Research Center at the University of Chicago was awarded $18.5 million to create a National Health Information Technology Resource Center to provide technical assistance, tools, and a best-practices repository as well as provide a focus for collaboration to grantees and other federal partners. In all, HHS awarded nearly $140 million in grants to promote the use of IT, develop state and regional networks, and encourage collaboration in advancing the adoption of electronic health records (Swartz, 2005).

In 2005, the Agency for Healthcare Research and Quality, part of HHS, awarded over $22 million in grants to 16 institutions in 15 states to aid in implementing healthcare IT projects emphasizing patient safety and healthcare quality. The grants were designed to encourage the sharing of information among providers, labs, pharmacies, and patients, with the specific goal of decreasing medication errors and duplicate testing. Eleven of the 16 grants were awarded to small and rural communities (Anonymous, 2005c).

In his 2004 State of the Union address, President Bush called for the transformation of electronic health records within the next ten years in the United States and urged more healthcare organizations to consider implementing such health information technologies as electronic healthcare records (EHR), computerized ordering of prescriptions and medical tests, clinical decision support tools, digital radiology images, and secure exchange of authorized information, emphasizing that all of these technologies have been shown to improve patient care quality and reduce medical errors (Abrahamsen, 2005). In President Bush's 2008 budget proposal, there is funding for a healthcare system and IT is the starting point for the system. Carolyn Clancy, Director of the Agency for Healthcare Research and Quality, agrees the data generated from the healthcare system could answer various medical inquiries and could draw on the data of EHRs of millions of individuals to advance the evidence base for clinical care. She further suggests the data could reveal why costs are increasing and what risks and benefits are associated with particular prescription drugs (Lubell, 2007).

Industry Standards

The development of industry standards for both data communications and data taxonomies may be the most profound of all the opportunities currently facing healthcare. As a crucial first-step in modernizing the U.S. healthcare system, all industry participants—providers, payers, and regulators—are being urged to adopt interoperable systems and common data standards for existing federal, state, and health networks along with standard practices to promote data sharing and protection of patient privacy (Swartz, 2006b). Standard data communications technology and standard data definitions are essential for such health information technologies as electronic health records and e-prescribing (Brailer, 2004).

A recent study of several disability compensation programs within the U.S. found each program uses its own terminology and disability definitions causing non-standard interpretation of terms, misinterpretation of data, and delay in the disability evaluation process. The study suggests defining and adopting a standard for disability evaluation could not only eliminate process inefficiencies in determining disabilities but could also facilitate innovative disability technology practices (Tulu, Hilton, & Horan, 2006).

Some standardization of data has been introduced with the Health Insurance Portability and Accountability Act (HIPPA) legislation, but for the most part, standardization projects are voluntary and lack assurances the standards will be adopted by all parties. The National Quality Forum (NQF) endorsed a voluntary consensus standard for patient safety events and has been adopted by more than 260 healthcare providers, consumer groups, professional associations, federal agencies, and research and quality improvement organizations. The taxonomy is the first standardized, integrative classification system adopted by a group of medical agencies, organizations, providers, and U.S. states. The standard establishes a common taxonomy for healthcare errors and other patient safety problems. It can be used to classify data collected in different reporting systems, allowing the data about patient safety events to be combined and analyzed (Anonymous, 2005d).

Standardization can result in greater levels of system integration, increased sharing of data between healthcare partners, greater information continuity throughout the healthcare industry, and more powerful data mining. Enterprise Resource Planning (ERP) systems can offer more online processing to all users and function, automate routine job processes, and redefine existing work processes (Landry et al., 2005). With integration of systems and standardization of data taxonomies, the healthcare industry will experience changes similar to those in other industries where "enterprise" systems have been adopted. Others agree a seamless support of information flow for healthcare processes that are increasingly distributed requires the ability to integrate heterogeneous IT systems into a comprehensive system (Lenz, Beyer, & Kuhn, 2007).

System standards resulting in a greater level of systems integration is a pressing need. Conn (2007a) reports the compromise reached by two rival standards groups for data communications standards can help to bridge the gap between physicians' offices and hospitals in the electronic health record systems they use. The Continuity of Care Document standard combines the independent works by two standards development organizations on creating electronic summaries of care for discharged patients.

EXTERNAL THREATS

Legal Compliance

The Health Insurance Portability and Accountability Act (HIPAA), enacted by Congress in 1996, is the most significant Federal legislation affecting the U.S. healthcare industry since the Medicare and Medicaid legislation of 1965. Title I of HIPAA legislates improved portability and continuity of health insurance coverage for American workers. Title II addresses "administrative simplification" requiring the development of standards for the electronic exchange of personal health information (PHI). Administrative simplification requires rules to protect the privacy of personal health information, the establishment of security requirements to protect that information, and the development of standard national identifiers for providers, health insurance plans, and employers. Two significant sections of HIPAA are (1) the Privacy Rule and (2) the Security Rule.

The Privacy Rule legislates in detail the collection, use, and disclosure of personal health information. To be in compliance with the Privacy Rule, covered entities must notify individuals of uses of their PHI, keep a record of all disclosures of PHI, and document and disclose their privacy policies and procedures. Covered entities must have designated agents for receiving complaints and they must train all members of their workforce in proper procedures.

The Security Rule complements the Privacy Rule and presents three types of security safeguards designated as administrative, physical, and technical. For each type, the Rule identifies various security standards and names (1) *required implementation specifications* which must be adopted and implemented as specified in the Act and (2) *addressable implementation specifications* which are more flexible and can be implemented by the covered entities as deemed appropriate.

Covered entities face potentially severe penalties for failure to comply with the complex legalities of HIPAA and this has caused much concern throughout the industry. Physicians, medical centers, and other healthcare providers have experienced increased paperwork and cost to incorporate the requirements of this legislation into their current methods of operation. Future adoptions of new information technologies will be subject to its specifications as well (American College of Physicians, 2006).

Loss of Patient Trust

The Institute of Medicine (IOM) of the National Academy of Sciences released a report in 1999 that caused much attention to be focused on the U.S. healthcare industry. The report stated medical errors caused between 44,000 and 98,000 preventable deaths annually, and medication errors alone caused 7,000 preventable deaths (Kohn, Corrigan, & Donaldson, 1999). Within two weeks of the report's release, Congress began hearings and the President ordered a government-wide feasibility study for implementing the report's recommendations for (1) the establishment of a Center for Patient Safety, (2) expanded reporting of adverse events, and (3) development of safety programs in healthcare organizations. According to a study by Healthgrades, a leading healthcare ratings organization, during the period 2000–2002 the estimated number of accidental deaths per year in U.S. hospitals had risen from the 98,000 reported by the IOM in 1999 to 195,000 (Shapiro, 2006).

On July 29, 2005, President Bush signed into law the Patient Safety and Quality Improvement Act, establishing a federal reporting database. This was the first piece of patient safety legislation since the 1999 IOM report. Under this act, hospitals voluntarily report "adverse patient events" to be included in the database, and "patient safety organizations" under contract with the Federal government, analyze the events and recommend improvements. The reports submitted by the hospitals remain confidential and cannot be used in liability cases. The most recent Healthgrades report (April 7, 2007) covering the period 2003–2005, indicates that patient safety incidents have increased over the previous period to 1.16 million among the 40 million hospitalizations covered under the Medicare program.

Healthcare must utilize all available means to maximize patient safety and retain patient trust. Healthcare is an information intensive industry and the delivery of high quality healthcare depends in part on accurate data, available at the point of decision. Handwritten reports, notes and orders, non-standard abbreviations, and poor legibility all contribute to substantial errors and injuries (Kohn et al., 1999). A doctor needs to know a patient's medical history, ancillary providers need to be able to read the doctor's orders and patients need to be able to understand what the doctor expects of them. IT solutions are available that address many of the data accuracy and availability problems in healthcare records (Poston, Reynolds, & Gillenson, 2007); however, the level of adoption for these technologies is not impressive. For example, a 2005 report predicts that by the end of 2007, only 59% of all medical groups will have implemented an Electronic Health Record (EHR) system (Gans, Kralewski, Hammons, & Dowd, 2005). And although Computerized Physician Order Entry (CPOE) systems reduce medication errors by 80% (Bates et al., 1999), a 2004 survey by Leapfrog found only 16% of hospitals, clinics, and medical practices expected to be utilizing CPOE by 2006.

Cost

One of the most immediate barriers to widespread adoption of technology is the high cost of implementation. A report by the Annals of Internal Medicine estimated that a National Health Information Network (NHIN) would cost $156 billion in capital investment over five years and $48 billion in annual operating costs. Approximately two-thirds of the capital costs would be needed to acquire the functionalities and one-third for interoperability. The present level of spending is only about one-fourth of the amount estimated for the model NHIN. While an NHIN would be expensive, $156 billion is equivalent to 2% of annual healthcare spending for 5 years (Kaushal et al., 2005). Industry reports from Datamonitor, Gartner, and Dorenfest & Associates predict increased spending on IT by healthcare providers at an annual rate of between 10% and 15% (Broder, 2004). A study conducted by Partners Healthcare System, Boston, concluded that a national healthcare information system would cost $276 billion, take 10 years to build, and require another $16.5 billion annually to operate. However, the study also concluded that such a system would save U.S. hospitals $77.8 billion annually because of more efficient communication (Anonymous, 2005a).

According to a study by RAND Health (Health Information Technology, 2005), the U.S. healthcare system could save more than $81 billion annually, reduce adverse healthcare events, and improve quality of care if it were to widely adopt health information technology. Patients would benefit from better health and payers would benefit from lower costs; however, some hospitals fear loss of revenue due to reduced patient length of stay. A recent study in Florida suggests that this fear may be unfounded. The results of the study suggest that there is a significant and positive relationship between increased levels of IT use and various measures of financial performance. The results indicated that IT adoption is consistently related to improved financial outcomes both overall and operationally (Menachemi, Burkhardt, Shewchuk, Burke, & Brooks, 2006).

Lopes (2007) agrees that fully integrated electronic medical records systems can replace paper records and allow hospitals and physicians to share medical information electronically to improve response and lower costs from duplication. However, the adoption of such systems is slowed by the high cost of new technology, the complexity of the systems, training, and an unwillingness to adapt work processes to include new information technologies.

DISCUSSION AND CONCLUSIONS

Table 1 summarizes the current SWOT analysis of IT implementation in the healthcare industry in the U.S. The healthcare industry faces multi-faceted challenges to improve patient safety and assure information security while containing costs and increasing productivity. The key area for addressing these concerns is more investment in IT to facilitate the flow of information and offer access to providers and partners along the healthcare supply chain, reduce medical errors, and increase efficiency. Implementation of IT networks to achieve the required level of information and data communications is complicated by the variety of systems already used by provider organizations as well as the lack of system integration within provider organizations.

Various benchmarking studies are helping to educate healthcare providers about IT expenditures and offer comparison reports on expenditures. The availability of systems far exceeds the budget of most organizations to adopt them. However, the improved revenue cycles and cost-benefit offered by IT investments are becoming easier to quantify in faster turnaround and processing of patient-related transactions, shared data, reduced duplication of efforts, and increased provider and customer satisfaction.

Information technology can help take the paper chart out of healthcare, and eliminate error, variance and waste in the care process. IT can help connect the appropriate persons, knowledge, and resources at the appropriate time and location to achieve the optimal health outcome, increase customer service and patient care with industry leading medication fill rates and timely deliveries, cut operation costs through advanced warehouse management, reduce internal labor costs, and improve enterprise efficiencies of healthcare organizations, all through tightly integrated applications.

Table 1. SWOT Analysis

Strengths • Improved Patient Safety • Greater Efficiency of Operation • Current Investment in IT	**Weaknesses** • Lack of System Integration • User Resistance • Slow IT Adoption
Opportunities • The Internet • Favorable External Environment • Industry Standards	**Threats** • Legal Compliance • Loss of Patient Trust • Costs

IT can ultimately transform the healthcare industry. Along with improved safety and greater patient trust, adopting IT in healthcare can only improve current conditions and help the United States improve healthcare in general. Concerns remain about how smaller practices can afford the costs of new systems that require them to move their paper medical records to electronic media. These costs and start-up expenses mean an unequal playing field for small practices versus larger healthcare systems with more money to spend on IT integration.

AREAS FOR FUTURE RESEARCH

IT applications in healthcare are reaching the growth phase of the lifecycle. The strengths, weaknesses, opportunities, and threats at this stage of the life cycle are clear, but few solutions have been proposed. Research is needed to forecast the SWOT issues as IT in healthcare moves from growth to maturity. Case studies in both large and small physician practices as well as in large and small healthcare systems are needed to better understand the IT implementation timeframe and costs. Studies that address ways to overcome human barriers to implementation are also needed.

Using the supply chain model, the healthcare information system needs to be studied as to access and applicability for other providers including pharmacists, dietitians, insurance companies, home health service and equipment providers, and other vendors to healthcare. If the healthcare system it to be truly integrated, these additional players must be included. Protection of patient medical information, access to information, data security, and assurances of privacy should also be studied. International suppliers and other options for outsourcing should be investigated as a cost containment strategy.

With the growing body of patient-related medical and healthcare data, other applications should be studied. Data is available to postulate ways to improve lifelong health and reduce the incidences of various diseases and ailments. Such data mining should be studied by those in the management information systems area to determine cause and effect and recommend changes. As employers seek to contain healthcare costs of their employees, such data can aid in more active involvement in reducing health risks and making lifestyle changes (i.e., smoking cessation programs, dietary counseling, healthy cafeteria food, work-place gyms).

Choosing the best approach to implement IT systems in healthcare settings is also an area for further study. These systems should meet healthcare goals in addition to functionality and integration criteria. Involving physicians and other clinicians in selecting IT systems can increase their support and lessen their resistance to technology. In fact, the more stakeholders are involved in IT selection and implementation planning, the greater their acceptance and rate of adoption will be.

Studies in IT implementation outside the healthcare industry need to be reviewed and analyzed to determine where other industries have had success in implementation or have developed tools that could aid the healthcare arena. Human resource studies of executive ownership and accountability can help the healthcare industry better prepare physicians and other practice managers to overcome the user resistance to IT.

REFERENCES

Abrahamsen, C. (2005). Washington taps into healthcare technology. *Nursing Management*, *36*(3), 52.

Adams, B., Berner, E. & Wyatt, J. (2004). Applying strategies to overcome user resistance in a group of clinical managers to a business software application: A case study. *Journal of Organizational and End User Computing*, *16*(4), 55-65.

Adams, D., Nelson, R. & Todd, P. (1992). Perceived usefulness, ease of use, and usage of information technology: A replication. *MIS Quarterly, 16*(2), 227-247.

American College of Physicians. (2006). Health Insurance Portability and Accountability Act Privacy Rule causes ongoing concerns among clinicians and researchers. *Annals of Internal Medicine*, *145*(4), 313-316.

Anderson, K. (2007). The ASRT Salutes Your Excellence, North Mississippi Medical Center. *ASRT Scanner*, *39*(5), 8-11.

Anonymous. (1997). Just how smart are smart cards? *Financial Executive, 13*(5), 8-10.

Anonymous. (2005a). Adoption of health information technology standards would save billions. (2005). *Healthcare Strategic Management*, *23*(2), 4-6.

Anonymous. (2005b). Healthcare information technology panel established. *Association of Operating Room Nurses, AORN Journa*l, *82*(6), 1018.

Anonymous. (2005c). Millions awarded to healthcare IT projects. *Quality Progress*, *38*(12), 12.

Anonymous. (2005d). Patient safety event taxonomy standard endorsed. *Association of Operating Room Nurses, AORN Journal*, *82*(6), 1040-1041.

Ansoff, H. (1965). *Corporate strategy*. New York: McGraw-Hill.

Balas E., Weingarten S., Garb C., Blumenthal D., Boren S., & Brown G. (2000). *Improving preventive care by prompting physicians. Arch Intern Mcd, 160*, 301-308.

Baldrige Award Recipient Profile: Malcolm Baldrige National Quality Award 2006 Award Recipient, Healthcare: North Mississippi Medical Center. Retrieved from http://www.nist.gov/public_affairs/releases/mississippi.html

Bates, D., Teich, J., Lee, J., Seger, D., Kuperman, G., et al. (1999). The impact of computerized physician order entry on medication error prevention. *Journal of the American Medical Informatics Association*, *6*(4) 313-321.

Bergmo, T. & Johannessen, L. (2006). The long road from potential to realized gains of information technology in healthcare – experiences from Norway. *International Journal of Economic Development*, *8*(3), 682-716.

Bowen, R., & Bassler, P. (2006). Can you hear me?. *Healthcare Informatics*. Retrieved from www.healthcare-informatics.com.

Brailer, D. (2004). Perspective: Translating ideals for health information technology into practice. *Health Affairs*, W318-W320.

Broder, C. (2004). U.S. healthcare providers plan IT spending increase. *iHealthBeat*. Retrieved, from http://www.ihealthbeat.org/index.cfm

Brodkin, J. (2007). A Growing Divide in Health-Care IT. *Network World*, *24*(6), 17.

Burke, D. & Menachemi, N. (2004). Opening the black box: Measuring hospital information technology capability. *Healthcare Management Review*, *29*(3), 207-217.

Caro, D. (2005). Forging e-health partnerships: Strategic perspectives from international executives. *Healthcare Management Review*, *30*(2), 174.

Chin, T. (2005). Are physicians at the infotech tipping point?. *American Medical News, 48*(10), 20-21.

Cohen, S. (2005). Emerging benefits of integrated IT systems. *Healthcare Executive*, *20*(5), 14-18.

Conger, T. & Chiavetta, D. (2006). The pulse of industry. *Industrial Engineer, 38*(1), 30-35.

Conn, J. (2007a). Agreement could put EHRs on fast track. *Modern Healthcare, 37*(9), 22-23.

Conn, J. (2007b). Eyes on the HER. *Modern Healthcare*, *37*(9), 37-38.

Conn, J. (2007c). Still the matter of money. *Modern Healthcare*, *37*(9), 36-37.

Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319-340.

DeShazo, C., Fessenden, R. & Schock, P. (2005). Realizing IT's Potential. *Physician Executive*, *31*(4), 26-29.

Fitch, C. & Adams, C. (2006). Managing mobile provision for community healthcare support: Issues and challenges. *Business Process Management Journal*, *12*(3), 299.

Five Large Companies Working to Provide Lifelong Personal Health Records for Employees. (2007). *Healthcare Financial Management*, *61*(2), 11.

Flower, J. (2005). The 16 challenges. *Hospitals & Health Networks*, *4*(4), 16-18.

Gans, D., Kralewski, J., Hammons, T., & Dowd, B. (2005). Medical groups' adoption of electronic health records and information systems. *Health Affairs (Project Hope)*, *24*(5), 1323-1333.

Gillette, B. (2004). IT investment shows marked increase, but mostly on the clinical side. *Managed Healthcare Executive*, *14*(9), 58.

Glaister, K. & Falshaw, J. (1999). Strategic planning: Still going strong?. *Long Range Planning*, *32*(1), 107-116.

Goldberg, R., Kuhn, G., Louise, A., & Thomas, H. (2002). Coping with medical mistakes and errors in judgment. *Annals of Emergency Medicine*, *40*(2), 287-292.

Grain, H. (2005). Information systems in the new world: An emerging national approach. *Australian Health Review*, *29*(3), 292-297.

Greene, J. (2007). The costs & benefits of IT. *Trustee*, *60*(2), 8-12.

Harris, K., Donaldson, J. & Campbell, J. (2001). Introducing computer-based telemedicine in three rural Missouri counties. *Journal of End User Computing*, *13*(4), 26-35.

Hartzema, A., Winterstein, A., Johns, T., de Leon, J., Bailey, W., McDonald, K., & Pannell, R. (2007). Planning for pharmacy health information technology in critical access hospitals. *American Journal of Health-System Pharmacy*, *64*(3), 315-321.

Health Information Technology: Can HIT lower costs and improve quality?. (2005). *Rand Health*. Retrieved from http://www.rand.org

Health Grades, Inc. (2007). HealthGrades quality study: Fourth annual patient safety in American hospitals study. Retrieved from http://www.healthgrades.com/media/dms/pdf/PatientSafetyInAmericanHospitalsStudy2007.pdf.

Helms, M. (1999). How to be successful in China: A SWOT analysis. *Competitiveness Review*, *9*(2), 1-10.

Helms, M. (2003). Japanese managers: Their candid views of entrepreneurship. *Competitiveness Review*, *13*(1), 24-34.

Hitt, M., Ireland, R., Camp, S., & Sexton, D. (2001). Guest editors' introduction to the special issue: Strategic entrepreneurship: Entrepreneurial strategies for wealth creation. *Strategic Management Journal*, *22*(6-7), 479-491.

Hofer, C. & Schendel, D. (1978). Strategy *formulation: Analytical concepts*. St. Paul, MN: West Publishing Company.

Hough, C., Chen, J. & Lin, B. (2005). Virtual health/electronic medical record: Current status and perspective. *International Journal of Healthcare Technology & Management*, *6*(3), 257.

Janz, B., Pitts, M. & Otondo, R. (2005). Information systems and healthcare II: Back to the future with RFID: Lessons learned – some old, some new. *Communications of the Association for Information Systems*, *15*, 1.

Kalyanpur, A., Latif, F., Saini, S., & Sarnikar, S. (2007). Inter-Organizational e-Commerce in healthcare services: The case of global teleradiology. *Journal of Electronic Commerce in Organizations*, *5*(2), 47-56.

Kaushal, R., Blumenthal, D., Poon, E., Jha, A., Franz, C., Middleton, B., et al. (2005). The costs of a national health information network. *Annals of Internal Medicine*, *143*(3), 165-173.

Kay, D. & Clarke, J. (2005). Raise care quality, lower costs successfully with IT partnership. *Journal of Nursing Administration*, *12*.

Khan, G. & Al-Buarki, E. (1992). Strategic planning in bahrain. *Management Decision, 30*(5), 3-10.

Kiser, K. (2001). Doctor, teach thyself. *Online Learning*, 23-28.

Kohn, L. (2001). The Institute of Medicine report on medical error: Overview and implications for pharmacy. *American Journal of Hospital Pharmacy*, *58*(22), 63-66.

Kohn, L., Corrigan, J., & Donaldson, M. (Eds.). (1999). *To err is human: Building a safer health system*. Institute of Medicine, Washington: National Academy Press.

Kotler, P. (2000). *Marketing management*. Upper Saddle River, NJ: Prentice-Hall.

Kuperman, G., Teich, J., Gandhi, T., & Bates. D. (2001). Patient safety and computerized medication ordering at Brigham and Women's Hospital. *Journal of Quality Improvement*, *27*(10), 509-520.

Landry, B., Mahesh, S. & Hartman, S. (2005). The impact of the pervasive information age on healthcare organizations. *Journal of Health and Human Services Administration*, *27*(3-4), 444-464.

Lenz, R. (2007). IT Support for healthcare processes – Premises, challenges, perspectives. *Data & Knowledge Engineering*, *61*(1), 39-58.

Lenz, R., Beyer, M. & Kuhn, K. (2007). *International Journal of Medical Informatics*, *76*(2-3), 201-207.

Lieber, S.(2007). IT goes global. *Modern Healthcare*, *37*(9), 30.

Lopes, G. (2007). Paper still rules patient records. *Washington Times* (DC), March 5.

Lubell, J. (2007). IT may play Research Role in '08. *Modern Healthcare, 37*(5), 4.

Mahmood, M., Hall, L. & Swanberg, D. (2001). Factors affecting information technology usage: A meta-analysis of the empirical literature. *Journal of Organizational Computing and Electronic Commerce*, *11*(2), 107-130.

Marino, A. & Tamburis, O. (2005). Technological innovation and organisational context information management in the high specialised hospital: The case of South Italy. *International Journal of Healthcare Technology & Management*, *6*(3), 276.

Martin, E., Brown, C., DeHayes, D., Hoffer, J., & Perkins, W. (2005). *Managing Information Technology*, 5th ed. Upper Saddle River, NJ: Prentice-Hall.

McDonald, M. (1999). *Marketing plans*. Oxford, UK: Butterworth-Heinemann Press.

Mekhjian, H., Kumar, R., Kuehn, L., Bentley, T., Teater, P., et al. (2002). Immediate benefits realized following implementation of physician order entry at an academic medical center. *Journal of the American Medical Informatics Association*, *9*(5), 529-539.

Menachemi, N., Burkhardt, J., Shewchuk, R., Burke, D. & Brooks, R. (2006). Hospital information technology and positive financial performance: A different approach to finding an ROI. *Journal of Healthcare Management*, *51*(1), 40-59.

Mintzberg H., Ahlstrand, B. & Lampel, J. (1998). *Strategy safari: A guide through the wilds of strategic management*. New York: Free Press.

Moody, L. (2005). E-Health Web portals: Delivering holistic healthcare and making home the point of care. *Holistic Nursing Practice*, *19*(4), 156.

Natesan, C. (2005). The impact of the Internet on healthcare services. *Journal of American Academy of Business*, *7*(2), 247-252.

Order entry rules: Healthcare enterprise achieves physician acceptance, reduced medication errors and improved patient outcomes through CIS and CPOE technology. (2002). *Health Management Technology*, *23*(7), 34, 38.

Ortiz, E. & Clancy, C. (2003). Use of information technology to improve the quality of healthcare in the United States. *Health Services Research*, *38*(2), 10-13.

Parker, P. (2004a). Move care to a higher level with emergency systems. *Nursing Management*, *35*(9), 82-83.

Parker, P. (2004b). Quantify technology's benefits. *Nursing Management*, *35*(2), 41.

Porter, M. (1991). Towards a dynamic theory of strategy. *Strategic Management Journal*, *12*, 95-117.

Poston, R., Reynolds, R., & Gillenson, M. (2007). Technology solutions for data accuracy and availability problems in healthcare records. *Information Systems Management*, *24*(1), 59-71.

Proctor, R. (1992). Structured and creative approaches to strategy formulation. *Management Research News*, *15*(1), 13-19.

Rhoads, C. (2004). Rx for success. *Computerworld*, *38*(34), 27-28.

Roberts, B. & Stimson, R. (1998). Multi-sectoral qualitative analysis: A tool for assessing the competitiveness of regions and formulating strategies for economic development. *Annals of Regional Science, 32*(4), 469-494.

Scalise, D. (2002). CPOE: Is it worth it?. *Hospitals & Health Networks*, *76*(1).

Schnaars, S. (1998). *Marketing strategy*. New York: Free Press.

Sensmeier, J. (2004). Transform workflow through selective implementation. *Nursing Management*, *35*(12), 46-52.

Sensmeier, J. (2005). Here and now: Healthcare's top 10 IT breakthroughs. *Journal of Nursing Administration*, October, 8.

Shane, R. (2002). Computerized physician order entry: Challenges and opportunities. *American Journal of Health-System Pharmacy*, *59*(3), 286-288.

Shapiro, S. (2006). HealthGrades patient safety study shows increase in U.S. hospital incidents. *Medical News Today*, Retrieved from http://www.medicalnewstoday.com/printerfriendlynews.php?newsid=40954.

Sharma, D. (2004). Technologies for the people: A future in the making. *Futures*, *36*(6-7), 733.

Simpson, R. (2005). In direct proportion: Ratios, IT, and trust. *Nursing Management*, *36*(2), 14.

Smith, J. (1999). The behavior and performance of young micro firms: Evidence from business in Scotland. *Small Business Economics*, *13*(3), 185-200.

Sternberg, D. (2004). Trends worth watching. *Marketing Health Services*, *24*(2), 44.

Sternberg, D. (2005). Strategic innovation. *Marketing Health Services*, *24*(2), 44.

Study: U.S. electronic medical-records market count grow four-fold by 2015. (2006). *Business Journal* (Central New York), *21*(7), 8.

Swartz, N. (2005). U.S. hands out technology grants for healthcare. *Information Management Journal*, *39*(1), 14.

Swartz, N. (2006a). E-records may end fraud. *Information Management Journal*, *40*(1), 16.

Swartz, N. (2006b). Tech leaders release e-health roadmap. *Information Management Journal, 40*(1), 18.

Taylor, S. & Todd, P. (1995). Understanding information technology usage: A test of competing models. *Information Systems Research*, *6*(2), 144-176.

Thompson, A. & Strickland, A. (1998). *Crafting and Implementing Strategy,* 10th ed. New York: McGraw Hill.

Tulu, B., Hilton, B. & Horan. T. (2006. Improving disability evaluation productivity: Linking innovative business models with information technology. *International Journal of Healthcare Technology & Management*, *7*(1-2), 168.

Valentin, E. (2001). SWOT analysis from a resource-based view. *Journal of Marketing Theory and Practice, 9*(2), 54-69.

Venkatesh, V. & Davis, F. (1996). A model of the antecedents of perceived ease of use: Development and test. *Decision Sciences*, *27*(3), 451-481.

VHA Inc. introduces new version of its comparative clinical measurement tool. (2007). *Healthcare Purchasing News*, *31*(3), 8-14.

Wells, L. (2007). Role of information technology in evidence-based medicine: Advantages and limitations. *Internet Journal of Healthcare Administration*, *4*(2), 5-10.

Zhang, J. & Kelvin, A. (1999). Strategic entry for products and technology into China. *Journal of International Marketing and Marketing Research*, *24*(2), 14-29, 85-98.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 3, Issue 1, edited by J. Tan, pp. 75-92, copyright 2008 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 1.13
The Core Governmental Perspectives of E-Health

Antti Syväjärvi
University of Lapland, Finland

Jari Stenvall
University of Lapland, Finland

INTRODUCTION

Public healthcare is facing huge future challenges in order to deal with rising costs, growing demands of customers, information flow, demographic changes, and aging population. The healthcare service sector can be seen as an information intensive area during an era of innovation and information technology (cf. Bellamy & Taylor, 1998). According to McLaughlin, Rosen, Skinner, and Webster (1999), it is common to assume that technological interventions are almost inevitable and it is humans' duty, at least to some extent, to follow the suggested development. In the organizational level of public healthcare, high expectations about the technology and its new possibilities are introduced. Additionally, the customers can seek support and advice for their healthcare needs from thousands online connections at any time of a day (e.g., Silber, 2003). The European Commission (2004) states how "eHealth offers European citizens important opportunities for improved access to better health systems" (p. 22). This trend has implications to human beings and governments.

The electronic health services produced by the information and communication technology (ICT) belong to the era of e-government. The e-government can be seen as an electronic exchange of information and services between different actors (cf. Mälkiä, Anttiroiko, & Savolainen, 2004; Oliver & Sanders, 2004). The development of information society throughout the last decades has brought up possibilities to adapt, modify, and reorganize healthcare practices and services (e.g., Gallivan, 2001; Turner, Fraser, Muir Grau, & Toth, 2002). The ICT has been used as a tool to reorganize best organizational practices, information management, and government. The ICT has also given a possibility to produce tailored healthcare services and to gain improvements in cost-effectiveness, access, safety, and quality of public healthcare services (Bates et al., 2001; Whitten et al., 2002).

In the future, the healthcare organizations in public sectors will confront many challenges by means of the ICT implementation. This situation is considered here both as the function of healthcare organizations and as the supply of knowledge intensive public health services. The current viewpoint presumes a paradigm that is structured on the basis of specific conceptualization. The purpose of this article is to conceptualize the complex topic of e-health from the governmental viewpoint and to clarify the best organizational practices. Special notation is also given for human resources, information management and the ICT implementation. Finally, some future trends are shortly discussed.

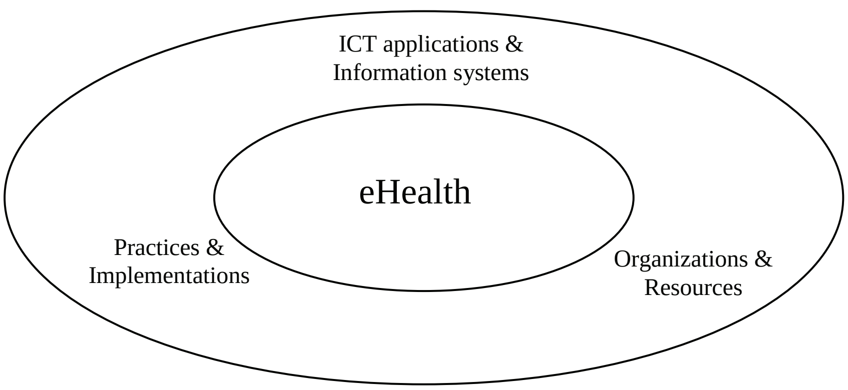
BACKGROUND

Experts define e-health differently and the term has some overlapping conceptual views. The term "e-health" is derived from the term "electronic commerce" (i.e., e-commerce), which was introduced in the mid-1990s to reflect the growing commercial use of the Internet. The e-health (cf. Eysenbach, 2001; European Commission, 2004; Silber, 2003), when recognized as a governmental issue, refers to the use of ICT applications or information systems to improve or enable health and healthcare services. Now it also refers to the main organizational and resource based factors before, during, and after an introduction of ICT. Finally, the term refers to best practices and implementations produced or needed to confront and manage with the ICT-based healthcare. The e-health concerns customers, patients, professionals, but additionally the whole primary healthcare, home care, and organizations like hospitals. It engages the terms like telemedicine, telematics, telehealth, medical and health informatics, interactive health communication, and so forth. To conclude, the e-health as a governmental concept give good reason for to describe the abovementioned combination by the means of three conceptual dimensions.

CONCEPTUAL DIMENSIONS OF THE E-HEALTH PARADIGM: THE GOVERNMENTAL PERSPECTIVE

In the first dimension, the e-health connection can be found by referring to ICT applications and information systems in healthcare and in supporting functions (cf. McGinnis, 1997; Turban, McLean, & Wetherbe, 1999). In many cases, this has included telemedicine and health informatics as patients have been informed, examined and treated over distance by using appropriate applications (e.g.,

Figure 1. E-health concept according to the core governmental dimensions



Hailey, Roine, & Ohinmaa, 2002). The e-health is seen now through issues like information system, standardization, and system quality (e.g., digital referrals, data security, access, and privacy, etc.). Also, the usability or user-friendly orientation should be placed under a close scrutiny. Hence, the applications and systems require understanding about technology and its effect on organizational structures, processes, and actors.

Secondly, it is considered that by e-health it is possible to renew and to produce more effective healthcare practices and implementations. Now the e-health is studied typically in conjunction with key processes like the management of ICT applications and information (Plsek & Wilson, 2001; Tachakra, El Habashy, & Dawood, 2001; Walker & Whetton, 2002) and by the means of best practices or ICT implementations in healthcare (e.g., Paré & Elam, 1999). The introduction of technology may support or change operational practices in organizations. The operational aspect engages also electronic healthcare services as the e-health services are multileveled. Services represent, for example, the possibility of consumers to interact online, the possibilities for institution-to-institution data transmissions, and the possibilities for peer-to-peer communication of consumers. However, the overall question is about the best practices and implementations in order to successfully complete the e-health in service sector.

As third dimension, the organization and its various resources (e.g., economical and humanistic) have been under intensive study. For example, Parente and Dunbar (2001) found that hospitals with integrated information systems have higher total and operating financial margins than those without these systems. Whitten et al. (2002) did a systematic review and found no good evidence that telemedicine is cost-effective. Additionally, the e-health can be seen in terms of human resources. This is the case, for example, with the acceptance of ICT-based organizational solutions (Mathieson, Peacock, & Chin, 2001) and with the professional human resources (e.g., Syväjärvi, Stenvall, Harisalo, & Jurvansuu, 2005). The interest has also been in health services and in policy to provide a diverse range of services (e.g., Silber, 2003; Turner et al., 2002). Both customers and professionals can be seen as key actors, because they use applications and participate in electronic service interaction (e.g., Hailey et al., 2002).

Hence, the e-health from governmental perspective can be defined as health service and information delivered through different technologies. Applications of the ICT and information systems support health, healthcare, and health services. In case of current governmental perspective, the dimension of ICT applications and information systems concentrates mainly on technological system, standardization, and usability issues. The ICT applications can be seen as architectures to understand and plan information system components in the form of an organizational infrastructure. The most crucial questions are the standardization of information systems and their usability in organizational environment.

It seems worldwide that open technical standards to provide e-health are lacking. McGinnis (1997) studied health informatics and stated the critical importance of data exchange standards for communication between healthcare providers. The usability instead is traditionally defined (cf. Isomäki, 2002) by attributes like learnability, efficiency, memorability, errors and satisfaction. However, without neglecting the human factors like learning, sensation, and perception, memory, problem solving, and so forth. An optimal system should be easy to learn, it should be efficient to use, the system should be easy to remember, the error rate with the system should be low, and the system should be pleasant to use. Corresponding arguments of usability and human-centered design are also offered by ISO quality standards.

Second governmental dimension of e-health was about practices and implementations. E-health as practices and implementations emphasizes the need to combine knowledge and the use of ICT in various ways. In this article, the management as an organizational process has the viewpoint of ICT-based e-health. McLaughlin et al. (1999) claimed that in hospitals it is needed a wider perspective than social, since managers and users need to have a commitment to longtime innovative process. Clark (1995) suggested that leaders have the main responsibility for development work, because the problem with technological changes was frequently that people and technology didn't meet or people didn't participate.

The management has also direct link to best practices in organization and its health service production. Martin (2000) claimed that the quality of public services and cost-effectiveness depend on management. Smith (2004) studied how to inspire and motive professional service teams and indicated the importance of discussion and collective responsibility. The e-health can be seen as a part of knowledge intensive public services in a knowledge intensive society. The previous conceptualization invoked to describe the current e-health practices and implementations more generally as seminal goals for public healthcare organizations. However, it should be reminded that information technology does not necessarily lead so radical organizational changes as expected (Bellamy & Taylor, 1998; Nicholson, 2003).

The third governmental dimension of e-health involved organizations and resources. The governments try to manage with the increasing costs of public administration. Simultaneously the need to improve the cost-effectiveness of public healthcare has become apparent. One way to resolve the dilemma has been the use of electronic health services. However, the results have been controversial. Whitten et al. (2002) made a massive evaluation and they included 24 of 612 identified articles presenting cost-benefit information. About 30% of those 24 tried to solve the economic function of telemedicine without a clear outcome.

However, in public organizations the ICT has had often a transformative role. The e-health is understood and projected as being able to reshape the way healthcare services are delivered. The demands set by stakeholders are obvious. Governments want accountability to ensure that the funds are used effectively. Healthcare providers want a better quality of work life through the ICT tools. Patients and customers want access to the high quality healthcare service. One response has been the virtual life of organizations (cf. Nicholson, 2003). The virtual life in current context can be defined as a group of people or organizations that interact through interdependent task guided by common good purpose. In addition, the actors have learned to interact across space, time and inter-professional or organizational boundaries by using the ICT-based e-health solutions (Syväjärvi et al., 2005).

ORGANIZATIONS AND E-HEALTH: IMPLEMENTATION AND HUMAN RESOURCES AS CRUCIAL FACTORS

Numerous organizational e-health projects with various goals are ongoing worldwide under public administration. For example, the European Commission launched an Action Plan for a European eHealth area in 2004. The aim is to support EU member states to materialize the benefits available from e-health, and to respond to the growing need for coordinated activities of e-health (European Commission, 2004). These kinds of activities denote a coherent agenda for the implementation of e-health. Hence, the e-health implementation may cause changes in organizational design and such changes typically require new organizational structures, processes, and personnel arrangements. From governmental perspective, the e-health either allows or forces an organization to rearrange especially its human resources. The focus is thus to analyze the relationship between e-health implementation and HR-centered organizational practice. Human resources are considered as a key area of e-health implementation.

Human resources have indeed an essential function both in knowledge intensive healthcare practices and in knowledge innovative services. The implementation of e-health has indeed organizational effects as it should concurrently support the patient mobility, facilitate the citizen-centered healthcare, but particularly to provide healthcare professionals new tools to sustain or improve the quality of public healthcare service. The implementation should allow greater flexibility, for example in case of inter-professional and organizational practices, to enhance information and knowledge exchange between humans. It is given here a synopsis of e-health implementation that contains previous governmental dimensions.

Many nations worldwide are known as countries with high technology. The extensive use of ICT and the adoption of various technological solutions represent everyday practice. In these knowledge intensive nations, the public policy has been largely based on the assumption that society will develop and succeed by focusing on ICT. For example in many European countries the information technology has been harnessed to promote social welfare. The ICT has been used for strengthening and developing the welfare society. The public service organizations and citizens thus have confronted the wave of e-government (e.g., Mälkiä et al., 2004; Oliver & Sanders, 2004). At the same time research has concentrated on the possibilities, quality, and costs of technology. However, what the implemented information technology means to public health service sector, public organizations, and especially to organizational human resources has usually received less attention. Under the issues of e-government and public services, there are several approaches to the human-centered e-health implementation.

Firstly, from the point of how the employees' attitudes and approval affect the success of implementation (Mathieson et al., 2001). This kind of approach has produced theoretical models to explain how implementation depends on reasoned action of humans and the technology acceptance. Considering a more professional view, for example, the doctors who have received training in technology and understand technological concepts are more likely to grasp the positive impact of technology in healthcare (Paré & Elam, 1999). Burge, Creps, and Wright (2001) assumed that doctors are key persons to adopt technology in healthcare organizations, however, the doctors are not enthusiastically accepting the technology if it does not help everyday work. Tachakra et al. (2001) suggest that telemedicine includes the ability to rearrange the distribution of work; especially concerning the tasks of doctors and nurses. The ICT-based e-health seems not to be a question of single profession as it reflects over entire professional work society.

Secondly, the ICT has been studied by the innovative theories. Again models have been presented how a technological innovation is adopted in an organization through various phases. Gallivan (2001) developed a frame of references how to merge a technical innovation into the organizational operation. Thus, there are themes like management support, high resources, organization culture, own responsibility of learning, broad work roles, work security, and individual characteristics, which either increase or decrease the innovative implementation. These studies argue that the success of information technology in organizations more generally reflects the systems of innovations' advance in organizations. Finally, as technology is implemented to organization more clear and exact strategies accompanied with leading spokesmen are required (Walker & Whetton, 2002).

Thirdly, the setup "e-health vs. implementation" stimulates the leadership and management dimensions. Clark (1995) suggests that the use of ICT is an organizational issue that needs efforts in the direction of change. The leaders have thus a responsibility to create and sustain participation, but also trust (cf. Harisalo & Stenvall, 2004). In good accordance, Plsek and Wilson (2001) found complexity in leadership and management of healthcare organizations. They assumed that leaders should found ways to make the change more attractive and meaningful for workers. This is done by securing the open information flow of good practices. The earlier observation about the complexity of knowledge intensive health services denotes the importance of information management. The information management refers to interdisciplinary procedures designed to provide and improve the appropriate information systems and resources by planning, organizing, executing, and evaluating necessary management operations. Thus, the human resource perspective both involves multi-professional diversity practices and presupposes the proper activities of human resource management (Syväjärvi et al., 2005).

Fourthly, the human resource centered ICT implementation considers the development of information systems according to usability factors, organizational practice, and the humanization of computerized information systems (Cummings & Worley, 2001; Isomäki, 2002). The usability of information systems is included in this framework. The information system development (ISD) can be introduced by different phases. It is stressed that the ISD tasks must be completed with knowledge concerning users and contextual analyses. Isomäki (2002) found that only a few information system designers have potential to contribute to the humanization of information systems. Cummings and Worley (2001) considered organizational change as a function of implementation process. They indicated how implementation causes organizational change and this can be controlled throughout singular phases like by information or system contracting and gathering, familiarizing, planning and accomplishing, evaluating and stabilizing. The humanization and e-health-based information system development can be seen as a part of organizational development (OD). Organization development in public health service sector applies this kind of knowledge and practice in order to help organizations to achieve effectiveness and social quality of work life.

The fifth and final aspect to comprise human resource and e-health implementation consists of the cognitive human capacity and service administration. The human capacity indicates the multilevel competencies of professionals. The service administration indicates both the service knowledge and management, which are approached in a customer-centered way. The e-health services are a challenge for both individuals and collective work society. Hence, individual, collective, and organizational competencies (e.g., Hamel & Prahalad, 1994; McLagan, 1997) are clearly required in different phases of the implementation. These competencies represent the foundation of service knowledge, which denotes the ability to arrange and manage both professionally and organizationally with the electronic public services. In present context this entirety (i.e., professional and organizational practices), the e-health implementation and customer relationships can be supported by appropriate and flexible service management. The public service management can comprise, for example, the dimensions of project management, human resource management, network management and customer relationship management. The concept of service management can be also seen as one value creator for service innovations or improvements, and with modifications, as a suitable management approach for organizations from various sectors (cf. Boyne, Martin, & Walker, 2004; Fitzsimmons & Fitzsimmons, 2004; Martin, 2000).

FUTURE TRENDS

In the near future organizations will once again face a dynamic and turbulent environment. That requires flexible and fast response ability from the healthcare organizations and public services. The ongoing era of e-government has enabled organizations to acquire such structures and processes that support new or additional ways of interaction between electronic service providers and customers. Hence, the future trends of e-health indicate implications for health service users, service suppliers and entire society. The purpose of this section is to discuss shortly some essential future trends of e-health. The future possibilities and needs of e-health are introduced according to present literature and reflected alongside the three main dimensions described in current conceptualization. Thus, the final section is based on future prospects and e-health's relationship to e-government.

The present view indicated that the public healt care services will be going through major electronic turnarounds. The e-health revolution accompanied with the future possibilities are far away from straightforward. Many different issues have to be considered in order to make successful ICT-based transitions in public services. In case of organizations and resources there is going to be a further increase in information intensity of the e-health. Public institutions and human actors are stressed and challenged over again with more sophisticated, and hopefully with more user-friendly and appropriate information technology. There will be a need for competent health professional as continuous cognitive pressure is evident. A clear e-health challenge for public healthcare organizations seems to be how to manage with differences in processes, with strong professional culture of healthcare personnel, with cognitive pressure set by ICT, and with hierarchical or traditional structures.

These prospects as seen as future trends of e-health indicate cooperation, but also knowledge and information integration both in organizational and professional level. For example, healthcare professional will probably have fast and unhindered access to the latest knowledge and to the other relevant information needed in clinical practice. It is also assumed that the number of virtual organizations and new innovative services will be increased. The organizational and resource-based future of e-health has to be affected more by customers and patients. The e-health services are targeted as support for the customers (i.e., for those who daily cope with individual health). Thus, as the impact of ICT on healthcare professionals is extensive, similarly the e-health should be harmonized to the specific needs of the users. For example, citizens will increasingly have an opportunity to access their own health data and to contact directly to health professional (e.g., online booking, e-prescribing, home-care monitoring, health information bank, etc.). It is expected that the customer-centered approach will be even more vital as a resource for the future development of e-health. In knowledge intensive society, more demanding customers can be recognized.

In the future, the applications appear to be even more plentiful for consultations with other organizations, professionals, and patients. Portable and mobile tools will be increasingly adopted in healthcare settings together with many web-based gateways or portals. For example, the question can be about the new innovative e-health systems for retrieval, feedback, participation, clinical data management, and in general about the electronic health recording systems (EHR). There will be quite many views about the needs of e-health-based information systems and probably only a few will get it right. The challenge will be the standardization of ICT applications and information systems. Also the users themselves should be equal partners in development teams. As indicated before, people will probably found themselves in the era of virtual life. This indicates the increasing use of collaboration-support information systems.

Finally, the e-health future has implication in the dimension of practices and implementations. The ICT can be seen as a powerful change agent for the whole healthcare sector and it is important to recognize the importance of non-ICT factors in e-health. In this sense, alongside the system implementation it seems to be significant to concentrate on workplace practices and implementations. The best practices and implementations will be valued even more in the complex future of e-health. The e-health interventions based on ICT re-engineering, human-centered information systems, human resource and change management, and practical work innovations will play crucial roles for wellness in workplaces.

It seems that practical work tools or models that support decision making, management, preparation and processes will have an enormous importance. Public healthcare organizations will continue to invest huge amounts of resources into the latest ICT. It seems that new investments on organizational practices and implementations should support both professionals and customers. For example, one challenge is to combine multi-professional practices and service security for the customers benefit. During the continuing era of e-government, the e-health concept will be even more inevitable and natural part of the public healthcare service.

CONCLUSION

The current governmental perspective of e-health has focused into three core dimensions. Current conceptualization has indeed some overlapping areas, but it is important that the approach to e-health is not done only by means of ICT factors or some single science. Hence, the e-health should follow the multi-scientific paradigm as the concept itself has been adopted by the experts of medicine, health science, applied information technology, social science, information and communication science, and by governmental and management science.

It was found how rich the topic of e-health is. The e-health was studied through the ICT implementation and human resources in the knowledge intensive healthcare service organizations. Hence, the implementation of e-health was recognized as a function of various practices and human resources together with future organizational possibilities in public service sector. Alongside the implementation procedures, the human resource was found to be fundamental. Thus, professional, inter-professional and customer perspectives were critical factors. Current perspectives go under the e-government. It combines the perspectives of governmental and management science, social science, health science, and applied information technology. The future trends of e-health thus involve various combined implications to customers and professionals, to service and system suppliers, and to organizations.

REFERENCES

Bates, D. W., Cohen, M., Leape, L. L., Overhage, J. M., Shabot, M. M., & Sheridan, T. (2001). Reducing the frequency of errors in medicine using information technology. *Journal of the American Medical Informatics Association, 8*(4), 299-308.

Bellamy, C., & Taylor, J. A. (1998). *Governing in the information age*. Buckingham, UK: Open University Press.

Boyne, G., Martin, S., & Walker, R. (2004). Explicit reforms, implicit theories, and public service improvement. *Public Management Review, 6*(2), 189-210.

Burge, L., Creps, L., & Wright, B. (2001). If you build it, will they come? *Health Management Technology, 22*(9), 14-20.

Clark, J. (1995). *Managing innovation and change. People, technology, and strategy*. London. Thousand Oaks, CA; New Delhi: SAGE Publications.

Cummings, T. S., & Worley, C. G. (2001). *Organization development and change* (7th ed.). Cincinnati: South-Western College Publishing.

European Commission. (2004). *eHealth—Making healthcare better for European citizens: An action plan for European eHealth area*. COM (2004) 356 final, Brussels.

Eysenbach, G. (2001). What is eHealth? *Journal of Medical Internet Research, 3*(2), e20. Retrieved from http://www.jmir.org/2001/2/e20

Fitzsimmons, J. A., & Fitzsimmons, M. J. (2004). *Service management: Operations, strategy, and information technology* (4th ed.). New York: McGraw-Hill.

Gallivan, M. (2001). Organizational adoption and assimilation of complex technological innovations: Development and application of new framework. *Database of Advances in Information systems, 3,* 51-85.

Hailey, D., Roine, R., & Ohinmaa, A. (2002). Systematic review of evidence for the benefits of telemedicine. *Journal of Telemedicine & Telecare, 8*(1), 1-30.

Hamel, G., & Prahalad, C. K. (1994). *Competing for the future*. Boston: Harvard Business School Press.

Harisalo, R., & Stenvall, J. (2004). *Trust as capital: The foundation of management*. In M. L. Huotari & M. Iivonen (Eds.) Trust as knowledge management and systems in organizations. Hershey, PA: Idea Group Publishing.

Isomäki, H. (2002). *The prevailing conceptions of the human being in information systems development: Systems designers' reflections*. Doctoral dissertation, University of Tampere, Finland.

Martin, S. J. (2000). Implementing best value: Local public services in transition. *Public Administration, 78*(1), 209-227.

Mathieson, K., Peacock, E., & Chin, W. W. (2001). Extending the technology acceptance model: The influence of perceived user resources. *Database for Advances in Information Systems, 32*(3), 86-112.

McGinnis, P. J. (1997). The scope and direction of health informatics. *Aviation, Space, and Environmental Medicine, 73*(5), 503-507.

McLagan, P. (1997). Competencies: The next generation. *Training & Development, 51,* 40-47.

McLaughlin, J., Rosen, P., Skinner, D., & Webster, A. (1999). *Valuing technology: Organisations, culture, and change*. London: Routledge.

Mälkiä, M., Anttiroiko, A. V., & Savolainen, R. (2004). *eTransformation in governance: New directions in government and politics*. Hershey, PA: Idea Group Publishing.

Nicholson, N. (2003). *What stays the same?* In S. Chowdhury (Ed.), *Organization 21C—Someday all organizations will lead this way* (pp. 241-255). Upper Saddle River, NJ: Prentice Hall.

Oliver, E. L., & Sanders, L. (2004). *E-government reconsidered: Renewal of governance for the knowledge age*. Canada: Canadian Plains Research Center, University of Regina.

Paré, G., & Elam, J. J. (1999). Physicians' acceptance of clinical information systems: An empirical look at attitudes, expectations, and skills. *International Journal of Healthcare Technology and Management, 1,* 46-61.

Parente, S. T., & Dunbar, J. L. (2001). Is the health information technology investment related to the financial performance of US hospitals? An exploratory analysis. *International Journal of Healthcare Technology and Management, 3*(1), 48-58.

Plsek, P. E., & Wilson, T. (2001). Complexity, leadership, and management in healthcare organisations. *British Medical Journal, 323*(7215), 746-750.

Silber, D. (2003). *The case for eHealth*. The Netherlands: The European Commission. European Institute of Public Administration.

Smith, G. (2004). *Leading the professionals. How to inspire & motivate professional service teams.* London and Sterling, VA: Kogan Page Limited.

Syväjärvi, A., Stenvall, J., Harisalo, R., & Jurvansuu, H. (2005). The impact of information technology on human capacity, interprofessional practice, and management. *Problems and Perspectives in Management, 1*(4), 82-95.

Tachakra, S., El Habashy, A., & Dawood, M. (2001). Changes in the workplace with telemedicine. *Journal of Telemedicine and Telecare, 7,* 277-280.

Turban, E., McLean, E., & Wetherbe, J. (1999). *Information technology for management.* New York: John Wiley & Sons, Inc.

Turner, A., Fraser, V., Muir Grau, J. A., & Toth, B. (2002). A first class knowledge service: Developing the national electronic library for health. *Health Information and Libraries Journal, 19*(3), 133-145.

Walker, J., & Whetton, S. (2002). The diffusion of innovation: Factors influencing the uptake of telehealth. *Journal of Telemedicine and Telecare, 8*(3), 73-75.

Whitten, P. S., Mair, F. S., Haycox, A., May, C. R., Williams, T. L., & Hellmich, S. (2002). Systematic review of cost effectiveness studies telemedicine interventions. *British Medical Journal, 14,* 1434-1437.

KEY TERMS

E-Government: An electronic exchange of information and services between different actors. Usually refers to local or state government activities that take place by digital processes and various networks.

E-Health: Health services and information delivered through different technologies in order to improve or enable health, healthcare and services with variable contents and transactions (cf. Eysenbach, 2001).

E-Health Usability: The design and technical features that allow electronic healthcare applications and information systems to be suitable for intentional use in a user friendly way.

Health Informatics: Knowledge, skills, and tools to support the delivery of healthcare information and services (cf. Silber, 2003).

Health Information System: An ICT-based routine constructed for the purposes of delivering health and medical care. Information system can be seen as physical process that supports an organization by providing information to achieve goals.

Information Management: Interdisciplinary procedures designed to provide and improve appropriate information systems and resources. Refer to planning, organizing, executing, and evaluating the necessary management operations.

Standardization of E-Health: The method to reduce or eliminate inconsistency, potential added costs and quality problems in applications and information systems of the electronic healthcare.

Chapter 1.14
Differences in Computer Usage for U.S. Group Medical Practices:
1994 vs. 2003

Marion Sobol
Southern Methodist University, USA

Edmund Prater
University of Texas at Arlington, USA

ABSTRACT

Research on the use of information technology in healthcare has focused on hospitals and Health Management Organizations (HMOs). However, little has been done to study the use of IT in group medical practices. In 1994, we conducted a pilot study of group medical practices and then repeated this pilot study in 2003 to obtain a longitudinal picture of the IT services used by these private practices. Researchers can use this to form ideas of the important issues and changes involved in IT usage in group medical practices over the past decade thus providing a needed benchmark to fill a gap in the existing literature and that can be used to compare domestic as well as international practices. For example, an expanded form of this study will be conducted in Taiwan this summer.

INTRODUCTION

Studies of the introduction of computer technology in medical settings have focused on hospitals (Griffith & Sobol, 2000; Sobol, Humphrey, & Jones, 1992; Sobol & Smith, 2001) and more recently on HMOs. In these studies, such issues as barriers to the introduction of technology in hospitals, returns to adoption of technology, and the market status of the adoption of different technologies have been studied. It was found that there are many barriers to the adoption in hospitals. The longitudinal issues of what the changes have been in the last decade have been studied with the results that certainly there has been an increase in adoptions over the past decade. These increases have occurred in both transactional, informational, and strategic uses of technology (Sobol & Woods, 2000). This trend is expected to increase. A survey in 2002 by Sheldon I. Dorenfest & Associates of Chicago indicated that IT spending on healthcare in 2002 would be $21.6 billion (Dorenfest, 2002).

While the focus has been hospitals and HMOs, very little has been done to study the use of IT in group medical practices both small and large. This is the case even though researchers have for years trumpeted the impact of IT on physician's practice (Rodger, Pendharkar, & Paper, 1996; Shine, 1996). In 1994, we conducted an initial study of group medical practices of three or more doctors; we completed a later study in 2003 to obtain a longitudinal picture of the IT services used by these private practices. While this is not yet the definitive study of IT in group medical practices, it can be used to form ideas of the important issues and changes involved in IT usage in the smaller group medical practices over the past decade thus providing a needed benchmark to fill a gap in the existing literature and to start an intensive study of changes in IT usage.

In this article, we look at the differences in computer usage, computer facilities, sources of computer information, and the satisfaction with computer usage in group medical practices from 1994 to 2003. We compare these characteristics and the amount of time spent on business issues by size of practice and years in practice for group medical practices studied in 1994 and 2003.

BACKGROUND

There has been a great deal of research on IT as well as healthcare. Unfortunately, much of this has been of limited use to practicing physicians. From the computer science side of research, most work has been done on theoretical computing structures. This includes work such as neural net applications of drug/plasma levels (Tolle, Chen, & Chow, 2000) or parsing methods for biomedical texts (Leroy, Chen, & Martinez, 2003). When trying to overlap IT and near term healthcare concerns, the research has tended to focus on public policy (Magruder, Burke, Hann, & Ludovic, 2005) or on hospitals. This includes work done on hospitals and adoption of computer-based IT (Sobol et al., 1992; Sobol & Woods, 2000), as well as the impact of IT use on hospital staffing and payroll (Sobol & Smith, 2001). Other work has focused on the barriers to IT adoption within healthcare (*Economist*, 2005; Sobol, Alverson, & Lei, 1999).

On the other end of the spectrum, some research has been conducted on issues surrounding IT in private practices. This research has tended to be very specific in nature, however. This includes whether or not medical practices should hire an IT person or outsource (Lowes, 2005) or the use of electronic billing systems by private practices (Burt, 2005). Other research tends to focus on a hot technology that is currently being embraced such as electronic medical records (Miller & Sim, 2004; Palattao, 2004). What is lacking is an overall benchmark or "snapshot" of overall IT use by private practice physicians. That is the goal of this research.

METHODOLOGY

In the summer of 1994, a mail survey was sent to a sample of 270 multiple physician groups within Maricopa County, Arizona, who were chosen from lists of a value-added reseller. These practices were medical groups containing three or more physicians. A total of 65, or a response rate of 24%, of usable replies were received. This is a good response for a mail survey. In the summer of 2003, 54 physicians were surveyed in group practices of three or more in the Arlington/Mansfield area of Tarrant County, Texas. The surveys were given to the business managers' offices and were returned by mail. Thus, in essence both were mail surveys utilizing the same questionnaire. Both counties (Maricopa and Tarrant) included large metropolitan areas (Phoenix and Fort Worth), were 71 to 77% white and had median per capita income of approximately $22,250 to $22,500. In both counties, 65 to 69% of the people were in the labor force. These statistics show both interview sites to be similar Southwestern areas. The group practices in Tarrant County were randomly chosen from a list of 525 physicians in group practices and provided a sampling rate of 10%. The size of the practices and the years in practice in both surveys were well-distributed over a wide spectrum of sizes and age of practice. In the following paragraphs, we will look at the use of computers in the group medical practice with the primary perspective of assessing how this usage has changed over the last decade.

First, we will use rank order correlation to determine if the orders of importance for various technologies (hardware and applications) have remained the same over the decade. Spearmans' rho was chosen for this because it makes no assumptions about the shape of the relationship between variables. Secondly, we will employ tests of the differences in proportions for software and hardware adoptions, important types of technology and applications, time spent on business applications, and sources of computer information, comparing 1994 and 2003. These tests will be developed for whole groups and will also focus on differences by the size of the practice (number of doctors) and the number of years in practice.

SURVEY RESULTS

Characteristics of the Samples

The practices in the 1994 sample tended to be larger in terms of the number of physicians than the 2003 sample. As we can see from Table 1A, about one quarter of the group practices in 1994 were less than five while 60.4% of those in 2003 were less than five. If we look at practices of less than 10, 50% of the 1994 sample can be compared to 75% of the 2003 group. Why this disparity? There are many different factors. For example, the Arlington area is a fast-growing area and new start-up practices are forming. However, these are

Table 1. Characteristics of samples, 1994 vs. 2003

A. Number of Physicians In Practice	1994	2003
Less than 5	24.1%	60.4%
5 but < 10	25.9	15.7
10 but < 25	31.6	15.7
25 but < 50	18.4	8.2
	100%	100%
B. Years in Practice		
0-5	23.1%	17.3%
6-10	16.9	11.5
11-15	18.5	17.3
16-20	12.3	17.3
21+	27.7	36.6
No answer	1.5	0.0
	100%	100%

Table 2. Current computer systems, 1994 vs. 2003

System	1994	2003
Individual PCs (not connected)	15.4%	17.5%
Network connected PCs	35.4	61.4*
Midrange or Mini Computer	16.9	8.8
Mainframe Computer	13.8	3.5*
Network PCs & Mainframe	3.1**	--
Network PCs & Mini	4.6**	--
Service Bureau	--	8.8**
Don't Know	10.8	--
	100%	100%

** Indicates that the differences are significant at the 0.05 level using a paired t-test, two tail.*
*** Reflects differences in available services.*

not just new physicians but include older practices that are moving to the faster-growing areas. In addition, during the past 10 years, many practices have become part of hospital-based groups in order to save on costs. This leveraging of resources among a large hospital network has meant that there is less of a need to have larger individual practices in order to distribute costs. There may be other factors, but these are hard to delineate. Because of this, for the studies in this paper, we will look at the variables by practice size so we can study small vs. large practices.

In terms of years in practice, the samples are similar. In Table 1B, we see that in 1994, about 23.1% of the practices had been in business five years or less, while 17.3% of the 2003 practices had been in practice 5 years or less. If we look at 6 to 10 years in practice, 16.9% of the 1994 group as opposed to 11.5% had been in practice 6 to 10 years. Adding these groups to form groups in practice 0 to 10 years we get 40% for 1994 and 27.8% for 2003. Thus, the 2003 practices tended to be smaller and somewhat older. To account for these differences, we will separate a number of our analyses by practice size and years in practice.

Current Computer Systems

The computer systems used have changed over the last decade. We can see from Table 2 that in 1994, 15.4% used individual PCs (non-connected, in even a local area network) while 17.5% in 2003 used individual PCs (non-connected). This remained about the same. However, there has been a big move to network-connected PCs.

Types of Business Applications Used

We have examined the types of computer equipment in group practice offices; we now turn to the types of applications used by these offices. This is important because when the doctors were asked if they were satisfied that the applications used met their business requirements, 86.7% of doctors in 1993 were. In 2003, the level of satisfaction had risen slightly to 91.0%. So while there have been complaints that physicians are not open to using information technology in hospital settings (Florien, 2003), physicians apparently are satisfied using IT in their own practices. Table 3 indicates the types of uses of computer applications for the 1994 and 2003 samples. The rank order of overall uses in 2003 is highly correlated with uses in 1994.

Since the earlier sample contained a larger number of larger medical practices, we will separate the study by group size to make the statistics more comparable (see Table 4). Generally, the order of importance of the usage of applications has

Table 3. Applications used in group medical practice, 1994 vs. 2003

Percent Who Used Application Applications	1994	2003
Personnel Scheduling	48.3%	26.9%*
Facility Scheduling	21.7	28.8
Patient Scheduling	51.7	55.8
Insurance Billing	95.0	86.5
Practice Billing	85.0	80.8
Business Record Keeping	78.3	65.4
Patient Record Keeping	53.8	46.7
Using Networking Software	51.9	30.0***
Using Hospital Network Software	32.7	11.7***
Vendor Networking	10.0	9.6
Expert Systems	10.0	11.5
Imaging Technology	--	11.5**
Voice Recognition	0	7.7**

† Spearman's Rank Order Correlation Coefficient rho = .949, p = .00001

** Significant at 0.05 level.*
*** Reflects differences in available services.*
**** Reflects the fact that practices were automatically networked.*
† There were at least three possible ways to compare the percentages: Spearmans' rho, Kendall's tau and gamma. Since the questions were the same but the samples were not paired, Spearman's rho was chosen.

Table 4. Application used by group practice size, 1994 vs. 2003

Applications	Groups Less Than 10		Groups 10 or Greater	
	1993	2004	1993	2004
Personnel Scheduling	42.3%	13.3%	48.0%	32.4%
Facility Scheduling	15.4	33.3	28.0	27.0
Patient Scheduling	53.8	53.3	50.0	56.8
Insurance Billings	100.0	100.0	92.0	81.1
Practice Billing	84.6	73.3	80.0	83.8
Business Record Keeping	80.8	53.3	80.0	70.3
Patient Record Keeping	34.6	66.7	52.0	48.6
Practice Networking	19.2	33.3	36.0	59.5
Hospital Networking	3.8	46.7	24.0	27.0
Vendor Networking	7.7	6.7	12.0	10.8
Use of Expert System	3.8	20.0	16.0	8.1
Imaging Technology	3.8	13.3	28.0	10.8
Voice Recognition	0	6.7	0	0
	rho = .767 p = .002		rho = .925 p = .00006	

remained the same (r = 0.949). This should come as a surprise since the basic business aspects of managing a business/practice have not changed. In 1995, a survey of physician managers found that the key issues they needed support in were personnel management, computing, budgeting, and financial management issues (Cordes, Rea, Rea, & Vuturo, 1995).

Practice networking is higher for large groups and seems to have risen for all groups since 1994. Hospital networking by computer for groups of less than 10 physicians *has risen considerably*

Table 5. Applications used by years in practice 1994 vs. 2003

Applications	Years in Practice					
	0 - 10		11 - 20		21 Plus	
	1994	2003	1994	2003	1994	2003
Personnel Scheduling	71.4%	13.3%	38.9%	38.9%	35.3%	26.3%
Facility Scheduling	33.3	33.3	11.1	33.4	17.6	21.1
Patient Scheduling	60.9	53.3	50.0	61.1	35.3	52.6
Insurance Billings	95.8	100.0	94.4	83.4	94.1	78.9
Practice Billing	87.5	73.3	77.8	83.4	88.2	84.2
Business Record Keeping	83.3	53.3	66.7	88.9	77.8	52.6
Patient Record Keeping	54.2	66.6	27.8	66.7	42.9	31.6
Practice Networking	37.5	40.0	27.8	72.3	17.6	47.4
Hospital Networking	8.3	46.7	11.1	38.9	17.7	15.8
Vendor Networking	12.5	6.7	11.1	11.1	5.8	10.5
Use of Expert Systems	0	20.0	16.7	11.1	17.7	10.5
Imaging Technology	16.7	13.3	16.7	5.6	17.6	15.8
Voice Recognition	4.2	0	0	16.7	0	5.3
	rho = .719		r ho = .771		r ho = .828	
	p = .00561		p= .003 p		= .00048	

from 3.8% to 46.7%. Vendor networking seems to have stayed the same.

We now consider the relationship between years in practice and usage of different computer applications (Table 5). We used rank order correlation to compare the relative amount of use of each of these applications by years in practice (Table 5), and the rankings have remained similar over the decade. However, different applications show differing amounts of use.

Importance of Different Types of Savings with Respect to Business Aspects of the Medical Practice

We then looked at the overall rankings of the importance of different types of savings or improvements to business practice that could be achieved by utilizing computers (Columns 5 & 6, Table 6). We asked respondents to rank the importance of different types of savings from computer adoptions on a scale from 1 to 5 (with 5 being very important). In more recent years (2003), respondents have been inclined to say that different savings are very important. On every one of 13 categories in Table 6, columns 5 and 6, the 2003 sample is 15% to 35% higher than their counterparts were in 1994. In both years, improving insurance claim processing was ranked as most important. In 1994, 70.2% said this, but by 2003, 85.8% found this type of savings most important. Other important savings issues in 2003, with 70% or more ranking them as important, were improving patient record keeping, access to patient or hospital information, increasing cash flow, and reducing administrative overhead.

The next area was the comparison of the importance of different types of savings by the age of the practice (see Table 7). These could be cost reduction or quality improvements. In Table 6, we classified the savings. Overall, the three groups (0 to 10, 11 to 20, 21+) order of importance (by use) of the applications was the same although the individual usage percentages differed widely.

Time Spent on Business Aspects of Practice

A very striking difference in the decade was the increased amount of time being spent by doctors on the business aspects of their practices. In 1994, 64.6% of the doctors said that they spent

Table 6. Importance of different types of savings from computer adoptions, 1994 vs. 2003 by a number of people in the group

		Less than 10		10 or More		All	
		1994	2003	1994	2003	1994	2003
CR	Reducing administrative overhead	59.3%	66.7%	44.0%	80.6%	54.0%	70.3%
CR	Improving insurance claim processing	70.4	86.7	68.0	83.3	70.2	85.8
QI	Improving patient record keeping	37.0	80.0	52.0	77.8	45.9	79.4
QI	Access to patient or hospital information	26.9	80.0	52.0	63.9	38.4	75.8
CR	Increasing cash flow	55.6	66.7	52.0	88.9	55.9	72.5
QI	Use of computer technology	22.2	66.7	33.3	66.7	23.4	66.7
QI	Use of image storage of patient records	14.8	60.0	8.0	52.8	13.3	58.1
CR	Reduce service bureau costs	28.0	33.3	21.4	50.8	24.5	37.8
QI	Enhance professional image	37.0	46.7	40.0	77.8	39.3	54.8
QI	Business training for support or staff	11.1	46.7	24.0	61.1	19.6	50.4
QI	Business management of practice	33.3	66.7	37.5	66.7	35.0	66.7
CR	Financial performance/ controls	44.4	66.7	52.0	72.2	47.5	68.1
QI	Business planning	22.2	60.0	36.0	69.4	27.5	62.4
		rho = .440 p = .133		rho = .775 p = .002		rho = .746 p = .003	

CR = cost reduction QI = quality improvements

less than 10% of their time on business aspects, while in 2003, only 23.1% spent less than 10% on business, indeed they spent far more time on these aspects. About a quarter of the doctors in 1994 said they spent up to 25% of their time on business, and 21.2% of the 2003 sample recorded this answer. The striking difference was that, in 2003, 55.7% of the doctors spent *more than 25%* of their time on business as compared to only 9.2% of the doctors in 1993, reporting that they spent more than 25% of their time on business issues (see Table 8).

Table 9 shows the importance of different types of savings by time spent on business aspects. One of the key changes from 1994 to 2003 was that, in 1994, both groups sited insurance claims processing as most important. On the other hand, in 2003, access to patient and hospital information had become most important thus reflecting a key shift in focus.

Table 7. Importance of different types of savings from computer adoptions, 1994 vs. 2003 by years in practice

	0 – 10 Years		11 – 20		21+	
	1994	2003	1994	2003	1994	2003
Reducing administrative overhead	54.0%	66.7%	70.6%	88.2%	37.5%	73.7%
Improving insurance claim processing	66.7	88.7	76.5	88.9	68.8	78.9
Improving patient record keeping	42.3	80.0	47.4	94.5	41.2	63.2
Access to patient or hospital information	28.0	80.0	31.6	82.7	58.9	47.4
Increasing cash flow	48.0	66.7	52.6	88.9	70.6	89.5
Use of computer technology	12.0	66.6	26.3	76.4	41.2	57.9
Use of image storage of patient records	8.0	60.0	15.8	58.4	17.7	47.4
Reduce service bureau costs	13.0	33.3	25.0	47.2	41.2	52.6
Enhance professional image	32.0	46.7	31.6	82.7	52.9	73.7
Business training for support or staff	12.0	46.7	21.1	58.4	29.4	63.2
Business management of practice	25.0	66.7	31.6	53.5	47.1	78.9
Financial performance/ controls	40.0	66.7	47.4	66.0	58.9	78.9
Business planning	16.0	60.0	36.8	66.7	35.3	73.7
	rho = .688		rho = .672		rho = .554	
	p = .009		p = .004		p = .049	

Table 8. Percent of doctor's time spent on the business aspects of their practice, 1994 vs. 2003

Percent of Time Spent in Business Aspects of Practice	1994	2003
Less than 10%	64.6%	23.1*
10% up to 25%	24.6	21.2
More than 25%	9.2	55.7*
No Answer	1.6	--
	100%	100%

Table 9. High importance of different types of savings from computer adoptions, 1994 vs. 2003 by time spent on business aspects

	Time Spent by Doctors on Business Aspects of Their Practice			
	Less Than 10 Percent		More Than 10%	
	1994	2003	1994	2003
Reducing administrative overhead	53.8%	16.7%	54.5%	27.5%
Improving insurance claim processing	64.1	25.0	81.8	27.9
Improving patient record keeping	46.1	50.0	45.5	64.9
Access to patient or hospital information	39.5	100.0	36.4	82.7
Increasing cash flow	59.0	83.3	50.0	80.2
Use of computer technology	23.8	50.0	22.7	70.3
Use of image storage of patient records	17.8	41.7	4.5	55.0
Reduce service bureau costs	31.5	41.7	11.1	55.0
Enhance professional image	38.5	41.7	40.9	35.0
Business training for support or staff	17.9	8.3	22.7	7.6
Business management of practice	34.2	0.0	36.4	12.6
Financial performance/controls	46.2	8.3	50.0	7.6
Business planning	20.5	8.3	40.9	9.9
	rho = .198		rho = -.163	
	p = .518		p = .594	

Correlating 1994 (< 10% vs. 10%+), rho = 0.884, p = 0.000006
Correlating 2003 (< 10% vs. 10%+), rho = 0.957, p = 0.0000003

Sources of Information on Running Business Aspects of Their Practice

The doctors in both surveys were asked, "Where do you get the information you need to run the business aspects of your practice?" This question suggested sources and allowed doctors to fill in blank lines. Also, it allowed respondents to check many answers. We will compare overall answers for the two surveys (Table 10). The ranking stayed approximately the same for 2003 as it had been in 1994 (rho = 0.813, p = 0.04). However in 2003, there was more reliance on associations and professional meetings and on professional journals. Use of consultants had tapered off, but vendors were more likely to be relied upon for information.

DISCUSSION

In viewing Table 2, we see that, in 1994, 35.4% of the offices had network-connected PCs while, in 2003, 61.4% had network-connected PCs. In 1994, 16.9% of the offices had mid-range or mini-computers while, in 2003, only 8.8% had these large computers. No doubt this is due to the enhanced capacities of today's PCs. No one in 2003 reported a network of PCs and mainframes, while, in 1994, about 7.7% used these facilities. The past decade has also shown a difference between the computing services that are available for physicians to use. For example, in 2003, 8.8% of the practices reported that they used outside service bureaus to handle their computer needs. Application Service Providers (ASPs) were not commonly available in 1994.

Table 10. Sources of information to run business

Percent Who Checked Answers Sources	1994	2003
Colleagues	55.4%	53.1%
Associations/Professional Meetings	46.2	67.3*
Professional Journals	46.2	59.2*
Business Publications/Newspapers	7.7	10.2
Consultants	55.4	34.7*
Universities	3.1	14.3
Vendors	21.5	38.8*
Other (CPAs, Billing Companies, Business Managers)	13.8	10.2
	rho = .813	
	p = .040	

Table 3 shows that, in general, the rank orders for application usage by practice size has remained the same in the last decade (rho = 0.767 and 0.925 respectively). Personnel scheduling by computer is done more often on the computer by large firms, but has gone down over the decade. Facility scheduling has gone up and patient scheduling has stayed the same or increased slightly. It seems that insurance billings for large firms using computers have decreased somewhat but for smaller firms 100% of insurance billing is done by computer. This difference may be due to age differences of physicians (older physicians tending to be in larger groups). We will assess the impact of age on IT usage later.

Looking at Table 4, we see the use of expert systems has gone up considerably in small practices from 3.8% to 20%. This is due to the fact that decision support systems have played an increasingly widespread role as the healthcare industry has embraced managed care (Dutta & Heda, 2000). This was predicted by McCauley and Ala (1992). One other area that has seen change is that of imaging technology. This may be due to the impact of telemedicine. In the early 1990s, imaging technology was very expensive, and only large groups and hospitals could afford it. Today, there is a minimal cost associated with it and new applications such as remote or telemedicine has attracted new physicians (Prater & Roth, 2003).

If we look at Table 5, we see that different applications have grown and fallen in importance depending on the age of the practice. The younger practices (0 to 10 years in practice) were more likely to use hospital networking, patient record keeping, and expert systems than their counterparts were in 1994 and 2003. They were less likely to use the computer for personnel scheduling, probably because they were in smaller practices. If we look at the middle group, 11 to 20 years in practice, they were more likely to do patient scheduling, patient record keeping, and practice networking than were their counterparts in 1994. So relatively new (10 years or less) and intermediate-aged practices (11 to 20 years) were more likely to use computers than the older practices (21+ years). In most other areas, the applications listed were utilized by the same percentage or less for each of the years in practice groups in 2003 as compared to 1994.

Table 6 shows that except for access to patient or hospital information, the same issues were very important in 1994 as in 2003 but not at the 70% level. No doubt because there was relatively little networking capacity in 1994, this issue was only ranked as very important by 38.4% of the sample. Assessing the importance of savings with relation to practice size (Table 6, Columns 1-4), the three most important issues for smaller practices (less than 10 physicians) in 2003 are improving insurance claims processing, improving patient record keeping, and access to patient hospital information. In 1994, the smaller practices chose improving insurance claims processing, increasing cash flow, and financial performance controls. So they were stressing the importance of the computer for billing and financial purposes. The smaller practices were not as likely to rank order these types of savings in the same orders in 1994 as they did in 2003 (rho = 0.44, p = 0.133). For the larger practices (10 or more physicians) in 2003, the most important savings were increasing cash flow, improving insurance claims processing, and reducing administrative overhead. For 1994, the top importance (for large practices) was for improving insurance claims processing. There were four other savings that tied for second place, improving patient record keeping, access to patient and hospital information, increasing cash flow, and financial performance controls. Uses devoted to quality improvements (QI) almost doubled in importance from 1994 to 2003. On the other hand, uses aimed at cost reduction increased in importance by 15 to 20 percentage points. Thus, it seems that early computer introductions were primarily devoted to cost reductions while later introductions may be more focused on quality improvements such as enhancing professional image, access to patient hospital information, and improving patient record keeping. Looking at the younger practices (0 to 10 years), we see that here again improving insurance claims processing was the most important issue in 1994 and 2003. Reducing administrative overhead was next in importance in 1994 and then increasing cash flow ranked third. In 2003, improved patient record keeping and access to patient and hospital information tied for second place. Thus in 2003, the informational and strategic uses of IT (such as long-term planning and marketing improvement) are becoming as important as transactional uses. However, the top uses were mainly those that led to cost reductions rather than quality improvements.

Using Table 7, and looking at the mid-term practices (11 to 20 years), again insurance claims processing was most important in both years, reduction of administrative overhead came in second in 1994 and third in 2003. Tying for first in 2003 was increased cash flow, and improving insurance claim processing. Tying for third in 2003 was a new variable enhanced professional image, and access to patient or hospital information. Here, we see the older doctors starting to focus on quality improvements. Finally, when we look at the older practices (21 years or more), increased cash flow comes out first for 2003 with improving claims processing, business management of the practice and business training for support or staff and financial controls vying for second place. Evidently, the older firms were sensing the need for updates and for more training and better business management practices. In 1994, the older practices ranked increasing cash flow first and improving insurance claims processing second. So over the years, the older practices have sensed the need for better business management and quality improvements. This need has probably been instrumental in the encouragement of the recent developments in management training and executive education for doctors and dentists (Glasser, 1997; Lazarus, 1999; Lipson, 1997).

With respect to the data from Table 8, we see that in 1994, 65% of the doctors spent less than 10% of their time of the business aspects of their practice. By 2003, 55.7% of the doctors reported that they spent more than 25% of their time on business aspects. With all of the different current issues such as HMOs, health insurance, Medicare, Medicaid, and so forth, doctors must personally make more decisions and consult with third parties to justify the use of medical techniques and billing. So we see that a decade ago, “physician” and “administrator” referred to two different people. Today, physicians realize that they have to embrace both titles, and their separate (and sometimes opposing) strategies (Lazarus, 1999; Lipson, 1997).

When we look at the importance of savings from computer adoptions and compare doctors who spend less than 10% of their time on the business aspects of their practice, we find that there is no correlation between the importance of various types of usage in the 1994 vs. 2003 (Table 9). In 1994, the three most important savings were improving insurance claims processing, increasing cash flow, and reducing administrative overhead. In 2003, access to patient hospital information, increasing cash flow, improving patient record keeping, and use of computer technology was the highest. Thus, the doctors in 2003 who spent little time on business stressed the patient information and record keeping. When we look at the doctors spending more than 10% of their time on business aspects in 1994 and 2003, we also find little correlation between the various types of usage. In fact, we find a somewhat negative correlation in what they found as the most important savings. In 1994, improving insurance claim processing, reducing administrative overhead, increasing cash flow, and financial performance controls were most important. In 2003, access to hospital patient hospital information and increasing cash flow were first and second in importance, and the use of computer technology came in third. These doctors in 2003 were also interested in reducing service bureau costs and image storing of patient records, which were not concerns for their counterparts in 1994 since image storage was not common in the early 1990s. Moreover, patient record keeping has grown tremendously in importance probably due to the new HIPPA legislation (Jonietz, 2003).

If we compare the two groups (under 10% and 10% plus) for 2003, there is a very high correlation (rho = 0.957 and p = 0.0000003, bottom of Table 9). Thus, for each year, separately, the time spent by doctors on business matters doesn't seem related to the importance they accord to different business aspects. However, over the decade, different business aspects have become more important.

This last assessment seems to be the overarching theme of the data. The changes that have occurred in IT usage during the last decade have been in response to physicians trying to adapt to greater business demands on them. Physicians are becoming more aware of this with new programs in medical schools focusing on business. Texas Tech, for example has instituted a combined MD and MBA program. As physicians become businesspeople as well as healers, they have had to embrace information technology in different ways. Apparently, the profession as a whole is adapting to this change. This is seen in Table 10, where the amount of business information provided by associations/professional meetings, professional journals and vendors has increased statistically significantly from 1994. On the other hand, the use of consultants has decreased significantly from 1994, where it was the number one source of information. This shows that physicians no longer have to pay experts to provide business information to them but are being supported by their profession.

CONCLUSION

Since the 1970s, researchers and practitioners have attempted to apply IT to the practice of medicine in order to increase efficiencies and decrease costs; however, the work is far from over. Technological leaps such as the Internet and the World Wide Web have been born and begun to impact medicine (Prater & Roth, 2002). The medical industry still has no clear common goals for IT and very few universally accepted standards. More than 90% of the $30 billion in health transactions are carried on via phone, fax or paper (Shine, 1996). As an example, in only 22% of clinics in North America can a clinician call up a medical record, input information, and enter orders (2002). To quote Tommy Thompson, the former

U.S. Secretary of Health and Human Services, “some grocery stores have better technology than our hospitals and clinics” (Turner, 2004). Other researchers have argued that we are beginning to see the first-level benefits of digitization such as increases in speed, control, accountability, and cost containment (Flower, 2003). This line of thinking is supported by the new federal 10-year initiative to “use Medicare as a vehicle for pilot programs ranging from handling prescriptions electronically to moving patient records online so that caregivers and patients can refer to them regardless of time or place” (Turner, 2004). Our paper shows that physicians have, in increasing numbers, embraced the use of IT in these areas. Thus, one of the benefits of this paper is to show the growth and trend lines of the various types of IT use by physicians for the past decade. We have also noted the move from primarily transactional uses of IT to informational and strategic uses.

Also, we have been able to look at the changes in infrastructure for group medical practices over the past decade. We have compared the different applications used over the decade by different-sized practices and by relatively new versus old practices. We have also measured the importance of different types of savings from computer adoptions and shown that, especially for small practices, the importance of these types of savings have changed significantly. Finally, we have shown the tremendous increase in time that physicians must devote to the business aspects of their practices.

We see a trend from the use of IT in transactional uses such as billing and insurance company dealings to more informational uses as patient record and contact with hospitals (Dutta & Heda, 2000). There has also been an increase in strategic emphasis of IT. This would include such uses as enhancing professional image use of expert systems and business planning, which have significantly increased in importance over the decade (Tables 3, 6). Moreover, we see big changes in infrastructure as discussed by Weill and Broadbent (1998). The number of practices with network-connected PC’s has doubled. Mainframe and mini-computers, which were used in 30.7% of practices in 1994, were used in only 12% of practices in 2003. Thus, the role of computers and the type of computer facilities employed in group medical practices have appreciably changed over the past decade.

One of the major problems of moving to IT systems is cost. A recent article in *USA Today* (Schmit, 2004) points out that wider use of computer software and hardware costs $10,000 to $20,000 per doctor and most of the benefits go to the insurers and hospital records. Doctor productivity drops 20% during the first 3 to 6 months after computer installation. Some possibilities for improvement are suggested by Bridges to Excellence, an employer coalition including GE and Ford Motor formed last year, which pays eligible doctors $50 per patient, per year, to use technology to improve healthcare.

While this paper provides a benchmark for the past decade’s use of IT by physicians, its best use is as a foundation for further research. First of all the consequences of the new HIPPA legislation may lead to more informational uses of IT in the physician’s offices (Jonietz, 2003). Second, while the overall use of IT in key clinical functions remains low (Networks, 2000, 2001), new technologies are being developed on an almost daily basis. Electronic patient records are being utilized more frequently (Hassey, Gerrett, & Wilson, 2001; Mondl, Szolonits, & Kohane, 2001). There is a need to determine the use of computerized physician order entry (CPOE) systems, disease registries, and pharmaceutical surveillance systems. Patient-centered Internet applications are expanding. An example is PatientSite, which allows access to patient’s medical records (except for clinical notes). It also allows patients to schedule appointments online. Other systems tie in hospital-based IT with private practitioners. An example would be eICUs where hospital specialists can remotely track as many as

105 patients in intensive care. If problems arise, direct intervention can be signaled to the patient's on-site medical staff, while the personal physician is being called (Turner, 2004). Another example is hospitals that provide online prescription renewal. These examples compose just a partial list. The key issue is that physicians are just beginning to utilize IT. In a recent study by BCG reported in the *Wall Street Journal* (Landro, 2002), only 42% of U.S. physicians are using electronic records or plan to do so in the near future. However, others have argued that, "at the end of the day, a physician's value will depend on whether he or she can still connect to patients in this cyberspace odyssey" (Healy, 2004).

So will the majority of physicians end up going the same route as the early movers or will they demand new IT capabilities? Will these new applications of IT come from the U.S. or overseas? Given that IT is interconnecting the world and governments worldwide strive to provide excellent medical services while containing costs, there is a need to glean information from worldwide practices. As of now, we do not know, but we will be applying an expanded form of this study in Taiwan this summer. The benchmarking done in this study, combined with periodic research assessing changes will provide the tools that researchers and practitioners need to best anticipate and manage the changes in medical practices that the future holds.

REFERENCES

Burt, C. W. (2005). Use of computerized clinical support systems in medical settings: United States, 2001-2003. *Medical Benefits, 22*(9), 9-11.

Cordes, D. H., Rea, D. F., Rea, J. L., & Vuturo, A. F. (1995). Training the future physician manager: Assessing the results. *Safety Science, 20*, 349-352.

Dorenfest, S. I. A., Ltd. (2002). *Clinical systems fuel IT spending.* Retrieved on August 15, 2002, from http://www.healthdatamanagement.com/html//ExpertStory.cfmDID=8959

Dutta, A., & Heda, S. (2000). Information systems architecture to support managed care business practices. *Decision Support Systems, 30*, 217-225.

Economist. (2005). The no-computer virus-IT in the health-care industry. *Economist, 375*(8424), 65-68.

Florien, E. (2003, November). IT takes on the ER. *Fortune*, *148*(11), 193-200.

Flower, J. (2003, Winter). Beyond the digital divide: Nothing about wiring physicians and hospitals is trivial. *Health Forum Journal, 46*, 8-15.

Glasser, J. P. (1997, June). Return on investment. *Healthcare Informatics*, 134-138.

Griffith, T. L., & Sobol, M. G. (2000). Negotiating medical technology implementation: Overcoming power and stakeholder diversity. *International Journal of Healthcare Technology and Management*, *2*(1-4), 375-392.

Hassey, A., Gerrett, D., & Wilson, A. (2001). A survey of validity and utility of electronic patient records in a general practice. *British Medical Journal, 322*(7299), 1401-1405.

Healy, B. (2004, August 2). 2004: A medical odyssey. *US News and World Report, 137*(3), 61.

Jonietz, E. (2003, April). Paperless medicine. *Technology Review, 106*, 59-66.

Landro, L. (2002, February 11). Unhealthy communication. *Wall Street Journal,* R12.

Lazarus, A. (1999, March). Down to business: A growing number of physicians recognize the value of an MBA. *Modern Physician*, *3*(3), 48-49.

Leroy, G., Chen, H., & Martinez, J. D. (2003). A shallow parser based on closed-class words to capture relations in biomedical text. *Journal of Biomedical Informatics, 36*, 145-158.

Lipson, R. A. (1997, November/December). Back to school. *Health Systems Review, 30*, 49-51.

Lowes, R. (2005). Expert help: Hire or outsource? *Medical Economics*, *82*(9), 23-25.

Magruder, C., Burke, M., Hann, N. E., & Ludovic, J. A. (2005). Using information technology to improve the public health system. *Journal of Public Health Management & Practice, 11*(2), 123-131.

Miller, R. H., & Sim, I. (2004). Physicians' use of electronic medical records: Barriers and solutions. *Health Affairs, 23*(2), 116-127.

Mondl, K. D., Szolonits, P., & Kohane, I. S. (2001). Public standards and patients' control: How to keep electronic medical records accessible but private. *British Medical Journal, 322*(7281), 283-287.

MRI. (2002). *MRI fourth annual survey of EHR trends and usage*. Newton, MA: Medical Records Institute.

Networks, H. H. (2000). *Health care's 100 most wired*. Hospitals & Health Networks.

Networks, H. H. (2001). *Health care's 100 most wired*. Hospitals & Health Networks.

Palattao, K. J. (2004). Essential EMR functions: A perspective from the front lines. *Health Management Technology, 25*(11), 22-25.

Prater, E. L., & Roth, W. (2002, August). Medicine and the Internet: Opportunities in cyberspace. *MGMA Connexion*, 24-27.

Prater, E. L., & Roth, W. (2003, January). Telemedicine: What's happening now? *MGMA Connexion*, 24-27.

Rodger, J. A., Pendharkar, P. C., & Paper, D. J. (1996). End-user perceptions of quality and information technology in health care. *The Journal of High Technology Management Research, 7*(2), 133-147.

Schmit, J. (2004, July 20). Healthcare's paper trail is costly route. *USA Today,* B1, 2.

Shine, K. I. (1996). Impact of information technology on medicine. *Technology in Society, 18*(2), 117-126.

Sobol, M. G., Alverson, M., & Lei, D. (1999). Barriers to the adoption of computerized technology in health care systems. *Topics in Health Information Management, 19*(4), 1-19.

Sobol, M. G., Humphrey, J. P., & Jones, T. (1992). Hospital size and adoption of computer based hospital information technologies. *Journal of Hospital Marketing, 7,* 173-195.

Sobol, M. G., & Smith, G. N. (2001). The impact of IT adoption on hospital staffing and payroll. *International Journal of Healthcare Technology and Management, 3*(1).

Sobol, M. G., & Woods, J. A. (2000). A ten year update on the marketing life cycle of hospital information technologies. *Health Marketing Quarterly, 15*(1/2), 71-85.

Tolle, K. M., Chen, H., & Chow, H.-H. (2000). Estimating drug/plasma concentration levels by applying neural networks to pharmacokinetic data sets. *Decision Support Systems, 30*, 139-151.

Turner, R. (2004, August 2). A high dose of tech. *US News and World Report*, 46-60.

Weill, P., & Broadbent, M. (1998). *Leveraging the new infrastructure: How market leaders capitalize on information technology.* Boston: Harvard Business School Press.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, edited by J. Tan, pp. 64-77, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 1.15
Medical Education in the 21st Century

Stefane M. Kabene
University of Western Ontario, Canada

Jatinder Takhar
University of Western Ontario, Canada

Raymond Leduc
University of Western Ontario, Canada

Rick Burjaw
University of Western Ontario, Canada

INTRODUCTION

As with many disciplines, the fields of healthcare in general and medicine, in particular, have made vast strides in improving patient outcomes and healthcare delivery. But, have healthcare professionals and medical academia been able to maximize the utilization of new technologies to improve the delivery of the right knowledge, to the right people, at the right time across geographical boundaries? In order to provide the best quality of care, regardless of patient or provider location, specific issues must be addressed.

Healthcare consumers and providers recognize that the system is often over worked, time constrained, poorly funded and desperately in need of a means to maintain up-to-date knowledge and efficient skills in order to deliver the best quality of care (Health Canada, 1998). We also know that there is a large disparity in both the quality and types of healthcare available between developed and developing countries (Lown, Bukuchi & Xavier, 1998). Within a single country there are also differences in healthcare services based upon location (rural vs. urban areas), wealth, age, gender and a host of other factors (Health Canada, 2004). However, because Information and Communication Technologies (ICT) can be a simple and cost effective tool, it can make desperately needed medical knowledge available to developing countries (Pakenham-Walsh, Smith & Priestly, 1997). Furthermore, it is becoming more difficult to get physicians and extended healthcare professionals to participate in face-to-face seminars in order to learn about the progress and changes in the delivery of healthcare. Time, travel requirements and cost are the biggest barriers to overcome. For rural areas and developing countries these issues are even more evident (Ernst and Young, 1998). Today, many institutions and countries are exploring and implementing ICT solutions to help reduce these inequities. The fact remains however that in the case of developing countries, a critical shortage of healthcare professionals remains (Fraser and McGrath, 2000). Adding to the problem is the fact that the telecommunications network, the backbone of ICT, in Africa is the least developed in the world (Coeur de Roy, 1997)

This article concentrates on two main aspects of ICT. First, it examines ways in which ICT can assist in information and knowledge transfer and second, it explores the challenges of ICT implementation.

ICT AND ITS ROLE IN MEDICAL EDUCATION

Providing the right medical knowledge and training to healthcare professionals can be a challenge in the best of circumstances. In developing countries dissemination of the best clinical practice protocols at an affordable cost regardless of the location of the targeted audience is even more daunting. While technology such as CD-ROM-based learning can be tremendously efficient in helping medical students learn fast and well, there may be a lack of individual access to the necessary infrastructure such as equipment and power. In these cases the material is often used in classrooms and the goal of facilitating individual learning and allowing students to go at their own pace may not be met (Pakenham-Walsh, 2003). It is important to keep in mind that even developed countries have, in spite of relatively easy access to the necessary technology, difficulties in properly managing its introduction and use in medical schools (Greenhalgh, 2001).

Tele-education can help in reaching remote communities in developing countries. However, in many regions, technological compatibility and training remain a challenge (Pakenham-Walsh, 2003). The birth of the Internet, in spite of its many imperfections, has dramatically changed the way information, communication and learning are delivered. Although there still exists an imbalance (Davison, Harris, Vogel & Vreede, 1999), in many ways the Internet Age can reduce the gap between developed and developing countries in terms of access to all types of information (United Nations ICT Task Force-1, 2004). While the full potential of the Internet as an ICT tool has yet to be defined it does allow access to information at a speed, quality, and cost previously unseen and defies the notion of geographical boundaries. The Internet allows access to medical information through online articles, video presentations, videoconferences, e-mail-based information and communication. Today, a medical student or healthcare practitioner in India, Africa or any developing country can gain access to the latest medical information from around the world. This new reality has set the foundation for a truly worldwide medical and healthcare community of practice. However, the lack of technological infrastructure and training in developing countries can affect the delivery and use of this health care information (Lown et al., 1998).

As mentioned earlier, users of ICT in the healthcare field face many challenges. Yet despite these challenges, we are at the beginning of a marvelous adventure that has the potential to create a healthcare arena for all with unrestricted access to information and knowledge for practitioners around the world. While this could be viewed by some as simple, utopian rhetoric, the group "Doctors without Borders (Medecins sans Frontières)" already demonstrates the willingness of medical communities to create a vast exchange of information and knowledge (Orbinski, 2000). The term "doctors without borders" captures the spirit of the new world order and sets the future stage for sharing healthcare information. This growing global community of healthcare workers may bring about the desperately needed improvement in the availability of healthcare information around the world (Jareg & Kaseje, 1998).

Despite the great potential, we need to both understand and address the limitations of ICT as a tool in the acquisition and transfer of information and knowledge. The Achilles' heel of the ICT user is not, in our opinion, in terms of access to information, but rather in terms of prior training in healthcare, possession of the necessary technical skills, and an efficient infrastructure (United Nations ICT Task Force-2, 2004). These are the basic requirements in understanding and transforming the rich information that is available into meaningful and useful medical knowledge. When the introduction of ICT is paired with proper access and training, it can lead to the successful creation of a community ready, willing and able to use the ICT to its fullest extent possible (United Nations ICT Task Force-2, 2004).

While such a community can create, share and apply important healthcare knowledge on a wide range of issues, it has to be done accurately and in a timely manner in order to be beneficial (Using ICT to Empower Communities, 2003).

An additional benefit of ICT is the degree to which it can help reduce the sense of isolation often felt by healthcare professionals, especially in rural areas. As a result, staff morale can be improved (Using ICT to Empower Communities, 2003; Ballantyne, 2003).

At this point, the question arises: Can a viable knowledge community exist through the use of ICT?

Before answering, one must differentiate between ICT as a tool on the one hand and the knowledge it can help foster and transfer across boundaries on the other hand. Unlike Marshall McLuhan, who stated that the medium was the message (McLuhan, 1994), we say that ICT is not the knowledge, but only its channel.

FOSTERING A KNOWLEDGE COMMUNITY

Up until now we have discussed the use of ICT as a method of sharing information and knowledge. However, there are conflicting opinions regarding the differences between information and knowledge and the transferability of knowledge (Wilson, 2002).

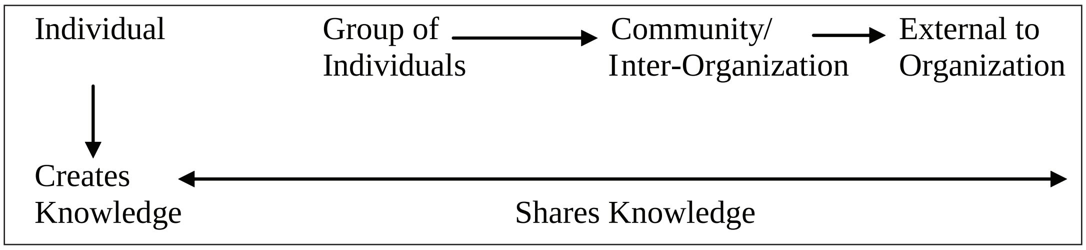
Exploring the Definition of Knowledge

What is knowledge? How can it be defined? These questions have been posed for thousands of years and answered with varying degrees of clarity. Plato described knowledge as justified true belief. Other definitions have been as simple and direct as what we know (Wilson, 2002), or as detailed as a mix of experience, values, information and insights that are applied in the minds of individuals as well as embedded in organizations (Davenport & Prusak, 2000).

Related to knowledge, but not interchangeable, are data and information. Although they are sometimes used interchangeably with each other and with knowledge, they are different. Data is considered the most basic level of discrete facts. Information is more sophisticated than data in that it is a set of related facts. Both data and information, while important, are not knowledge. They generally have no meaning attached to them and they are much more likely to be objectively measured and agreed upon by those preparing and using them. They are also likely to be more generally available than knowledge.

While related to each other, these three concepts are not interchangeable. Rather, there is a hierarchy. It is possible to have data and informa-

Diagram 1. Hierarchy of knowledge creation and sharing within a community



tion without knowledge but *it is not possible to have knowledge without data and information*.

If there is no knowledge without first having access to data and information, then ICT, by allowing the transfer of data and information anywhere, anytime, allow individuals across the world to have the fundamental resources needed for the creation of knowledge. This makes ICT a critically important aspect of knowledge creation and transfer.

Diagram 1 describes a hierarchy of knowledge creation and sharing within a community. Note that knowledge also flows back to the individual.

In this paradigm, ICT, although not perfect, becomes the arteries of a mega community linked by these common objectives.

Classical medical education activities delivered through traditional means such as seminars, perceptorships, residency programs, and workshops seem quite inefficient and costly, poorly coordinated, supply driven, and the content of the information and learning provided is frequently not relevant to the diverse needs of today's rural healthcare workers, especially in the third world (Ballantyne, 2003).

ICT, by allowing healthcare professionals to interact together and with the public, can dramatically change the way medicine is practiced by allowing the timely and unlimited exchange of an ever-increasing quality and quantity of information and knowledge. The role of ICT in the healthcare domain has been so important that in some countries, such as Malaysia, it has been regulated and important financial and material support is provided (Minges & Gray, 2004). The number of online programs of continuing medical education has, for example, increased by 110% in 1999 compared to 1998 (Reynolds, 2002). This shows a definite trend towards the acceptance and use of ICT in the field.

In addition to providing fast and easily accessible information, ICT helps create ties across geographical borders. When asked about their expectations of the use of ICT, healthcare professionals ranked sharing of information and experience number one (Roundtable on ICT for Continuing Medical Education, 2003). The act of sharing helps to create a deeper sense of community (Bwalya, 2003; Grunwald, 2003; Roos, 2001).

Moreover, when it comes to poorer and rural areas, ICT can assist in reducing costs by minimizing redundancy and mismanagement of treatment by enabling healthcare professionals to have readily available clinical information about their patients. An economic argument can be made for greater use of ICT tools in that it will lead to more efficient use of available healthcare resources (Denz, 2003).

In rural areas of developing countries, the successful implementation of ICT requires addressing the management, maintenance and user support issues (Verboom, 2003) which are obviously important aspects and may be easier to deal with in cities and developed countries.

ICT has completely and forever changed the way healthcare education can be delivered (Finkelstein, 2003; Greenhalgh, 2001).

In order to be successful, the use of ICT tools in the healthcare field must be driven by a vision and by policies that recognize the potential. As well, the involvement and buy in of the multiple stakeholders-medical academia, governments at national, regional and local levels, and healthcare practitioners and consumers is critical.

CHALLENGES FOR ICT IN DEVELOPING COUNTRIES AND RURAL AREAS

Table 1 summarizes some of the critical ICT issues faced by developing countries.

The inequities described in Table 1 undoubtedly create a digital divide (Cullen, 2003) making an efficient use of ICT difficult in developing countries (Moghaddan & Lebedeva, 2004). In healthcare, this is unacceptable as it could mean the difference between life and death. Healthcare professionals as a community without borders must be accepted and supported and ICT tools must be developed and made available by international agencies. Access to information is instrumental to the success of healthcare systems in developing and transitional economies (Pakenham-Walsh, 2000). In developed countries however, resistance to change seems to be the main issue where older generations of healthcare professionals are less familiar with the use of ICT and sometimes prefer not to use them (Peterson, 1999).

CONCLUSION

ICT tools are indispensable in providing desperately needed information and knowledge to healthcare professionals regardless of their geographical location. Much has been done and, while improvements continue, more remains to be accomplished (Dash, Gowman, Traynor, Jones & Tait, 2003).

The creation of an international task force to encourage access to ICT tools by healthcare professionals across the world is essential. It will help reduce the gap in medical knowledge, improve infrastructure and quality of care among rural and urban areas as well as amongst developing and developed countries. Quality of care should become a right, not a privilege. We must always keep in mind the patient and improved patient outcomes on a worldwide scale (Towle, 1998).

We believe that healthcare ICT communities will be instrumental in achieving this objective.

Table 1. Critical ICT issues faced by developing countries

-Scarcity and/or costs of telephone lines -Unreliability of telephone lines and Internet connectivity -Lack of skilled workers to properly use the technology -Lack of vision and support from policy makers -Political instability leading to: Prohibitive costs Poor infrastructure Poor funding Electricity shortage -Low priority for ICT against the urgent need to spend the money on medication instead

REFERENCES

Ballantyne P. (2003). *Using ICT to enhance the skills of rural health workers.* iConnect online. Retrieved from http://www.iconnect-online.org/

Bwalya, Y., (2003) *Ensuring local ownership, participation and sustainability of projects.* International Institute for Communication and Development. Retrieved from http://www.iicd.org/stories/articles/Story.import5200

Coeur de Roy, O. (1997). The African challenge: Internet, networking and connectivity activities in a developing environment. *Third World Quarterly,* 18(5), 883-899.

Cullen, R. (2003). The digital divide: A global and national call to action. *The Electronic Library,21*(3), 247-258.

Dash, P., Gowman, N., Traynor, M., Jones, B. G., & Tait, G. (2003). Increasing the impact of health services research. *British Medical Journal, (International edition), 327*(7427), 1339-1342.

Davenport, T. H. & Prusak, L. (2000). *Working knowledge: How organizations manage what they know.* Harvard Business School Press.

Davison, R., Roger Harris, R., Vogel, D., & Vreede, G. (1999). Information technology in developing countries. *Proceedings of the 32nd Hawaii International Conference on System Sciences.*

Denz, M. D. (2003). *Telemedicine in daily clinical work. Now? Tomorrow?* International Communication Union. Retrieved from http://www.itu.int/itudoc/itu-t/workshop/ehealth/addinfo/info000.html

Ernst & Young LLP. (1998). *Healthcare issues facing the "G20" countries: A meeting between healthcare governors and health ministers.* Retrieved from http://www.ey.com/global/download.nsf/International/G20_Health_Meeting_WEF2003/file/G20%20 Health%20Meeting%20WEF2003.pdf

Finkelstein, M. (2003, September). The morphing of the American academic profession: Every societal transformation in the role of higher education has brought in its wake a concomitant transformation in the faculty role-Featured Topic. *Liberal Education*, 1-4.

Fraser, H. S. F., & McGrath J. D. (2000). Information technology and telemedicine in sub-Saharan Africa. *British Medical Journal, (International edition), 321*(7259), 465-467.

Garner, P. (1998). Implementing research findings in developing countries-getting research findings into practice, part 8. *British Medical Journal, 317*, 531-535.

Greenhalgh, T. (2001). Computer assisted learning in undergraduate medical education-education and debate. *British Medical Journal, 322,* 40-44.

Grunwald, T. (2003). *Characteristics of community networking (CN).* Retrieved from http://www.making thenetwork.org/toolbox/tools/twelve.htm

Health Canada. (2004). *Health care in Canada.* Retrieved from http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page =PG_263_E& cw_topic=263 &cw_re=AR_4_ E#full

Health Canada and The Canadian Medical Association. (1998). *Moving toward a quality health system: Key challenges and strategies.* Retrieved from http://www.hc.sc.gc.ca/hppb/healthcare/pubs/mtqhs/

Internet Healthcare Coalition. (n.d.). *eHealth code of ethics.* Retrieved from http://www.ihealth coalition.org /ethics/ehcode.html

Jareg, P., & Kaseje, D. C. O. (1998). Growth of civil society in developing countries: Implications for health. *The Lancet, 351*(9105), 819-823.

Kalu, K. (2000). Perspectives on technology and economic development in Nigeria. *Journal of Third World Studies, Americus, 17*(1), 91-115.

Lown, B., Bukachi, F., & Xavier, R. (1998, October). Health information in the developing world. *The Lancet,* S34-39.

McLuhan, M., & Lapham, L. H. (1994). *Understanding media: The extensions of Man.* MA: MIT Press. Reprint edition (October 20, 1994).

Minges, M., & Gray, V. (2004). *Multimedia Malaysia: Internet case study.* Retrieved from http://www.itu.int/ITU-D/ict/cs

Moghaddam, F. M., & Lebedeva, N.M. (2004). Carriers, dual perceptions, and the information communication revolution. *Educational Technology, Research and Development, 52*(1), 83-88.

Nonaka, I., & Takeuchi, H. (1995) *The knowledge creating company.* Oxford: Oxford University Press.

Orbinski, J. (2000). *Medecins sans frontières (doctors without borders).* CMAJ. Retrieved from http://www.cmaj.ca/misc/audio/orbinski_0200.shtml

Pakenham-Walsh, N. (2000). Improving access to reliable information in developing countries. Letter to the editor. *British Medical Journal, 323*, 1-2.

Pakenham-Walsh, N. (2003). *ICT in continuing medical education in eastern and southern Africa.* Retrieved from http://www.dgroups.org/groups/ict4cme/index.cfm?op=dsp_show msg&list name=ict4cmr &msgid=71252&cat_id=2753

Pakenham-Walsh, N., Smith, R., & Priestley, C. (1997). Meeting the information needs of health workers in developing countries. *British Medical Journal, (International edition), 314*(7074), 90.

Peterson, M. (1999). Realizing the promise: Delivering pulmonary continuing medical education over the Internet. *Chest, 115,* 1429-1436.

Polanyi, M. (2000). *The tacit dimension.* New York: State University of New York Press.

Reynolds, S. (2002, January 29). Wired physicians cite Internet as preferred format for CME according to Medsite research survey; 97% Say Online Continuing Medical Education is Profession's tool of the Future. *Business Wire.*

Roos, G. (2001). *Establishing communities of teachers and learners online in South Africa.* SchoolNet SA-Educator Development for ICT. Retrieved from http://www.school.za/edict/impact/forum2004-1.htm

Round table on ICT for Continuing Medical Education for Rural Health Workers in Uganda. (2003). Kampala, November 2003. *Proceedings of the Round Table Workshop.* November 2003. Kampala. Uganda. Copyright: Makerere University and IICD, 2003.

Towle, A. (1998). Changes in health care and continuing medical education for the 21st century-continuing medical education, part 1. *British Medical Journal, 316,* 301-304.

United Nations ICT Task Force-1. (2004). *Wireless internet institute unveils wireless universal connectivity initiatives.* Retrieved from http://www.unis.un vienna.org/unis/pressrels/2004/pi1563.html

United Nations ICT Task Force-2. (2004). *Council adopts ministerial declaration calling education, health crucial to human resources development.* Retrieved from http://www.unis.unvienna.org/unis/pressrels/2002/ecosoc6011.html

Using ICT to empower communities - The WVK Bunyala Experience. (2003, August). AfriAfya, African Network for health knowledge management and communication, 1(4). Retrieved from http://www.afriafya.org/Documents/AfriAfya%20August%202003%20Issue4.pdf

Verboom, J. J. (2003). *Continuing medical education using ICT.* KiNcite. Retrieved from http://www.kincite.com/DigitalCommunities/ContinousMedicalEducation_ICTs.pdf

Wilson, T.D. (2002). The nonsense of knowledge management. *Information Research, 8*(1). Paper No. 144.

KEY TERMS

CD-Rom (Compact Disc Read-Only Memory): Optical data storage medium using the same physical format as audio compact discs, readable by a computer with a CD-ROM drive.

CME (Continuing Medical Education): All learning by healthcare providers, after basic training.

Community:A unified body of individuals, people with common interests living in a particular area.

Digital Divide: The phrase has been applied to the gap that exists in most countries between those with ready access to the tools of information and communication technologies (ICT), and those without such access or skills.

E-Health: Used to characterize not only "Internet medicine," but also virtually everything related to computers and medicine.

Information Technology (IT) or Information And Communication Technology (ICT): The technology required to convert, store, process, transmit, and retrieve information.

Internet, The: The vast collection of interconnected networks that all use TCP/IP protocols.

Preceptorship: Teaching or tutoring.

System: A group of devices or an artificial objects or an organization forming a network, especially for distributing something or serving a common purpose.

Telehealth: The use and transmission of video, voice and text data for a multitude of health–related issues, including, health management, patient care, and health worker training and education, individual and patient education on health maters.

This work was previously published in Encyclopedia of Developing Regional Communities with Information and Communication Technology, edited by S. Marshall, W. Taylor, and X. Yu, pp. 488-493, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.16
Knowledge Management in Healthcare

Sushil K. Sharma
Ball State University, USA

Nilmini Wickramasinghe
Cleveland State University, USA

Jatinder N.D. Gupta
University of Alabama in Huntsville, USA

ABSTRACT

Healthcare organizations are facing many challenges in the 21st Century due to changes taking place in global healthcare systems. Spiraling costs, financial constraints, increased emphasis on accountability and transparency, changes in education, growing complexities of biomedical research, new partnerships in healthcare and great advances in IT suggest that a predominant paradigm shift is occurring. This shift is necessitating a focus on interaction, collaboration and increased sharing of information and knowledge which is in turn leading healthcare organizations to embrace the techniques of Knowledge Management (KM) in order to create and sustain optimal healthcare outcomes. This chapter describes the importance of knowledge management systems for healthcare organizations and provides an overview of knowledge management technologies and tools that may be used by healthcare organizations.

INTRODUCTION

Knowledge Management (KM) is an essential tool in today's emerging healthcare system. Hospitals that seek to deploy KM systems need to understand the human element in the process. Earlier, success factors were only restricted to a few healthcare variables such as patient care and cost, but over the years, technology (both clinical and administrative) has evolved as a differentiating variable, thus redefining the doctrines of competition and the administration of healthcare treatments. For example, in today's healthcare environment we are now treating patients with an emphasis on prevention and managing the patient through good health throughout their life. Such an approach requires significant investment in knowledge assets. One of the key objectives of a KM system is to insulate a hospital's intellectual knowledge from degeneration (Elliot, 2000).

Most hospitals are unaware of their acquired knowledge base. Further, knowledge capital is often lost from a hospital through employee attrition, high turnover rates, cost-saving measures and improper documentation (Chase, 1998). Specific KM tools and metrics help focus the hospital on acquisition, retrieval and storage of knowledge assets both tangible and/or other for activities such as learning, strategic planning and decision making (Oxbrow, 1998). This goes a long way in crafting a coherent and well-designed growth plan for the hospital (Allee, 1997, 1999). KM treats intellectual capital as a managed asset. Improved patient care is directly proportional to a hospital's intellectual assets. The tactical expertise and experience of individual workers should be fully captured and reflected in strategy, policy and practice at all levels of the hospital management and patient care activity (Conklin, 1998). The intangible asset of knowledge of the employee can nurture radical innovation in advance planning, change management, hospital culture and well balanced approaches. Fostering a knowledge-sharing attitude and competency of patient care processes are vital for any KM program in healthcare (Burca, 2000; Matheson, 1995). Hospitals managing and sharing their knowledge assets effectively will have benefits of cycle time reduction, cost reduction, improved return on investment, higher satisfaction index, and better medical and paramedical education levels (Antrobus, 1997; Atkins et al., 2001).

KNOWLEDGE MANAGEMENT

Knowledge Management (KM) is an emerging, interdisciplinary business model dealing with all aspects of knowledge within the context of the firm, including knowledge creation, codification, sharing and how these activities promote learning and innovation (Choo, 1998). Unfortunately there's no universal definition of KM, just as there's no agreement as to what constitutes knowledge in the first place (Beckman, 1999). For this reason, it's best to think of KM in the broadest context:

KM is a discipline that promotes an integrated approach to identifying, managing, and sharing all of an enterprise's information assets, including database, documents, policies and procedures, as well as unarticulated expertise and experience resident in individual workers (Wickramasinghe, 2003). There are many dimensions around which knowledge can be characterized such as storage media, accessibility, typology and hierarchy. Each of these dimensions is explained in this chapter (Brailer, 1999; Broadbent, 1998; Skyrme, 2001, 1999, 1998; Davenport & Prusak, 1997, 1998).

Knowledge Storage Media

There are several media in which knowledge can reside including: the human mind, an organization, a document and/or a computer. Knowledge in the mind is often difficult to access; organizational knowledge is often dispersed and distributed; document knowledge can range from free text to well-structured charts and tables; while computer knowledge can be formalized, sharable and often well structured and well organized. In order to effectively manage KM it is important to pay careful attention to the most useful storage media.

Knowledge Accessibility

Intellectual and knowledge-based assets fall into one of three major categories (Nonaka, 1994; Nonaka & Nishguchi, 2001; Sharma & Wickramasinghe, 2004):

- *Tacit (human mind, organization):* accessible indirectly only with difficulty through knowledge elicitation and observation of behavior.
- *Implicit (human mind, organization):* accessible through querying and discussion, but informal knowledge must first be located and then communicated.
- *Explicit (document, computer):* readily available, as well as documented into formal knowledge sources that are often well organized, often with the help of IT.

In order for effective KM to ensue, it is necessary to understand these categories of knowledge as well as their subtle nuances.

Knowledge Typologies

Typologies are defined, characterized and described in terms of knowledge type-conversion, structural features, elementary properties, purpose and use, and conceptual levels. Knowledge typologies play an integral role in a robust KM system.

Knowledge Hierarchy

A further dimension considers the premise that knowledge can be organized into a hierarchy. Several authors draw distinctions between data, information and knowledge (Allee, 1997; Devenport & Prusack, 1998; Leonard, 1998).

- *Data:* Facts, images and sounds.
- *Information:* Formatted, filtered and summarized data.
- *Knowledge:* Instincts, ideas, rules and procedures that guide action and decisions and are context dependent.

Strictly speaking, KM is a process that identifies and creates knowledge assets. In healthcare, KM optimally utilizes these assets to fulfill the core healthcare objectives. Knowledge assets in hospitals are intangible but they can be defined as knowledge that a medical/paramedical/non-medical person has with respect to patient care, medical needs, operating environment and technologies, something that he or she can utilize in routine medical and healthcare management. KM identifies and maps intellectual assets within the hospital, thereby generating precious knowledge capital for competitive advantage within the medical institution. Since knowledge is dynamically imbedded in organizational and healthcare networks and processes as well as in the staff that use them, hospitals need to have a built-in KM system that "crisscrosses" with its healthcare networks (Jackson, 2000). The result is that employees will then be better informed and continuously updated on the latest tools and best practices. It is important that the techniques adopted to enable KM must take into account some basic factors such as the type of hospital, its culture and its needs to ensure a successful KM system for a healthcare setting (Gokce, 2002; Johnson, 1998; Keeling & Lambert, 2000).

THE NEED FOR KM IN HEALTHCARE

The health sector is large, accounting for between 6% to 12% of GDP across OECD countries. Though the use of healthcare services varies between nations, public expectations of them globally have risen dramatically everywhere since 1950 and the trend is still upwards. Fresh demands arise from the appearance of new drugs and the invention of new technology, from advances in prevention and diagnosis as well as therapy, and from new categories of demand, such as care of the elderly (Eisenberg, 2002). The health sector is complex and includes a range of key actors: patients, providers, practitioners, payers, purchasers, pharmaceutical industry, and professors. The interaction among these actors shapes what counts as relevant knowledge as well as how it is produced, mediated, and used (Conner, 2001). Further, the domain of medical knowledge has expanded to such a degree that a human mind can no longer retain it all. There are now some 20,000 medical journals in the world. A professor of medicine spends on average one day a week to remain abreast of studies in his/her field of interest as well as for his/her research. What can a generalist physician do? How much time can he/she devote to "keeping up"? In France, there are some 7,000 prescription drugs based on some 3,500 active ingredients. A physician has the right to prescribe them all. Can he/she be familiar with all of them? He/she must also be aware of some 300 medical references, some 800 biological tests, more than 1,000 imagery tests, and more than 1,500 surgical interventions. If he/she prescribes six drugs, he/she must also be aware of some 720 potential sources of interaction. The figure reaches 3,328,800 if 10 drugs are prescribed. In addition to the therapeutic value of each molecule, the physician should also know their price and potential effect on specific population groups (diabetics, the obese, children, the elderly, etc. The growth in knowledge has necessarily led to specialization, which too has meant "balkanization" and lack of coordination, especially in hospitals but also in private practice (Halpern, Perry & Narayan, 2001; Dean, 2002). As knowledge is shared, responsibility and decisions about treatment should be shared as well (Eisenberg, 2002). Hospital Information Management (HIM) professionals, like other healthcare personnel, have always sought, used, and valued knowledge about their practice. Managers hire experience because they understand the value of knowledge that has been developed and proven over time. Unfortunately, they are bombarded each day with information in the form of e-mails, voice mails, faxes, reports, memos, and so on — much of which is repetitive or simply not useful. On the other hand, the same professionals spend a great deal of time looking for the information they need by accessing the Web, sending e-mails, making phone calls, and scouring computerized reports. It is in this process that KM can make a difference. Studies have shown that managers get two-thirds of their information from face-to-face meetings or phone conversations. Only one-third comes from documents or other knowledge bases. Unlike material assets, which decrease in value as they are used, the value of knowledge increases with use. For example, new ideas on records storage or retrieval breed other new ideas, and shared knowledge stays with the giver while it enriches the receiver (Jadad, Haynes, Hunt & Browman, 2000). The potential for new ideas arising from the store of knowledge in any healthcare organization is to provide a common entry point for corporate knowledge, such as formularies, clinical road maps, and key financial indicators (Einbinder, Klein & Safran, 1997). Thus what we can see is that the need for KM in healthcare is critical and becomes significant when it begins to focus on the needs of individual users, departmental indicators, and key processes in order to capture and display relevant, useful, and usable knowledge in a customized fashion (Sorrells & Weaver, 1999a, 1999b, 1999c).

APPROACHES TO CAPTURE, STORE AND SHARE KNOWLEDGE

For companies that need to leverage their corporate knowledge, the following four initiatives may help you establish a knowledge sharing system of your own. These initiatives draw upon a predominate, repository model but are also relevant to other models (Morrissey, 1998).

Build the Infrastructure Using Appropriate Technology

Technology enables connectedness to take place in ways that have never before been possible. Harnessing intellectual capital can be expedited through a network-computing infrastructure. Technology has emerged to support each different approach to knowledge management. Document management systems expedite document storage and retrieval. Web-casts allow synchronous communication between experts while discussion groups enable asynchronous interaction. Learning management systems track an employee's progress with continuous learning while data warehousing mines powerful SQL databases, which organize and analyze highly structured information. Paramount to the successful use of these technologies is naturally a flexible, robust IT infrastructure (Sharma & Wickramasinghe, 2004).

Build a Conceptual Infrastructure with Competencies as the Backbone

Technology is important in harnessing intellectual assets, but integrated solutions encompass more than that. You must rethink the conceptual infrastructure of your business. For example, you may need to: ensure intellectual assets reflect your vision and values; articulate the theoretical framework for your processes; establish a taxonomy or categorization scheme to organize your information; create cross references that reflect relationships between entries; or index your information using attributes or meta-tags. The notion of competence plays a critical role in knowledge indexing and sharing. Karl Erik Sveiby, noted Swedish expert on managing and measuring knowledge-based assets, observes, "The concept of competence, which embraces factual knowledge, skills, experience, value judgments and social networks is the best way to describe knowledge in the business context." Once competencies or target proficiencies are defined, they become the backbone, which connects users to useful, relevant knowledge.

Create a Repository of Reusable Components and Other Resources

Before the Industrial Revolution, products were handcrafted; each piece was unique and couldn't be reused. The genius of the Industrial Revolution centered on making reusable parts and components became standardized and interchangeable. The Information Revolution is similar. Instead of crafting a unique solution each time, knowledge sharing creates a warehouse of "stored parts" — e.g., standardized and interchangeable components which can be reused and adapted: skills, best practices, models and frameworks, approaches and techniques, tools, concepts, specific experiences, presentation aids, white papers, etc. Adding to this, resources such as directories of experts indexed by their field can help you gain access to knowledge outside of your core competencies.

Set High Standards for Quality and Usability

Ensure that your information complies with high quality standards because it is the foundation upon which to build a knowledge-centric healthcare organization. In addition, it is important to make sure that the system meets the users' needs, which may involve reworking or restructuring information. Leonard Caldwell observed: "Critical information must first be reorganized so that information is presented in a way that mirrors users' needs and parallels a thought process occurring within a job function or task." Establishing consistent patterns helps end-users find information quickly. Online coaching can provide users with tips and techniques on how to modify, customize, or tailor information. Leveraging organizational knowledge is not an option — it is an imperative if one is to flourish in the marketplace. It can lead into a new phase of quality and innovation.

It will reduce cycle time and gives a competitive advantage as a company. The synergy will also contribute to growth as individuals. Companies who have the foresight to manage their knowledge capital now will have an advantage in the future (Herbig, Bussing & Ewert, 2001).

KM TOOLS AND TECHNOLOGIES

The paraphernalia of the information revolution — computers, communications networks, compact discs, imaging systems and so on — are now widely expected to make a vital contribution to helping doctors and other medical professionals do their work better (Gokce, 2002). New information technologies include:

- Electronic patient records, which are more up to date, easier to access, and more complete than paper ones;
- Standardized medical terminologies and languages, both within and across natural language communities;
- Methods and tools to support faster dissemination of information via the Internet that leads to new scientific understanding of diseases and their treatment;
- More timely and reliable methods and tools to support better communication and coordination among members of healthcare teams;
- A creative approach to KM can result in improved efficiency, higher productivity and increased revenues in practically any business function.

The technical goal of KM initiatives is to give the organization the ability to mine its own knowledge assets, which could include creating such tools as a centralized search capability, automatic indexing and categorization, content analysis and preparation, data analysis, and customizable features integrated in a digital dashboard. Process improvement is a precursor to providing a knowledge-centered environment. Before an organization can foster collaboration and knowledge sharing, the organization must possess an understanding of information flows and of the overall knowledge infrastructure. There is no such thing as the perfect KM product. Instead, different tool sets can be integrated with the organization's legacy systems (Heathfield & Louw, 1999). Technical issues that KM projects must address include:

- Setting up electronic delivery strategies for information
- Identifying information sources and services
- Building decision support tools and data-mining templates
- Establishing enterprise-wide business rules
- Implementing process improvement techniques

Knowledge Mapping

Knowledge management is rapidly becoming a critical success factor for competitive organizations. Carrying out knowledge management effectively in an industrial environment requires support from a repertoire of methods, techniques and tools, in particular knowledge engineering technology adapted for knowledge management. Knowledge mapping creates high-level knowledge models in a transparent graphical form. Using knowledge maps, management can get an overview of available and missing knowledge in core business areas and make appropriate knowledge management decisions. Knowledge mapping is a good example of a useful knowledge management activity with existing knowledge acquisition and modeling techniques at its foundations (Strawser, 2000). Knowledge mapping is a technique rather than a product. A knowledge map could be used as a visual example of how information is passed from one part of an organization or group to another and is usually a good place to start understanding what types of intellectual assets the organization has at its disposal. Most organizations that have implemented KM applications provide a context and framework for the way knowledge is gained. KM is usually an integral part of continuous quality improvement or total quality management projects. Several consultants offer knowledge-mapping methodologies (Heathfield & Louw, 1999).

Process-Based Knowledge Map

A process-based knowledge map is essentially a map or diagram that visually displays knowledge within the context of a business process. In other words, the map shows how knowledge should be used within the processes and sources of this knowledge. The overview of the business process is prepared before the knowledge and the sources are mapped to this process. Any type of knowledge that drives the process or results from execution of the process can be mapped. This could include tacit knowledge (knowledge that resides in people such as know-how, experience, and intuition) and explicit knowledge (codified knowledge that is found in documents); customer knowledge; knowledge in processes; etc.

Intelligent Agents

In the early days of online information retrieval systems, individuals met with search intermediaries who were trained to use the online systems. The intermediaries were often knowledgeable about the information seeker's area of interest. Today, technology in the form of personal computers and the Internet provides users with the means to access the online databases from their own offices. However, distributed sources of online information, e.g., the World Wide Web (WWW), compound the problem of information searching. Both novice and experienced users still need support with the search process and the integration of information. To address this problem, agents have been developed for information management applications. The goal of intelligent search agents is to allow end-users to search effectively, be it either a single database of bibliographic records or a network of distributed, heterogeneous, hypertext documents. The approaches range from desktop agents specialized for a single user to networks of agents used to collect data from distributed information sources. Intelligent agents use a combination of profiling techniques, search tools, and recognition algorithms to "push" information to the decision maker on a regular basis. Because intelligent agents use a standard Web analogy, users can quickly set up "net casts" of internal information to automatically receive knowledge bases when they become available. For example, a physician can request that lab results be forwarded to his or her individual dashboard as soon as the lab has completed the procedure. One cautionary note: using push technologies can result in an information flood if filters are not configured to reduce unwanted or unnecessary data (Strawser, 2000).

Web Browsers

Web browsers such as Microsoft's Internet Explorer are practical because of their cost and relative ease of use, and they have become the preferred presentation layer for accessing knowledge bases. The productivity potential inherent in browsers is similar to that of wireless phones. The freer the knowledge worker is of place, time, medium, and device, the less time is spent on the process of messaging, and the more time is available for results. The less time spent on process, the shorter the knowledge cycle, which can be a significant productivity advantage (Strawser, 2000).

KM Applications

Most new KM applications consist of two major elements (often integrated into one interface): a means for employees looking for specialized knowledge to hook up with other people in the organization with the same knowledge (usually via a web application), in other words, an easy way for employees to tap into tacit knowledge resources (people); and a means for employees looking for specialized knowledge to search relevant documents/data (also usually via a web application), in other words, an easy way for employees to tap into explicit knowledge. KM applications are usually designed to support a particular information set in an organization, such as length-of-stay margin management, physician profitability, or accounts receivable recovery. Some consulting firms offer a base set of templates as a core application, focusing on desktop applications, whereas others have developed information sets that provide encyclopedic knowledge of a particular healthcare segment, such as physician issues (Strawser, 2000).

Workflow Applications

In the software world, and in particular in the imaging subset of that world, the need arose to send a particular (bit-mapped) document to a particular workstation or users on a network. This simple routing and distribution was called "workflow." It was quickly learned that these work items could be "tracked," which allowed the accumulation of data about not only where an item had been but also on what happened to it along the way. Who worked on it, how long it was there, the status it had leaving a point, and where it went next was starting to look much more like the manufacturing model. If data was also accumulated about the workers in this operation (how many did they complete, to what status, how long did it take for each) the model was complete. Workflow applications, such as Lotus Notes or Outlook 2000, also play an important role in a KM implementation. For example, a KM solution based on Office 2000 could serve as a nurse triage application, integrating automatic call distribution and transaction-processing systems, in concert with a knowledge base of typical responses to patient questions and symptoms (Strawser, 2000). These described technologies and tools can be exploited to create data banks, knowledge banks and a KM software as shown in *Table 1*, which will be a great help for providing better solutions for healthcare.

These data banks can be readily accessed through a network (Internet or intranet) if the user has at his/her disposal a workstation, software, and passwords that give access to these networks and their different sites. The system should be made flexible so that it can be adaptable in time and space, and can be customized: each physician can, when he/she wishes, consult one of the

Table 1. Data banks, knowledge banks and KM software

Data banks	(Involving text and images as well as figures) which act as medicine's memory to be used for research (clinical, pharmaceutical, epidemiological, etc.)
Knowledge banks	(Bibliographies, sites for exchanges among professionals, etc.), which make it possible to have access at any moment to the state of the art and can help in making medical decision (diagnosis or treatment)
Software	To help with diagnosis and prescribing which does not replace the physician but acts to extend their knowledge

banks or receive specific information on the fields which he/she has chosen in advance. Access need not be limited to physicians alone, but can also be made available to health professionals and the public, and the latter could adapt behavior to help prevent the onset of illness. Of course, these data banks have to be organized, updated continuously and meet users' expectations by allowing them to ask questions and to engage in discussions among themselves, so that it is always possible to evaluate the quality of the information consulted (Fitchett, 1998). The knowledge bank can make it possible to consult experts located elsewhere and to transmit images and other elements of a patient's file to a colleague to obtain an opinion, a practice currently known as "tele-medicine" (Hansen & Nohria, 1999). Software can help to offer software diagnostic and prescription aids to extend this knowledge (Confessore, 1997; Wyatt, 2000; Timpson, 1998).

CONCLUSIONS

Until recently, most of the knowledge experience and learning about KM had been accessible to only a few practitioners. However, during the past three years an explosion of interest, research, and applications in KM has occurred. There is some concern among the practitioners that KM might suffer a fate similar to business reengineering, artificial intelligence, and total quality management. That is, interest in the discipline must last long enough to iron out the bugs while simultaneously delivering significant business value. The irony is that just when the discipline works well, potential users often have lost interest in the fad, point to the inevitable early failures, and thus miss out on the real benefits. Unless the very ambitious and interesting KM initiatives in healthcare evolve differently, even if they work in a technical sense, they will not work in the economic sense and the healthcare system will continue to be what it is today: an immense apparatus for reimbursing healthcare costs. However, there will still be a health/social services network and a medical profession that is familiar with the tools and power of the Internet. The experience will not be totally negative, even if it is likely to reach goals that are different from those set at the outset. Although considerable progress has been achieved in KM across a broad front, much work remains to fully deliver the business value that KM promises. Ultimately, in order to realize the enormous potential value from KM, organizations must motivate and enable creating, organizing and sharing knowledge.

This chapter has served to highlight the need for healthcare in general to embrace the strategies, protocols, tools, techniques and technologies of knowledge management in order to contend with key challenges pertaining to access, quality and value of healthcare delivery. The following chapters in the book will highlight specific areas within the healthcare industry and key issues regarding knowledge management that are pertinent. This will serve to develop a detailed understating of the essential requirements for creating knowledge-based healthcare organizations.

REFERENCES

Allee, V. (1997). *The Knowledge Evolution: Building Organizational Intelligence*. London: Butterworth-Heinemann.

Allee, V. (1999). Knowledge Management: Moving the Care Model from a "Snapshot" to a "Story." *Health Forum Journal*. Retrieved from the World Wide Web: http://www.healthforum.com/hfpubs/asp/ArticleDisplay.asp?PubID=7&ArticleID=1305&Keyword=knowledge.

Antrobus, S. (1997). Developing the nurse as a knowledge worker in health - learning the artistry of practice. *Journal of Advanced Nursing*, 25 (4), 823-829.

Atkins, J., Cooper, L., Hughes, E., Kingston, P. & Mayne, A. (2001, January). The healing power of KM. *Knowledge Management*, 1.

Beckman, T. J. (1999). The current state of knowledge management. In J. Liebowitz (Ed.), *Knowledge Management Handbook*. New York: CRC Press.

Brailer, D. J. (1999, January). Management of knowledge in the modern health care delivery system. *Joint Commission Journal on Quality Improvement*, 25 (1), 1999.

Broadbent, M. (1998). The phenomenon of knowledge management: What does it mean to the information profession? *Information Outlook*, 2 (5), 23-36.

Burca, S. (2000). The learning healthcare organization. *International Journal for Quality in Healthcare*, 12 (6), 457-458.

Chase, R. L. (1998). Knowledge navigators. *Information Outlook*, 2(9), 18.

Choo, C. W. (1998). *The Knowing Organization: How Organizations Use Information to Construct Meaning, Create Knowledge, and Make Decisions.* New York: Oxford University Press.

Confessore, S. J. (1997). Building a learning organization: communities of practice, self-directed learning and continuing medical education. *Journal of Continuing Education in the Health Professions*, 17(1), 5-11.

Conklin (1996). Capturing organizational memory. In R. M. Baecker (Ed.), *Groupware and Computer-Supported Cooperative Work* (pp. 561-565). San Mateo, CA: Morgan Kaufmann.

Conner, M. (2001). Developing network-based services in the NHS. *International Journal of Health Care Quality Assurance including Leadership in Health Services*, 14(6-7), 237-244.

Davenport, T. H. & Prusak, L. (1997). *Information Ecology: Mastering the Information and Knowledge Environment*. New York: Oxford University Press.

Davenport, T. H. & Prusak, L. (1998). *Working Knowledge: How Organizations Manage What They Know*. Boston, MA: Harvard Business School Press.

Dean, B. (2002). Learning from prescribing errors. *Quality and Safety in Healthcare*, 11(3), 258-260.

Einbinder, J. S., Klein, D. A. & Safran, C. S. (1997). Making effective referrals: A knowledge management approach. *Proceedings AMIA Annual Fall Synposium*, (pp. 330-334).

Eisenberg, H. (2002). Transforming Hospitals into Knowledge Management Centres [presentation]. Retrieved from the World Wide Web: http://www.syntrek.com/kmppt/.

Elliot, S. (2000). Sharing knowledge & best practices: The hows and whys of tapping your organization's hidden reservoirs of knowledge. *Health Forum Journal*. Retrieved from the World Wide Web: http://www.healthforum.com/hfpubs/asp/ArticleDisplay.asp?PubID=7&ArticleID=1343&Keyword=knowledge.

Fitchett, J. (1998). Managing your organization's key asset: Knowledge. *Health Forum Journal*. Retrieved from the World Wide Web: http://www.healthforum.com/hfpubs/asp/ArticleDisplay.asp?PubID=7&ArticleID=12869& Keyword=knowledge.

Gokce, C. (2002, July/August). Knowledge management in the NHS. *Managing Information*, 9 (6).

Halpern, S., Perry, S. C. & Narayan, S. (2001, January/February). Developing clinical practice environments supporting the knowledge work of nurses. *Computers in Nursing*, 19 (1), 17-23.

Hansen, M. T. & Nohria, N. (1999). Thomas Tierney. What's your strategy for managing knowledge? *Harvard Business Review,* 77(2), 106-116.

Heathfield, H. & Louw, G. (1999, June). New challenges for clinical informatics: Knowledge management tools. *Health Informatics Journal*, 5(2), 67-73.

Herbig, B., Bussing A. & Ewert, T. (2001). The role of tacit knowledge in the work context of nursing. *Journal of Advanced Nursing*, 34 (5), 687-695.

Jackson, J. R. (2000). The urgent call for knowledge management in medicine. *The Physician Executive*, 26 (1), 28-31.

Jadad, A. R., Haynes, B. Hunt, D. & Browman, G. P. (2000). The Internet and evidence-based decision-making: A needed synergy for efficient knowledge management in health care. *CMAJ*, 162 (3), 362-365.

Johnson, D. E. L. (1998, July). Knowledge management is new competitive edge. *Health Care Strategic Management,* 16 (7), 2-3.

Joyce, L. (2000). Translating knowledge into good practice. *Professional Nurse,* 16 (3), 960-963.

Keeling, C. & Lambert, S. (2000). Knowledge management in the NHS: Positioning the healthcare librarian at the knowledge intersection. *Health Library Review*, 17 (3), 136-143.

Kenner, C. (2002). *Knowledge Management and Nursing Education*. Retrieved from the World Wide Web: http://eknowledgecenter.com/articles/1011/1011.htm.

Krogh, G., Ichijo, K. & Nonaka, I. (2000). *Enabling Knowledge Creation: How to Unlock the Mystery of Tacit Knowledge and Release the Power of Innovation*. New York: Oxford University Press.

Leonard, D. (1998). *Wellsprings of Knowledge: Building and Sustaining the Sources of Innovation*. Boston, MA: Harvard Business School Press.

Matheson, N. W. (1995). Things to come: Postmodern digital knowledge management and medical informatics. *Journal of the American Medical Informatics Association,* 2 (2), 73-8.

Morrissey, J. (1998). Principles of knowledge management. *Modern Healthcare,* 28 (7), 42.

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organizational Science*, 5, 14-37.

Nonaka, I. & Nishiguchi, T. (2001). *Knowledge Emergence*. Oxford: Oxford University Press.

Nutkis, D. S. (1997, November). Webcasting: Knowledge management coming to health care providers. *Surgical Services Management,* 3 (11), 18-21.

Nutley, S. M. & Davies, H. T. O. (2001). Developing organizational learning in the NHS. *Medical Education*, 35, 35-42.

O'Conner, G. T. et al. (1996). A regional intervention to improve the hospital mortality associated with coronary artery bypass graft surgery. *Journal of the American Medical Association*, 6 (4), 55-73.

Odden, J. R. (1998). Developing web-based knowledge management systems for healthcare call centers. *Journal of Healthcare Information Management*, 12 (2), 87-95.

OECD. (2000). *Knowledge Management in the Learning Society, OECD*. Retrieved from the World Wide Web: www.brint.com/km.

Open Clinical. (2000). *The Medical Knowledge Crisis and Its Solution Though Knowledge Management [draft white paper].* Retrieved October 25, 2000 from the World Wide Web: http://www.openclinical.org/docs/whitepaper.pdf.

Oxbrow, N. & Abell, A. (1998). Putting knowledge to work: What skills and competencies are required? In *Knowledge Management: A New*

Competitive Asset. Washington, DC: SLA State-of-the-Art Institute, 25.

Pfeffer, J. & Sutton, R. I. (2000). *The Knowing-Doing Gap: How Smart Companies Turn Knowledge into Action*.

Sharma, S. & Wickramasinghe, N. (2004). A framework for building a learning organization in the 21st Century. (forthcoming). *International Journal of Innovation and Learning.*

Skyrme, D. J. (1998, March). Knowledge Management: Solutions: The Role of Technology. *ACM SIGGROUP Bulletin*, Special Issue on Knowledge Management at Work.

Skyrme, D. J. (2001). *Capitalizing on Knowledge: From e-Business to k-Business.* Butterworth-Heinemann.

Skyrme, D. J. & Amidon, D. M. (1999). The knowledge agenda. In J. D. Cortada & J. A. Woods (Eds.), *The Knowledge Management Yearbook* (pp. 108-125). Butterworth-Heinemann.

Sorrells, J. J. (1999). The role of the chief nurse executive in the knowledge-intense organization of the future. *Nursing Administration Quarterly*, 23 (3), 17-25.

Sorrells, J. J. & Weaver, D. (1999a, July/August). Knowledge workers and knowledge-intense organizations, Part 1: A promising framework for nursing and healthcare. *Journal of Nursing Administration*, 29 (7/8), 12-18.

Sorrells, J. J. & Weaver, D. (1999b, September). Knowledge workers and knowledge-intense organizations, Part 2: Designing and managing for productivity. *Journal of Nursing Administration*, 29 (9), 19-25.

Sorrells, J. J. & Weaver, D. (1999c). Knowledge workers and Knowledge-intense organizations, Part 3: Implications for preparing healthcare professionals. *Journal of Nursing Administration*, 29 (10), 14-21.

Strawser, C. L. (2000). Building effective knowledge management solutions. *Journal of Healthcare Information Management*, 14 (1), 73-80. Retrieved from the World Wide Web: http://www.himss.org/asp/ContentRedirector.asp?ContentID=948.

Timpson, J. (1998). The NHS as a learning organization: Aspirations beyond the rainbow? *Journal of Nursing Management*, 6 (5), 273-274.

Wickramasinghe, N. (2003). Practising what we preach: Are knowledge management systems in practice really knowledge management systems? *Business Process Management Journal*, 9 (3), 295-316.

Wyatt, J. C. (2000). Clinical questions and information needs. *Journal of the Royal Society of Medicine*, 93 (4), 168-171.

This work was previously published in Creating Knowledge-Based Healthcare Organizations, edited by N. Wickramasinghe, J.N.D. Gupta, and S. Sharma, pp. 1-13, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 1.17
Healthcare Knowledge Management

Kostas Metaxiotis
National Technical University of Athens, Greece

INTRODUCTION

The healthcare environment is changing rapidly, and effective management of the knowledge base in this area is an integral part of delivering high-quality patient care. People all over the world rely on a huge array of organizations for the provision of healthcare, from public-sector monoliths and governmental agencies to privately funded organizations, and consulting and advisory groups. It is a massive industry in which every organization faces a unique combination of operational hurdles. However, what every healthcare system has in common is the high price of failure. Faced with the prospect of failing to prevent suffering and death, the importance of continuously improving efficiency and effectiveness is high on the agenda for the majority of healthcare organizations (Brailer, 1999). Taking also into consideration that the amount of biological and medical information is growing at an exponential rate, it is not consequently surprising that knowledge management (KM) is attracting so much attention from the industry as a whole.

In a competitive environment like the healthcare industry, trying to balance customer expectations and cost requires an ongoing innovation and technological evolution. With the shift of the healthcare industry from a central network to a global network, the challenge is how to effectively manage the sources of information and knowledge in order to innovate and gain competitive advantage. Healthcare enterprises are knowledge-intensive organizations which process massive amounts of data, such as electronic medical records, clinical trial data, hospitals records, administrative reports, and generate knowledge. However, the detailed content of this knowledge repository is to some extent "hidden" to its users, because it is regularly localized or even personal and difficult to share, while the healthcare data are rarely transformed into a strategic decision-support resource (Heathfield & Louw, 1999). KM concepts and tools can provide great support to exploit the huge knowledge and information resources and assist today's healthcare organizations to strengthen healthcare service effectiveness and improve the society they serve.

The key question which remains is the following: *How can we make knowledge management work in healthcare?* The answer is given in the following sections.

The Healthcare Industry: A Brief Overview

The health care industry is one of the largest single industries all over the world and the largest one in the United States. It has increased by over 65% since 1990 and is expected to double by the year 2007.[1] The IT industry is strategically positioned to become a powerful ally to the healthcare industry as it strives to adopt well-managed cost-efficient strategies. Advanced information technologies can give healthcare providers the opportunity to reduce overall healthcare expenses by lowering the costs of completing administrative and clinical transactions. Nevertheless, in comparison to other industry sectors, the healthcare industry has been slow to embrace e-business solutions and other advanced information technologies, as presented in Table 1.

The same study revealed that the healthcare industry spends substantially more on overhead and computer facility maintenance than other industry sectors. In 1997, for instance, the healthcare industry allotted 12% of its budget to maintain existing infrastructure—6% more than the industry norm. The high level of investment in this area by healthcare organizations indicates that many providers operate with the aid of old systems, which require constant repair and maintenance.

At this stage, it is worth emphasizing that the healthcare context differs from other information systems application domains in that it often concerns sensitive and confidential information and leads to critical decisions on people's lives (or quality of life). Thus, stakeholder conflicts have more of an impact than in other areas such as business, trade, and manufacturing. Healthcare is an area with quite intense differences of values, interests, professional backgrounds, and priorities among key stakeholders. Given the complexity of the context, health informatics in general cannot simply focus on technical or information systems aspects alone. It has to take account of their relationship with clinical and managerial processes and practices, as well as deal with multiple stakeholders and organizational cultures and accompanying politics.

Concluding, it should be stressed that healthcare is not only a significant industry in any economy (Folland, Goodman, & Stano, 1997), but also a field that needs effective means to manage data as well as information and knowledge. Man-

Table 1. Percentage of IT implementation in industry (Computer Economics, 1999)

Industry Sector	% in Place
Transportation	57.2
Banking and Finance	52.9
Insurance	48.1
State & Local Government	37.5
Trade Services	36.8
Retail Distribution	35.5
Process Manufacturing	34.9
Discrete Manufacturing	33.3
Wholesale Distribution	33.3
Utilities	26.9
Federal Government	25.0
Healthcare	21.8
Professional Services	21.7

aged care has emerged as an attempt to stem the escalating costs of healthcare (Wickramasinghe & Ginzberg, 2001) and improve the quality of services.

THE BACKGROUND OF KM IN HEALTHCARE

An increasing concern with improving the quality of care in various components of the healthcare system has led to the adoption of quality improvement approaches originally developed for industry. These include *Total Quality Management* (Deming, 1986), an approach that employs process control measures to ensure attainment of defined quality standards, and *Continuous Quality Improvement* (Juran, 1988), a strategy to engage all personnel in an organization in continuously improving quality of products and services. Nowadays, the importance of knowledge management is growing in the information society, and medical domains are not an exception. In Yu-N and Abidi (1999), managing knowledge in the healthcare environment is considered to be very important due to the characteristics of healthcare environments and the KM properties. We should always keep in mind that medical knowledge is complex and doubles in amount every 20 years (Wyatt, 2001).

The healthcare industry is nowadays trying to become a knowledge-based community that is connected to hospitals, clinics, pharmacies, physicians, and customers for sharing knowledge, reducing administrative costs, and improving the quality of care (Antrobus, 1997; Booth, 2001). The success of healthcare depends critically on the collection, analysis, and exchange of clinical, billing, and utilization information or knowledge within and across the organizational boundaries (Bose, 2003).

It is only recently that initiatives to apply KM to the healthcare industry have been undertaken by researchers. Firstly, in the second half of the 1980s, several authors tried to apply artificial intelligence (AI)—with doubtful success—to medicine (Clancey & Shortliffe, 1984; Frenster, 1989; Coiera, Baud, Console, & Cruz, 1994; Coiera, 1996). MYCIN is probably the most widely known of all medical (and not only) expert systems thus far developed (Shortliffe, 1976). And this is despite the fact that it has never been put into actual practice. It was developed at Stanford University solely as a research effort to provide assistance to physicians in the diagnosis and treatment of meningitis and bacteremia infections. PUFF, DXplain, QMR, and Apache III are also some of the most well-known medical expert systems that were developed and put into use (Metaxiotis, Samouilidis, & Psarras, 2000).

De Burca (2000) outlined the conditions necessary to transform a healthcare organization into a learning organization. Fennessy (2001) discussed how knowledge management problems arising in evidence-based practice can be explored using “soft systems methodology” and action research. Pedersen and Larsen (2001) presented a distributed health knowledge management (DKM) model that structures decision support systems (DSSs) based on product state models (PSMs) among a number of interdependent organizational units. The recurrent information for the DSS comes from a network-wide support for PSMs of the participating organizations.

Ryu, Hee Hp, and Han (2003) dealt with the knowledge sharing behavior of physicians in hospitals; their study investigated the factors affecting physicians’ knowledge sharing behavior within a hospital department by employing existing theories, such as the Theory of Reasoned Action and the Theory of Planned Behavior. Torralba-Rodriguez and colleagues (2003) presented an ontological framework for representing and exploiting medical knowledge; they described an approach aimed at building a system able to help medical doctors to follow the evolution of their patients, by integrating the knowledge offered by physicians and the knowledge collected from intelligent alarm systems. Also, Chae, Kim, Tark, Park, and Ho (2003) presented an analysis of healthcare quality indicators using data mining for developing quality improvement strategies.

Reviewing the literature, it is concluded that a KM-based healthcare management system should have the following objectives (Shortliffe, 2000; Booth & Walton, 2000):

- To improve access to information and knowledge at all levels (physicians, hospital administrators and staff, consumers of health services, pharmacies, and health insurance companies) so that efficiencies and cost reductions are realized.
- To transform the diverse members (care recipients, physicians, nurses, therapists, pharmacists, suppliers, etc.) of the healthcare sector into a knowledge network/community of practice.
- To enable evidence-based decision making to improve quality of healthcare.

Table 2 presents important Web sites dedicated to the promotion and application of KM to healthcare.

THE KNOWLEDGE MANAGEMENT PROCESS IN HEALTHCARE

In order to examine whether knowledge management can really succeed in healthcare, we can analyze this proposition in terms of examining the knowledge management process and the likelihood of success for the healthcare organizations in achieving these steps in the process. The KM process consists of four key stages, as shown in Figure 1 (Schwartz, Divitini, & Brasethvik, 2000).

Knowledge identification and capture refer to identifying the critical competencies, types of knowledge, and the right individuals who have the necessary expertise that should be captured. Then, this captured knowledge is shared between individuals, departments, and the like. The knowledge application stage involves applying knowledge—which includes retrieving and using knowledge—in support of decisions, actions, and problem solving, and which ultimately can create new knowledge. As new knowledge is created, it needs to be captured, shared, and applied, and the cycle continues.

Knowledge Identification and Capture in Healthcare

One way to identify the critical knowledge that should be captured and determine the experts in the healthcare organization who have the knowledge on a specific issue (e.g., disease, therapy) is to conduct a knowledge audit. The knowledge audit helps to identify the types of knowledge needed and the appropriate sources (e.g., patient records, medical research literature, medical procedures, drug references) in order to develop a knowledge management strategy for the organization.

On the other hand, the use of intranets is suggested as basic tools for the capture of implicit knowledge. St Helens & Knowsley Health Informatics Service—which covers 320,000 patients—designed and developed an intranet structure with the aim to generate the potential to capture organizational implicit knowledge (Mimnagh, 2002). The real challenge has been to create a health- community- wide intranet that implements directory services, communities of practice, and lessons learnedt in a way which builds on existing activity and looks for the synergistic effect of adding a KM focus to ongoing work.

Vast amounts of medical knowledge reside within text documents, so that the automatic extraction of such knowledge would certainly be beneficial for clinical activities. Valencia-Garcia and colleagues et al. (2004) presented a user-centered approach for the incremental extraction of knowledge from text, which is based on both

Table 2. Important Web sites dedicated to KM in healthcare

Web Site	Description
www.nelh.nhs.uk/knowledge_management.asp	The National Electronic Library for Health has a link dedicated to knowledge management. It describes how to manage explicit knowledge and outlines revolutions in KM in healthcare.
www.who.int	The World Health Organization has launched the Health Academy, which aims to demystify medical and public health practices, and to make the knowledge of health specialists available to all citizens through Web-based technology. The academy will provide the general public with the health information and knowledge required for preventing diseases and following healthier lifestyles.
www.cochrane.org	The Cochrane Collaboration is an international non-profit and independent organization, dedicated to making up-to-date, accurate information about the effects of healthcare readily available worldwide. The major product of the collaboration is the Cochrane Database of Systematic Reviews, which is published quarterly.
www.AfriAfya.org	AfriAfya, African Network for Health Knowledge Management and Communication, is a consortium formed by well-known agencies such as Aga Khan Health Service in Kenya, CARE International, SatelLife HealthNet, PLAN International, and the Ministry of Health in Kenya to harness the power of information and communication technology for community health.
www.hc-sc.gc.ca/iacb-dgiac/km-gs/english/kmhome.htm	The goal of knowledge management at Health Canada is to use the knowledge that resides in the department—in the minds of its staff, in the relationships they have with other organizations, and in their repositories of information—to fulfill their mission: to help the people of Canada maintain and improve their health.
www.ucl.ac.uk/kmc/index.html	The Knowledge Management Centre is part of the School of Public Policy of University College London (UCL). The Knowledge Management Centre's aim is to improve clinical practice, patient outcomes, and health service innovation and efficiency by promoting better health knowledge management by serving as a resource center and making efficient use of its resources internally and across a network of collaborators.

knowledge technologies and natural language processing techniques. The system was successfully used to extract clinical knowledge from texts related to oncology and capture it.

Concluding, a key question is whether people would be willing to give up their competitive edge to have their knowledge captured via online repositories, lessons learned, best practices, and the like. This possible dilemma is especially valid in the healthcare sector.

Knowledge Sharing in Healthcare

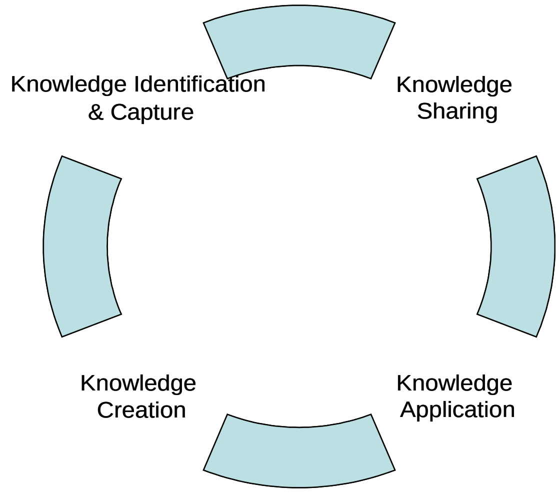
Productive organizations have the ability to create an environment where specialized knowledge, skills, and abilities of all employees are leveraged to achieve advancements in service industry. However, healthcare organizations cannot be considered as a good example of such organizations. A healthcare organization is a collection of professional specialists who contribute to the delivery of patient care, but also often act competitively inside the organization, without being willing to transfer knowledge because of associated status and power within the organization and the society.

Taking also into account that people in general are not likely to share their knowledge unless they think it is valuable and important, it becomes clear why doctors and physicians are not willing to share and transfer their knowledge. In addition, due to minimal interdisciplinary training, the transfer of tacit knowledge which occurs through apprenticeship style work patterns—for example, internships where junior doctors work alongside a senior clinician in surgery or intensive care—remains problematic (Beveren, 2003).

Effective knowledge management requires a "knowledge sharing" culture to be successful. Especially in healthcare, it is crucial that doctors and physicians understand the benefits of knowledge sharing on a number of levels: benefits to the organization, benefits to patients, and benefits to them personally. The more you can clearly demonstrate these benefits, the more people are likely to be open to change. Doctors and physicians need to be recognized and rewarded in a formal way (e.g., promotions, cash awards) to make knowledge sharing a reality in healthcare.

The Wisecare (Workflow Information Systems for European Nursing Care) project—an EC-funded initiative (1997-1999)—has promoted knowledge sharing using the Internet and online communities. Wisecare provided nurses with a vast amount of information and knowledge about clinical practice through both the Wisecare Web site and data collection tool. This information has been specifically selected to meet their clinical needs and meant nurses had access to relevant knowledge extremely quickly.

Figure 1. The knowledge management process cycle



Lesson learned systems can also be an effective knowledge sharing approach to be used in healthcare (Yassin & Antia, 2003).

Knowledge Application in Healthcare

Knowledge application refers to taking the shared knowledge and internalizing it within one's perspective and worldviews. For the healthcare organizations the reality is that technology can only fulfill some of their needs. And how well it fulfills them depends critically on managing the knowledge behind them—content management, assigning knowledge roles, and so forth. Tom Davenport (2002), a prominent author on knowledge management, is often quoted as offering the following rule of thumb: your investment in technology in terms of both cost and effort should stay under one-third of the total knowledge management effort—otherwise you are going wrong somewhere.

Knowledge-enabling technologies which can effectively be applied to healthcare organizations are:

- Groupware
- Intranet
- Collaborative tools (e.g., discussion boards, videoconferencing)
- Portals
- Taxonomies

Abidi (2001) presented the Healthcare Enterprise Memory (HEM) with the functionality to acquire, share, and operationalize the various modalities of healthcare knowledge. Davenport (2002) outlined how Partners Health Care System in Boston implemented an enormously successful expert-intervention KM solution. Case studies from the UK's National Health Service (NHS) and the Department of Health illustrated the drive towards modernization and more effective collaborative working among public-sector healthcare systems (Ark Group, 2002).

Knowledge Creation in Healthcare

In general, knowledge creation may take the form of new products or services, increased innovation, and improved customer relationships. In the healthcare setting, knowledge creation can take place in terms of improved organizational processes and systems in hospitals, advances in medical methods and therapies, better patient relationship management practices, and improved ways of working within the healthcare organization. Given the various constraints and barriers occur in the healthcare sector, it takes longer for a new idea to be implemented in the healthcare setting versus that in the business sector.

A few examples of knowledge creation technologies that can be used in healthcare are:

- **Data Mining:** Tools that analyze data in very large databases, and look for trends and patterns that can be used to improve organizational processes.
- **Information Visualization:** Computer-supported interactive visual representations of abstract data to help improve understanding.

CONCLUSION

Knowledge is a critical tool for health, and knowledge management is the capacity to translate research results (knowledge) into policies and practices that can improve the quality of life and lengthen survival. Managing knowledge in a healthcare organization is like trying to knit with thousands of strands of knotted wool; data is held in a number of locations, managed by a variety of people, and stored in every imaginable format. Perhaps in no other sector does knowledge management have such a high promise.

Delivering healthcare to patients is a very complex endeavor that is highly dependent on information. Healthcare organizations rely on information about the science of care, individual patients, care provided, results of care, as well as its performance to provide, coordinate, and integrate services. The traditional single physician-patient relationship is increasingly being replaced by one in which the patient is managed by a team of health care professionals each specializing in one aspect of care. Hence, the ability to access and use the electronic healthcare record (EHCR) of the patient is fundamental. In addition, the transformation of healthcare data into a strategic decision-support resource is fundamental too.

KM can be approached in numerous ways to serve particular needs and conditions. Successful KM practices typically need to be supported by complementary efforts in different domains. IT-related support activities and infrastructures are very important. They serve vital functions, are complex, costly, and often take time to design and implement. In the case of healthcare, building the infrastructure for a KM practice requires extensive effort due to the peculiarities of the health sector (e.g., legal and ethical issues, complex procedures for provision of healthcare, doctors' behavior, etc.).

Coming back to the original question—*How can we make knowledge management work in healthcare?*—and by examining the knowledge management process, we can see that there are positive and negative points as to whether KM will truly work in the healthcare sector. Some people in healthcare think that KM is a passing fad like Total Quality Management, Business Process Reengineering, and other administration-backed initiatives. It is unfortunate to think in this light, as knowledge sharing should be encouraged so that lessons can be learned. KM solutions can facilitate the transfer of patient medical information, access to new treatment protocols as they emerge, knowledge exchange among experts, and so on.

Future research needs to be devoted to measuring the success of KM in healthcare organizations, showing quantitative benefits, and producing a "Return on Investment" index. Measurement is the least-developed aspect of KM because of the inherent difficulty to measure something that cannot be seen, such as knowledge (Bose, 2004). However, this is a very crucial issue since the future usage of KM is heavily dependent on both the quality of the metrics and whether output generated by this metric management would provide tangible value addition to the healthcare organizations. Integration of KM with e-health is also another direction for further research.

REFERENCES

Antrobus, S. (1997). Developing the nurse as a knowledge worker in health—learning the artistry of practice. *Journal of Advanced Nursing, 25*(4), 823-829.

Ark Group. (2002). Inside knowledge. *Knowledge Management, 6*(4).

Beveren, J. (2003). Does health care for knowledge management? *Journal of Knowledge Management, 7*(1), 90-95.

Booth, A. (2001). Managing knowledge for clinical excellence: Ten building blocks. *Journal of Clinical Excellence, 3*(4), 25-32.

Booth, A., & Walton, G. (2000). *Managing knowledge in health services.* Library Association Publishing.

Bose, R. (2003). Knowledge management-enabled healthcare management systems: capabilities, infrastructure and decision support. *Expert Systems with Applications, 24,* 59-71.

Bose, R. (2004). Knowledge management metrics. *Industrial Management & Data Systems, 104*(6), 457-468.

Brailer, D. (1999). Management of knowledge in the modern health care delivery system. *Joint Commission Journal on Quality Improvement, 25*(1), 1-7.

Chae, Y., Kim, H., Tark, K., Park, H., & Ho, S. (2003). Analysis of healthcare quality indicators using data mining and decision support systems. *Expert Systems with Applications, 24,* 167-172.

Clancey, B.C., & Shortliffe, E.H. (1984). *Medical artificial intelligence: The first decade.* Boston: Addison-Wesley.

Coiera, E. (1996). Artificial intelligence in medicine: The challenges ahead. *Journal of the American Medical Association, 6,* 363-366.

Coiera, E., Baud, R., Console, L., & Cruz, J. (1994). *The role of knowledge based systems in clinical practice.* Holland: IOS Press.

Davenport, T. (2002). Making knowledge work productive and effective. *Knowledge Management, 6*(1).

Davenport, T., & Beck, J. (2002). The strategy and structure of firms in the attention economy. *Ivey Business Journal, 66*(4), 48-54.

De Burca, S. (2000). The learning healthcare organization. *International Journal for Quality in Healthcare, 12*(6), 456-458.

Deming, W. (1986). *Out of the crisis.* Cambridge, MA: MIT Press.

Fennessy, G. (2001). Knowledge management in evidence based healthcare: Issues raised when specialist information services searched for the evidence. *Health Informatics Journal, 7*(1), 1-5.

Folland, S., Goodman, A., & Stano, M. (1997). *The economics of health and health care*. Englewood Cliffs, NJ: Prentice-Hall.

Frenster, J. (1989). *Expert systems and open systems in medical artificial intelligence.* Physicians' Educational Series.

Heathfield, H., & Louw, G. (1999). New challenges for clinical informatics: Knowledge management tools. *Health Informatics Journal*, *5*(2), 67-73.

Juran, J. (1988). *Juran's quality control handbook* (4th ed.). New York: McGraw-Hill.

Metaxiotis, K., Samouilidis, J.-E., & Psarras, J. (2000). Expert systems in medicine: Academic illusion or real power? *Journal of Informatics in Primary Care,* (February), 3-8.

Mimnagh, C. (2002). Towards capturing implicit knowledge: A practical application of intranet development. *Proceedings of the Conference in Advances in Clinical Knowledge Management 5,* London.

Pedersen, M., & Larsen, M. (2001). Distributed knowledge management based on product state models—the case of decision support in healthcare administration. *Decision Support Systems, 31,* 139-158.

Ryu, S., Hee Hp, S., & Han, I. (2003). Knowledge sharing behavior of physicians in hospitals. *Expert Systems with Applications, 25*, 113-122.

Schwartz, D., Divitini, M., & Brasethvik, T. (2000). Knowledge management in the Internet age. In D. Schwartz, M. Divitini, & T. Brasethvik (Eds.), *Internet-based knowledge management and organizational memory* (pp. 1-23). Hershey, PA: Idea Group Publishing.

Shortliffe, E. (2000). *Medical informatics.* New York: Springer-Verlag.

Shortliffe, E.H. (1976). *Computer-based medical consultations: MYCIN.* Elsevier.

Torralba-Rodriguez, F., Fernandez-Breis, J., Garcia, R., Ruiz-Sanchez, J., Martinez-Bejar, R., & Gomez-Rubi, J. (2003). An ontological framework for representing and exploiting medical knowledge. *Expert Systems with Applications, 25,* 211-230.

Valencia-Garcia, R., Ruiz-Sanchez, J., Vivancos-Vicente, P., Fernandez-Breis, J., & Martinez-Bejar, R. (2004). An incremental approach for discovering medical knowledge from texts. *Expert Systems with Applications, 26*(3), 291-299.

Wickramasinghe, N., & Ginzberg, M. (2001). Integrating knowledge workers and the organization: The role of IT. *International Journal of Healthcare Quality Assurance, 14*(6), 245-253.

Wyatt, J. (2001). *Clinical knowledge and practice in the information age: A handbook for health professionals.* Royal Society of Medicine Press.

Yassin, K., & Antia, B. (2003). Quality assurance of the knowledge exchange process: A factor in the success of child health programs in developing countries. *International Journal of Health Care Quality Assurance, 16*(1), 9-20.

Yu-N, C., & Abidi, S. (1999). Evaluating the efficacy of knowledge management towards healthcare enterprise modeling. *Proceedings of the International Joint Conference on Artificial Intelligence.*

KEY TERMS

Collaborative Tools: Electronic tools that support communication and collaboration—people working together; essentially they take the form of networked computer software.

Distributed Knowledge Management Model: The model which combines the interdependence of one partial product state model to others with the idea of knowledge acquisition rather than just the operational exchange relationship.

Evidence-Based Medicine: Healthcare based on best practice which is encoded in the form of clinical guidelines and protocols.

Groupware: Specific software which allows groups of people to share information and to coordinate their activities over a computer network.

Healthcare Enterprise Memory: A KM info-structure which supports the functionality to acquire, share, and operationalize the various modalities of knowledge existent in a healthcare enterprise.

Information Visualization: Computer-supported interactive visual representations of abstract data which help improve understanding.

Taxonomy: A hierarchical structure for organizing a body of knowledge; it gives a framework for understanding and classifying knowledge.

ENDNOTE

1 The Health Care Financing Administration, *National Health Expenditures* (1998).

This work was previously published in Encyclopedia of Knowledge Management, edited by D. Schwartz, pp. 204-210, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.18
Knowledge Management in Hospitals

Kevin C. Desouza
University of Illinois at Chicago, USA

ABSTRACT

The medical field in recent years has been facing increasing pressures for lower cost and increased quality of healthcare. These two pressures are forcing dramatic changes throughout the industry. Managing knowledge in healthcare enterprises is hence crucial for optimal achievement of lowered cost of services with higher quality. The following chapter focuses on developing and fostering a knowledge management process model. We then look at key barriers for healthcare organizations to cross in order to fully manage knowledge.

INTRODUCTION

The healthcare industry is information intensive and recent trends in the industry have shown that this fact is being acknowledged (Morrissey, 1995; Desouza, 2001). For instance, doctors use about two million pieces of information to manage their patients (Pauker, Gorry, Kassirer & Schwartz, 1976; Smith, 1996). About a third of doctor's time is spent recording and combining information and a third of the costs of a healthcare provider are spent on personal and professional communication (Hersch & Lunin, 1995). There are new scientific findings and discoveries taking place every day. It is estimated that medical knowledge increases fourfold during a professional's lifetime (Heathfield & Louw, 1999), which inevitably means that one cannot practice high quality medicine without constantly updating his or her knowledge. The pressures toward specialization in healthcare are also strong. Unfortunately, the result is that clinicians know more and more about less and less. Hence it becomes difficult for them to manage the many patients whose conditions require skills that cross traditional specialties. To add to this, doctors also face greater demands from their patients. With the recent advances of e-health portals, patients actively search for medical knowledge. Such consumers are increasingly interested in treatment quality issues and are also more aware of the different treatment choices and care possibilities.

Managing knowledge in healthcare enterprises is hence crucial for optimal achievement of lowered cost of services with higher quality. The fact that the medical sector makes up a large proportion of a country's budget and gross domestic product (GDP), any improvements to help lower cost will lead to significant benefits. For instance, in 1998 the healthcare expenditure in the US was $1.160 billion, which represented 13.6% of the GDP (Sheng, 2000). In this chapter, we look at the knowledge management process and its intricacies in healthcare enterprises.

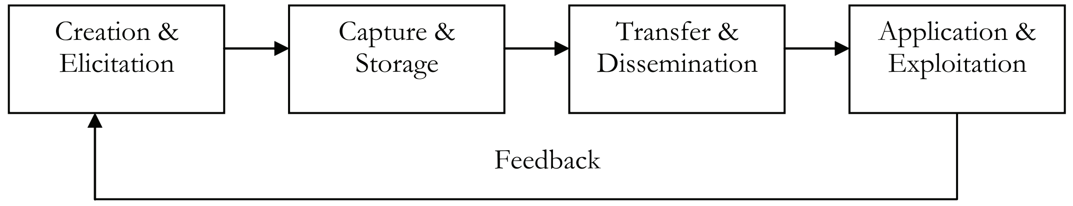
KNOWLEDGE MANAGEMENT PROCESS

Knowledge management from a process perspective is concerned with the creation, dissemination, and utilization of knowledge in the organization. Therefore, a well-structured process needs to be in place to manage knowledge successfully. The process can be divided into the following steps: beginning with knowledge creation or elicitation, followed by its capture or storage, then transfer or dissemination, and lastly its exploitation. We now elaborate on the various stages of the process:

Creation and Elicitation

Knowledge needs to be created and solicited from sources in order to serve as inputs to the knowledge management process. For the first scenario where knowledge has to be created, we begin at the root — data. Relevant data needs to be gathered from various sources such as transaction, sales, billing, and collection systems. Once relevant data is gathered, it needs to be processed to generate meaningful information. Transaction processing systems take care of this task in most businesses today. Just like data, information from various sources needs to be gathered. An important consideration to be aware of is that information can come from external sources in addition to internal sources. Government and industry publications, market surveys, laws and regulations, etc., all make up the external sources. Information once gathered needs to be integrated. Once all necessary information is at our disposal, we can begin analyzing it for patterns, associations, and trends — generating knowledge. The task of knowledge creation can be delegated to dedicated personnel, such as marketing or financial analysts. An alternative would be to employ artificial intelligence-based computing techniques for the task such as genetic algorithms, artificial neural networks, and intelligent agents (Desouza, 2002a). Data mining and knowledge discovery in data bases (KDD) relate to the process of extracting valid, previously unknown and potentially useful patterns and information from raw data in large data bases. The analogy of mining suggests the sifting through of large amounts of low grade ore (data) to find something valuable. It is a multistep, iterative inductive process. It includes such tasks as: problem analysis, data extraction, data preparation and cleaning, data reduction, rule development, output analysis and review. Because

Figure 1. Knowledge management process



data mining involves retrospective analyses of data, experimental design is outside the scope of data mining. Generally, data mining and KDD are treated as synonyms and refer to the whole process in moving from data to knowledge. The objective of data mining is to extract valuable information from data with the ultimate objective of knowledge discovery.

Knowledge also resides in the minds of employees in the form of know-how. Much of the knowledge residing with employees is in tacit form. To enable for sharing across the organization, this knowledge needs to be transferred to explicit format. According to Nonaka and Takeuchi (1995), for tacit knowledge to be made explicit there is heavy reliance on figurative language and symbolism. An inviting organizational atmosphere is central for knowledge solicitation. Individuals must be willing to share their know-how with colleagues without fear of personal value loss and low job security. Knowledge management is about sharing. Employees are more likely to communicate freely in an informal atmosphere with peers than when mandated by management. Desouza (2003b) studied knowledge exchange in game rooms of a high-technology company and found significant project-based knowledge exchanged.

Capture and Storage

To enable distribution and storage, knowledge gathered must be codified in a machine-readable format. Codification of knowledge calls for transfer of explicit knowledge in the form of paper reports or manuals into electronic documents, and tacit knowledge into explicit form first and then to electronic representations. These documents need to have search capabilities to enable ease of knowledge retrieval. The codification strategy is based on the idea that the knowledge can be codified, stored and reused. This means that the knowledge is extracted from the person who developed it, is made independent of that person and reused for various purposes. This approach allows many people to search for and retrieve knowledge without having to contact the person who originally developed it. Codification of knowledge, while being beneficial for distribution purposes, does have associated costs. For instance, it is easier to transfer strategic know-how outside the organization for scrupulous purposes. It is also expensive to codify knowledge and create repositories. We may also witness information overload in which large directories of codified knowledge may never be used due to the overwhelming nature of the information. Codified knowledge has to be gathered from various sources and be made centrally available to all organizational members. Use of centralized repositories facilitates easy and quick retrieval of knowledge, eliminates duplication of efforts at the departmental or organizational levels and hence saves cost. Data warehouses are being employed extensively for storing organizational knowledge (Desouza, 2002a).

Transfer and Dissemination

One of the biggest barriers to organizational knowledge usage is a blocked channel between knowledge provider and seeker. Blockages arise from causes such as temporal location or the lack of incentives for knowledge sharing. Ruggles' (1998) study of 431 US and European companies shows that "creating networks of knowledge workers" and "mapping internal knowledge" are the two top missions for effective knowledge management.

Proper access and retrieval mechanisms need to be in place to facilitate easy access to knowledge repositories. Today almost all knowledge repositories are being web-enabled to provide for the widest dissemination via the Internet or intranets. Group Support Systems are also being employed to facilitate knowledge sharing, with two of the most prominent being IBM's Lotus Notes and Microsoft's Exchange. Security of data sources and user friendliness are important considerations that need to be considered while providing access to knowledge repositories. Use of passwords and secure servers is important when providing access to knowledge of a sensitive nature. Access mechanisms also need to be user-friendly in order to encourage use of knowledge repositories.

Exchange of explicit knowledge is relatively easy via electronic communities. However, exchange of tacit knowledge is easier when we have a shared context, co-location, and common language (verbal or non-verbal cues), as it enables high levels of understanding among organizational members (Brown & Duguid, 1991). Nonaka and Takeuchi (1995) identify the processes of socialization and externalization as means of transferring tacit knowledge. Socialization keeps the knowledge tacit during the transfer, whereas externalization changes the tacit knowledge into more explicit knowledge. Examples of socialization include on-the-job training and apprenticeships. Externalization includes the use of metaphors and analogies to trigger dialogue among individuals. Some of the knowledge is, however, lost in the transfer. To foster such knowledge sharing, organizations should allow for video and desktop conferencing as viable alternatives for knowledge dissemination.

Exploitation and Application

Employee usage of knowledge repositories for purposes of organizational performance is a key measure of the system's success. Knowledge will never turn into innovation unless people learn from it and learn to apply it. The enhanced ability to collect and process data or to communicate electronically does not — on its own — necessarily lead to improved human communication or action (Walsham, 2001). Recently the notion of communities of practice to foster knowledge sharing and exploitation has received widespread attention. Brown and Duguid (1991) argued that a key task for organizations is thus to detect and support existing or emergent communities. Much of knowledge exploitation and application happens in team settings and workgroups in organizations, hence support must be provided. Davis and Botkin (1994) summarize the six traits of a knowledge-based business as follows:

1. The more *they* (customers) use knowledge-based offerings, the smarter *they* get.
2. The more *you* use knowledge-based offerings, the smarter *you* get.
3. Knowledge-based products and services adjust to changing circumstances.
4. Knowledge-based businesses can customize their offerings.
5. Knowledge-based products and services have relatively short life cycles.
6. Knowledge-based businesses react to customers in real time.

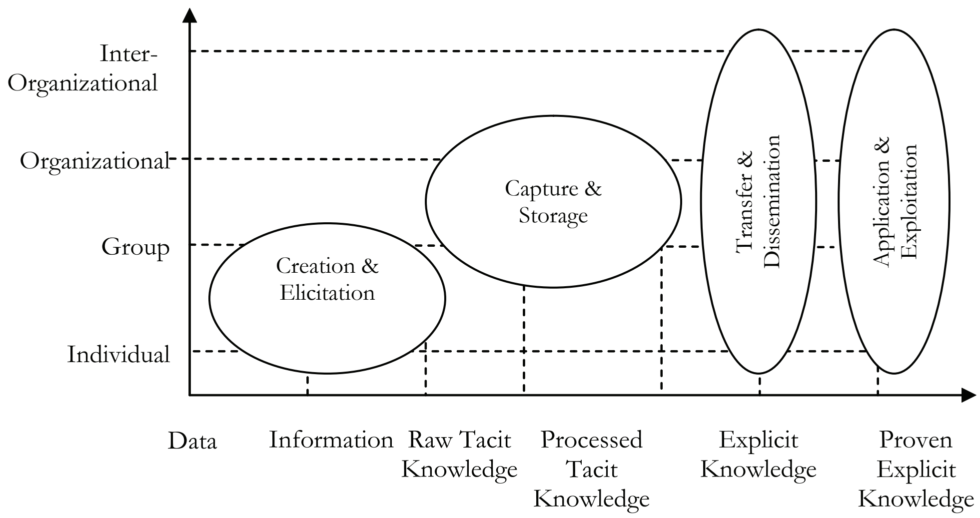
KNOWLEDGE MANAGEMENT IN HOSPITALS

We now apply the generic discussion of knowledge, knowledge management, and the process in the context of healthcare enterprises. For purposes of this chapter, we focus our attention on hospitals, although much of the discussion can be applied to other healthcare enterprises, such as pharmaceutical companies, insurance providers, etc.

Knowledge in Hospitals

In healthcare, we have the presence of both explicit and tacit forms of knowledge. Explicit knowledge is available in medical journals, research reports, and industry publications. Explicit knowledge can be classified under: internal and external. Internal are those that are relevant to the practice of medicine, such as medical journals and research reports. External are legal, governmental, and other publications that do not directly affect patient treatment methodology but govern general medical practices. Three dimensions in health information outlined by Sorthon, Braithewaite and

Figure 2. Staged look at knowledge management



Lorenzi (1997) include management information, professional information, and patient information. Overlap and commonalties are identified, but fundamental differences exist in the types of information required for each dimension, the way the information is used, and the way standards are maintained. The achievement of a comprehensive and integrated data structure that can serve the multiple needs of each of these three dimensions is the ultimate goal in most healthcare information system development. Tacit knowledge is found in the minds of highly specialized practitioners, such as neurosurgeons or cardiac arrest specialists. Much of tacit knowledge resides in the minds of individuals. Seldom does efficient knowledge sharing take place. One exception to this is where practitioners exchange know-how at industry or academic conferences. This, however, happens on an all-too-infrequent basis.

Knowledge Management in Hospitals

In the following section we step through the various stages of the knowledge management process.

Knowledge Creation and Elicitation

Creation and elicitation of knowledge can be handled in one of two modes: controlled or free form. In the controlled scenario, we can have an individual department responsible for overseeing knowledge gathering from the various functional areas. This department can be in lieu of the current medical records department in most hospitals, which are responsible for centrally storing patient information. We can also have a variation in which control is divested to each department. In this method, each department will be responsible for coordinating knowledge-sharing efforts from their constituents. For instance, a person in the pharmaceutical department will be responsible for gathering all knowledge on drugs administered to patients. In the second approach, i.e., free form, each individual is responsible for contributing to the organization's knowledge resource. A strong locus of control is absent. Individuals as end users of organizational resources share equal burden to contribute into the asset. A variation of the free-form strategy can be one in which each group, rather than individuals, are responsible for knowledge creation and sharing. An example would be a group of neurosurgeons that research new knowledge on surgical practices. Each of the above-mentioned strategies has associated pros and cons. For instance, with a controlled strategy, we need to have dedicated individuals responsible for knowledge creation and elicitation from group members. In the free-form strategy, while we do not have the overhead of a dedicated person, we lose a structured knowledge-creation process. Choosing a given strategy is a function of the hospital's resources. Along with soliciting internal knowledge, a hospital should also acquire relevant knowledge from external entities such as government, regulatory bodies, and research organizations. Gathering of external knowledge is crucial for hospitals due to the high nature of external pressures and their involvement in day-to-day operations.

As portrayed in *Figure 2*, knowledge creation and elicitation takes place at the individual and group level. Much of the elements gathered at this stage might be raw data, which after processing, becomes information. Information is then applied on experiences, learned associations, and training of employees to generate knowledge. This knowledge remains in tacit form until it is called upon for sharing with peers. Tacit knowledge stored with employees is raw to a large degree, as it is has not been checked for quality or validated against standards.

The other option in the healthcare industry is to generate knowledge through discovery. Data mining and other statistical techniques can be used to sift through large sets of data, and discover hidden patterns, trends, and associations. For instance, in the medical domain all resources are not only very expensive but also scarce. Optimal planning and usage of them is, hence, not a luxury but a requirement. To illustrate this let us take the case of simple blood units. Annually well over 12 million units of blood are transferred to patients (Clare et al., 1995; Goodnough, Soegiarso, Birkmeyer & Welch, 1993) with a cost-per-unit ranging from $48 to $110 with a national average of $78 (Sheng, 2000; Forbes et al., 1991; Hasley, Lave & Kapoor, 1994). Ordering of excess units of blood for operations is the primary cause of waste and corresponding increases in transfusion costs (Jaffray, King & Gillon, 1991). Blood ordered that is not used takes it out of supply for at least 48 hours (Murphy et al., 1995). Even though blood can be returned, it needs to be tested and routed which costs on average $33. In recent years, supply of blood has been decreasing in recent years due to an aging population and increased complexity of screening procedures (Welch, Meehan & Goodnough, 1992). Given these circumstances, any improvements in blood-need prediction can realize significant benefits. Data mining techniques such as artificial neural networks have been employed to sift through large medical databases and generate predictive models which can better forecast blood transfusion requirements. Kraft, Desouza and Androwich (2002a, 2002b, 2003a, 2003b) examine the discovery of knowledge for patient length-of-stay prediction in the Veterans Administration Hospitals. Once such knowledge is generated, it can be made available to external sources.

Knowledge Capture and Storage

Once gathered, knowledge needs to be captured and stored to allow for dissemination and transfer. Two strategies are common for capture and storage: codification and personalization. The codification strategy is based on the idea that knowledge can be codified, stored and reused. This means that the knowledge is extracted from the person who developed it, is made independent of that person and reused for various purposes. This approach allows many people to search for and retrieve knowledge without having to contact the person who originally developed it (Hansen et al., 1999). Organizations that apply the personalization strategy focus on dialogue between individuals, not knowledge objects in a database. To make the personalization strategies work, organizations invest heavily in building networks or communities of people. Knowledge is shared not only face-to-face, but also by e-mail, over the phone and via videoconferences. In the medical domain, the codification strategy is often emphasized, because clinical knowledge is fundamentally the same from doctor to doctor. For instance, the treatment of an ankle sprain is the same in London as in New York or Tokyo. Hence it is easy for clinical knowledge to be captured via codification and to be reused throughout the organization.

Knowledge capture has been one of the most cumbersome tasks for hospitals. Until rather recently much of the patient knowledge was stored in the form of paper reports and charts. Moreover, the knowledge was dispersed throughout the hospital without any order or structure. Knowledge was also recorded in different formats, which made summarization and storage difficult.

Recently we have seen advancements in the technology of Electronic Medical Records (EMRs). EMRs are an attempt to translate information from paper records into a computerized format. Research is also underway for EMRs to include online imagery and video feeds. At the present time they contain patients' histories, family histories, risk factors, vital signs, test results, etc. (Committee on Maintaining Privacy and Security in Healthcare Applications of the National Information Infrastructure, 1997). EMRs offer several advantages over paper-based records, such as ease of capture and storage. Once in electronic format, the documents seldom need to be put through additional transformations prior to their storage.

Tacit knowledge also needs to be captured and stored at this stage. This takes place in multiple stages. First, individuals must share their tacit know-how with members of a group. During this period, discussions and dialogue take place in which members of a group validate raw tacit knowledge and new perspectives are sought. Once validated, tacit knowledge is then made explicit through capture in electronic documents such as reports, meeting minutes, etc., and is then stored in the knowledge repositories. Use of data warehouses is common for knowledge storage. Most data warehouses do have web-enabled front-ends to allow for optimal access.

Knowledge Transfer and Dissemination

Knowledge in the hospital once stored centrally needs to be made available for access by the various organizational members. In this manner knowledge assets are leveraged via diffusion throughout the organization. One of the biggest considerations here is security. Only authorized personnel should be able to view authorized knowledge. Techniques such as the use of multiple levels of passwords and other security mechanisms are common. However, organizational security measures also need to be in place. Once the authorized users get hold of the knowledge, care should be taken while using such knowledge, to avoid unscrupulous practices. Moreover, employees need to be encouraged to follow basic security practices, such as changing passwords on a frequent basis, destroying sensitive information once used, etc. Ensuring security is a multi-step process. First, the individual attempting to access information needs to be authenticated. This can be handled through use of passwords, pins, etc. Once authenticated, proper access controls need to be in place. These ensure that a user views only information for which he or she has permission. Moreover, physical security should also be ensured for computer equipment such as servers and printers to prevent unauthorized access and theft.

Disseminating healthcare information and knowledge to members outside the organization also needs to be handled with care. Primarily physicians, clinics, and hospitals that provide optimal care to the patients use health information. Secondary users include insurance companies, managed care providers, pharmaceutical companies, marketing firms, academic researchers, etc. Currently no universal standard is in place to govern exchange of healthcare knowledge among industry partners. Hence, free flow of healthcare knowledge can be assumed to a large degree. From a security perspective, encryption technologies should be used while exchanging knowledge over digital networks. Various forms are available such as public and private key encryptions, digital certificates, virtual private networks, etc. These ensure that only the desired recipient has access to the knowledge. An important consideration while exchanging knowledge with external entities is to ensure that patient identifying information is removed or disguised. One common mechanism is to scramble sensitive information such as social security numbers, last and first names. Another consideration is to ensure proper use by partners. Knowledge transferred outside the organization (i.e., the hospital) can be considered to be of highest quality as it is validated multiple times prior to transmittal.

Medical data needs to be readily accessible and should be used instantaneously (Schaff, Wasserman, Englebrecht & Scholz, 1987). The importance of knowledge management cannot be stressed enough. One aspect of medical knowledge is that different people need different views of the data. Let us take the case of a nurse, for instance. He or she may not be concerned with the intricacies of the patient's condition, while the surgeon performing the operation will. A pharmacist may only need to know the history of medicine usage and any allergic reactions, in comparison to a radiologist who cares about which area needs to be x-rayed. Hence, the knowledge management system must be flexible to provide different data views to the various users. The use of intelligent agents can play an important role here through customization of user views. Each specialist can deploy customized intelligent agents to go into the knowledge repository and pull out information that concerns them, thus avoiding the information overload syndrome. This will help the various specialists attend to problems more efficiently instead of being drowned with a lot of unnecessary data. Another dimension of knowledge management is the burden put on specialists. A neurosurgeon is paid twice as much, if not more, than a nurse. Hence, we should utilize their skills carefully to get the most productivity. Expert systems play a crucial role here in codifying expertise/knowledge. When a patient comes for treatment, preliminary test and diagnosis should be handled at the front level. Expert systems help by providing a consultation environment whereby nurses and other support staff can diagnose illness and handle basic care, instead of involving senior-level doctors and specialists. This allows for the patients that need the care of experts to receive it and also improves employee morale through less stress.

Knowledge Application and Exploitation

The last stage, and the most important, is the application and exploitation of knowledge resources. Only when knowledge stored is used for clinical decision-making does the asset provide value. As illustrated in *Figure 3*, knowledge application and exploitation should take place at all levels from the individual to inter-organizational efforts. We draw a distinction here between application and exploitation. Applications are predefined routines for which knowledge needs are well defined and can be programmed. For instance, basic diagnosis when a patient first enters the hospitals, these efforts include calculation on blood pressure, pulse rates, etc. Knowledge needed at this level is well defined and to a large extent is repetitive. On the other hand, exploitation calls for using knowledge resources on an ad-hoc basis for random decision-making scenarios. For instance, if a hospital wants to devise an optimal nurse scheduling plan, use of current scheduling routines, plus knowledge on each individual's skill sets can be exploited for devising the optimal schedule. Decisions like these, once handled, seldom repeat themselves on a frequent basis.

With knowledge management being made easy and effective, quality of service can only increase. A nurse, when performing preliminary tests on a patient, can provide them with better information on health issues through consultation with an expert system. Primary care doctors normally refer patients for hospital care. Some of the primary care doctors may work for the hospital (Network) and the rest are independent of the hospital (Out-of-Network). Today there are a lot of inefficiencies associated with referring patients to hospitals. If a patient is referred, he or she has to contact the hospital personnel who then first take in all patient information and then schedule an appointment. The normal wait time can be any where from one to four weeks depending on seriousness. With the Internet revolution today, all patients, doctors, and hospitals can improve the process tremendously through the deployment of dedicated intelligent agents. Each doctor can be provided with a log-on and password to the hospital's web site. Upon entry to the web site, the doctor can use search agents to browse through appointment schedules, availability of medical resources, etc. These agents can then schedule appointments directly and electronically receive all documentation needed. Hospitals within a certain location can set up independent networks monitored by agents whereby exchange of medical knowledge and resources can take place. Patients can use search agents to browse through hospital web sites, request prescriptions, learn about medical treatments, view frequently asked questions, etc. Intelligent agents can also be trained to learn patient characteristics. Once this takes place, they can be deployed to monitor various medical web sites and send relevant information to the patient in the form of e-mails. Expert systems can be deployed to help the user navigate through the various knowledge bases through recommendations. If a user chooses the main category of "common cold," the expert system can ask for symptoms, suggest medications, etc. Patients can then use these notifications to improve the quality of their health. Intelligent agents also help in improving quality of service through providing only relevant decision-making information. Personnel can then act quickly and reduce time lags. An added benefit of a successful knowledge management system is less burden and stress on personnel. Hospitals are characterized for being highly stressful and always "on pins and needles" when it comes to employees. Through artificial intelligence, much of the routine details can be automated. This reduces the burden on personnel. Also, specialists and highly valued personnel can concentrate efforts on selected matters, the rest can be handled by junior level staff and intelligent systems. This makes for a more welcoming atmosphere.

IMPENDING BARRIERS TO KNOWLEDGE MANAGEMENT

The medical field has to overcome a few hurdles in order to realize the potential benefits of open connectivity for knowledge sharing among the partners of the supply chain and internal personnel such as doctors, surgeons, nurses, etc. We now highlight three of the most prominent issues:

Unified Medical Vocabulary

The first barrier is the development of a unified medical vocabulary. Without a unified vocabulary, knowledge sharing becomes close to impossible. There is diversity of vocabulary used by medical professionals, which is a problem for information retrieval (Lindberg, Humphreys & McCray, 1993). There are also differences in terminology used by various biomedical specialties, researchers, academics, and variations in information accessing systems, etc. (Houston, 2000). To make matters more complex, expertise among users of medical information also varies significantly. A researcher in neuroscience may use precise terminology from the field, whereas a general practitioner may not. Medical information also must be classified differently based on tasks. Researchers may need information summarized according to categories, while a practitioner or doctor may need patient-specific details that are accurate (Forman, 1995).

To help bridge some of the gap in terminology, we have two main medical thesauri in use. Medical Subject Headings (MeSH) and Unified Medical Language System (UMLS) are meta-thesauri developed by the National Library of Medicine (NLM) (Desouza, 2001). UMLS was developed in 1986 and has four main components: meta-thesaurus, specialist lexicon, semantic net, and information sources map. The meta-thesaurus is the largest and most complex component incorporating 589,000 names for 235,000 concepts from more than 30 vocabularies, thesauri, etc. (Lindberg et al., 1993). Approaches to organizing terms include human indexing and keyword search, statistical and semantic approaches. Human indexing is ineffective as different experts use varying concepts to classify documents, plus it is time-consuming for large volumes of data. The probability of two people using the same term to classify a document is less than 20% (Furnas, Landauer, Gomez & Dumais, 1987). Also different users use different terms when searching for documents. Artificial Intelligence-based techniques are making headway in the field of information retrieval. Houston et al. (2000) used a Hopfield network to help in designing retrieval mechanism for the CANCERLIT study. The issue of standardization of terminology continues to be a great debate. The Healthcare Financing Association (HCFA) is adopting some Electronic Data Interchange (EDI) standards to bring conformity to data (Moynihan, 1996). We can expect more standards to be released in the next few years to enable sharing of data.

Security and Privacy Concerns

With sharing of data comes the inherent risk of manipulation and security issues. Security of patients' data and preventing it from entering the wrong hands are big concerns in the field (Pretzer, 1996). Strict controls need to be put in place before open connectivity can take place. Patients' data are truly personal and any manipulation or unauthorized dissemination has grave consequences. Sharing of patient-identifiable data between members of the healthcare supply chain members is receiving serious scrutiny currently. Government and other regulatory bodies will need to set up proper laws to help administer data transmission and security (Palmer et al., 1986). The recent Health Insurance Portability and Accountability (HIPPA) Act can be seen as one of many governmental interventions into the healthcare industry for the protection of consumer privacy rights. Enterprises will have to go to the basics of ethics and operate carefully.

Organizational Culture

In every organization we can see the application of the 20/80 rule. Knowledge providers make up 20% of the workforce, as they possess experiences and insights that are beneficial to the organization. The remaining 80% are consumers of this knowledge (Desouza, 2002a). The providers are often reluctant to share and transfer knowledge as they fear doing so will make them less powerful or less valuable to the organization (Desouza, 2003a, 2003b). Between departments we also find knowledge barriers, in which one group may not want to share insights collected with the other. To help alleviate some of these issues, management should strive to provide incentives and rewards for knowledge-sharing practices. A highly successful approach is to tie a portion of one's compensation to group and company performance, thus motivating employees to share knowledge to ensure better overall company performance. Additionally, Foucault (1977) noted the inseparability of knowledge and power, in the sense that what we know affects how influential we are, and vice versa our status affects whether what we know is considered important. Hence, to alleviate this concern, an enterprise-wide initiative should be carried out making any knowledge repository accessible to all employees without regard to which department or group generated it.

A key dimension of organizational culture is leadership. A study conducted by Andersen and APQC revealed that one crucial reason why organizations are unable to effectively leverage knowledge is because of a lack of commitment of top leadership to sharing organizational knowledge or there are too few role models who exhibit the desired behavior (Hiebeler, 1996). Studies have shown that knowledge management responsibilities normally fall with middle managers, as they have to prove its worth to top-level executives. This is a good and bad thing. It is a good thing because normally middle-level managers act as liaisons between employees and top-level management, hence they are best suited to lead the revolution due to their experience with both frontline, as well as higher-level authorities. On the other hand it is negative, as top-level management does not consider it important to devote higher-level personnel for the task. This is changing, however. Some large companies are beginning to create the position of chief knowledge officer, which in time will become a necessity for all organizations. A successful knowledge officer must have a broad understanding of the company's operation and be able to energize the organization to embrace the knowledge revolution (Desouza & Raider, 2003). Some of the responsibilities must include setting up knowledge management strategies and tactics, gaining senior management support, fostering organizational learning, and hiring required personnel.

It is quite conceivable that healthcare enterprises will start creating the positions of chief knowledge officers and knowledge champions. Top management involvement and support for knowledge management initiatives cannot be underestimated. This is of pivotal importance in hospitals, as their key competitive asset is medical knowledge.

CONCLUSIONS

Some researchers and practitioners have expressed concern about knowledge management being a mere fad (Desouza, 2003b). To deliver promised values, knowledge management must address strategic issues and provide for competitive advantages in enterprises. McDermott (1999) noted that companies soon find that solely relying on the use of information technology to leverage organizational knowledge seldom works. The following chapter has introduced knowledge management and its process. We have also applied it to healthcare enterprises focusing on hospitals. Finally, we justified the knowledge management process by looking through two main strategic frameworks.

Knowledge management initiatives are well underway in most healthcare enterprises, and we can expect the number and significance of such efforts to increase over the next few years. Areas of future and continued research include: automated search and retrieval techniques for healthcare information, intelligent patient monitoring systems, and optimal knowledge representation semantics.

REFERENCES

Brown, J. S. & Duguid, P. (1991). Organizational learning and communities of practice: Towards a unified view of working learning and innovation. *Organization Science*, *2*(1), 40-57.

Clare, M., Sargent, D., Moxley, R. & Forthman, T. (1995). Reducing health care delivery costs using clinical paths: A case study on improving hospital profitability. *Journal of Health Care Finance, 21*(3), 48-58.

Committee on Maintaining Privacy and Security in Health Care Applications of the National Information Infrastructure. (1997). *For the record: Protecting Electronic health information*. Washington, DC: National Academy of Sciences.

Davis, S. & Botkin, J. (1994). The coming of knowledge-based business. *Harvard Business Review*, 165-70.

Desouza, K. C. (2001). Artificial intelligence for health care management. *Proceedings of the First International Conference on Management of Health care and Medical Technology*. Enschede, The Netherlands.

Desouza, K. C. (2002a). *Managing knowledge with artificial intelligence: An introduction with guidelines for non-Specialists.* Westport, CT: Quorum Books.

Desouza, K. C. (2002b). Knowledge management in hospitals: A process oriented view and staged look at managerial issues. *International Journal of Health Care Technology Management, 4*(6).

Desouza, K. C. (2003a). Barriers to effective use of knowledge management systems in software engineering. *Communications of the ACM, 46*(1), 99-101.

Desouza, K. C. (2003b). Facilitating tacit knowledge exchange. *Communications of the ACM, 46*(6), 85-88.

Desouza, K. C. & Raider, J. J. (2003). Cutting corners: CKO & knowledge management, *Business Process Management Journal*. (Forthcoming).

Forbes, J. M. et al. (1991). Blood transfusion costs: A multicenter study. *Transfusion, 31*(4), 318-323.

Forman, P. N. (1995). Information needs of physicians. *Journal of the American Society of Information Sciences, 46*(10), 729-736.

Foucault, M. (1977). Two lectures. In C. Gordon (Ed.), *Power/Knowledge: Selected Interviews and Other Writings.* New York: Pantheon Books.

Furnas, G. W., Landauer, T. K., Gomez, L. M. & Dumais, S. T. (1987). The vocabulary problem in human-system communication. *Communication of the ACM, 30*(11), 964-971.

Goodnough, L. T., Soegiarso, R. W., Birkmeyer, J. D. & Welch, H. G. (1993). Economic impact of inappropriate blood transfusions in coronary artery bypass graft surgery. *The American Journal of Medicine, 94*(5), 509-514.

Hansen, M. T., Nohria, N. & Tierney, T. (1999). What's your strategy for managing knowledge? *Harvard Business Review, 77*(2), 106-116.

Hasley, P. B., Lave, J. R. & Kapoor, W. N. (1994). The necessary and the unnecessary transfusion: A critical review of reported appropriateness rates and criteria for red cell transfusions. *Transfusion, 34*(2), 110-115.

Heathfield, H. & Louw, G. (1999). New challenges for clinical informatics: Knowledge management tools. *Health Informatics Journal, 5*, 67-73.

Hersch W. R. & Lunin L. F. (1995). Perspectives on medical informatics: Information technology in health care — Introduction and overview. *Journal of American Social Infrastructure Science*, 46, 726-727.

Hiebeler, R. (1996). Benchmarking knowledge management. *Strategy and Leadership, 24*(2).

Houston, A. L. et al. (2000). Exploring the use of concept spaces to improve medical information retrieval. *Decision Support Systems, 30*(2), 171-186.

Jaffray, B., King, P. M. & Gillon, J. (1991). Efficiency of blood use and prospects of autologous transfusion in general surgery. *Annals of the Royal College of Surgeons of England, 73*(4), 235-238.

Kraft, M., Desouza, K. C. & Androwich, I. (2002a). Knowledge discovery in clinical databases: issues and process. In M. Khosrow-Pour (Ed.), *Issues & Trends of Information Technology Management in Contemporary Organizations - 2002 Information Resource Management Association Conference* (Vol. 1, pp. 168-171). Hershey, PA: Idea Group Publishing.

Kraft, M., Desouza, K. C. & Androwich, I. (2002b). Length of stay predictions for spinal cord injury patients using neural networks and nursing diagnoses data: methodological issues, results and applicability. *Proceedings of the Second International Conference on Management of Health Care and Medical Technology*, Chicago, Illinois, July 28-30, 2002.

Kraft, M., Desouza, K. C. & Androwich, I. (2003a). Data mining in health care information systems: Case study of a veterans administration spinal cord injury population. *Proceedings of the Thirty-Sixth Hawaii International Conference on System Sciences (HICSS-36)*, Los Alamos, CA: IEEE Press, Big Island, Hawaii, January 6 - 9, 2003.

Kraft, M., Desouza, K. C. & Androwich, I. (2003b). Health care information: From administrative to practice databases, In G. Grant (Ed.), *ERP & Data Warehousing: Issues and Challenges*. Hershey, PA: Idea Group Publications.

Lindberg, D. A. B., Humphreys, B. L. & McCray, A. T. (1993). The unified medical language system. *Methods of Information in Medicine*, *32*, 281-291.

McDermott, R. (1999). Why information technology inspired but cannot deliver knowledge management. *California Management Review, 41*(4), 103-117.

Morrissey, J. (1995). Managed care steers information systems. *Modern Health Care*, *25*(8).

Moynihan, J. J. (1996). A milestone for paperless claims processing. *Health Care Financial Management, 50*(1), 68-69.

Murphy, W. G. et al. (1995). Blood use for surgical patients: A study of Scottish hospitals transfusion practices. *Journal of Royal College of Surgeons of Edinburgh*, *40*(1), 10-13.

Nonaka, I. & Takeuchi, H. (1995). *The Knowledge-creating Company*. New York: Oxford University Press.

Palmer, R. H., Kane, J. G., Churchill, W. H., Goldman, L. & Komaroff, A. L. (1986). Cost and quality in the use of blood bank services for normal deliveries, cesarean sections, and hysterectomies. *Journal of the American Medical Association*, *256*(2), 219-223.

Pauker, S. D., Gorry, G., Kassirer, J. & Schwartz, W. (1976). Towards the simulation of clinical cognition: Taking a present illness by computer. *American Journal on Medicine*, *65*, 981-996.

Pretzer, M. (1996). Why should you have a health lawyer's convention? *Medical Economics*, *73*(16), 160-168.

Ruggles, R. (1998). The state of the notion: Knowledge management in practice. *California Management Review*, *40*(3), 80-89.

Schaff, R., Wasserman, G., Englebrecht, R. & Scholz, W. (1987). Medical treatment assistance with an interactive drug information system. *Medical Expert Systems Using Personal Computers*, 45-51.

Sheng, O. R. (2000). Decision support for health care in a new information age. *Decision Support Systems, 30*(2), 101-103.

Smith, R. (1996). What clinical information do doctors need? *British Medical Journal, 313,* 1062-1068.

Sorthon, F., Braithwaite, J. & Lorenzi, N. (n.d.). Strategic constraints in health informatics: Are expectations realistic? *International Journal of Health Planning and Management*, *12*, 3-13.

Walsham, G. (2001). *Making a World of Difference: IT in a Global Context*. Chichester, UK: Wiley.

Welch, H. G., Meehan, K. R. & Goodnough, L. T. (1992). Prudent strategies for elective red blood cell transfusions. *Annals of Internal Medicine*, *116*(5), 393-402.

ENDNOTE

[1] A previous version of this article was published in: Desouza, K. C. (2002). Knowledge management in hospitals: A process oriented view and staged look at managerial issues. *International Journal of Healthcare Technology and Management,* 4(6). I would like thank Yukika Awazu for her assistance with this research project.

This work was previously published in An Creating Knowledge-Based Healthcare Organizations, edited by N. Wickramasinghe, J.N.D. Gupta, and S. Sharma, pp. 14-28, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 1.19
An Overview of Efforts to Bring Clinical Knowledge to the Point of Care

Dean F. Sittig
Medical Informatics Department, Kaiser Permanente Northwest, USA
Care Management Institute, Kaiser Permanente, USA
Oregon Health & Sciences University, USA

ABSTRACT

By bringing people the right information in the right format at the right time and place, state of the art clinical information systems with imbedded clinical knowledge can help people make the right clinical decisions. This chapter provides an overview of the efforts to develop systems capable of delivering such information at the point of care. The first section focuses on "library-type" applications that enable a clinician to look-up information in an electronic document. The second section describes a myriad of "real-time clinical decision support systems." These systems generally deliver clinical guidance at the point of care within the clinical information system (CIS). The third section describes several "hybrid" systems, which combine aspects of real-time clinical decision support systems with library-type information. Finally, section four provides a brief look at various attempts to bring clinical knowledge, in the form of computable guidelines, to the point of care.

To be effective, (clinical decision support) tools need to be grounded in the patient's record, must use standard medical vocabularies, should have clear semantics, must facilitate knowledge maintenance and sharing, and need to be sufficiently expressive to explicitly capture the design rational (process and outcome intentions) of the guideline's author, while leaving flexibility at application time to the attending physician and their own preferred methods. (Shahar, 2001)

INTRODUCTION

By bringing people the right information in the right format at the right time and place, informatics helps people make the right clinical decisions. Cumulatively, these better decisions improve health outcomes, such as quality, safety, and the cost-effectiveness of care. This improvement has been a mantra of informatics at least since the landmark article by Matheson and Cooper in 1982. This chapter provides an overview of the efforts over the years to develop systems capable of delivering such information at the point of care. Such an overview should help illustrate both the opportunities and challenges that lie ahead as we struggle to develop the next generation of real-time clinical decision support systems for use at the point of care.

VISION TO ACHIEVE

The ultimate goal is to provide patient-specific, evidence-based, clinical diagnostic and therapeutic guidance to clinicians at the point of care; this guidance should be available within the clinical information system (CIS) that defines their current workflow. In addition, we must have the tools necessary to enable clinicians, without specialized programming knowledge, to enter, review, and maintain all the clinical knowledge required to generate this advice. Finally, we must have the ability to rapidly change the clinical knowledge, test it, and make it available to clinicians without having to wait for a regular CIS updating schedule.

Questions That Must Be Answered Prior to Creating Such Systems

What information or knowledge is required to help the clinician make the right decision to achieve the desired health outcome?

Who will be the information's recipient (e.g., physician, nurse, pharmacist, or even a specific individual such as the patient's primary care physician)?

When in the patient care process is the intervention applied, for example, before, during (which can be broken down into sub-activities such as order entry or progress note creation) or after the patient encounter?

How is the intervention triggered and delivered? For example, does the system or the clinician initiate it? How much patient-specific data (if any) is needed to trigger system-initiated interventions? How much is the intervention output customized to the clinical workflow stage, the clinician, and or the patient? For system-initiated interventions, how can the threshold be set to minimize nuisance alerts? How intrusive should the information be (Krall, 2001)?

Will the clinicians find the information useful (Krall, 2002a, 2002b)? What can we do to minimize the number of false positive alerts?

Where will the clinician be when receiving the intervention, for example, with the patient at the bedside or in the office? What should happen if the information becomes available at some future point when the clinician is no longer with the patient to whom the information pertains?

Which medium will be used to convey the message, for example, e-mail inbox, wireless and/or handheld device, pager, CIS/CPOE screen, or printed pre-visit encounter sheet?

Is there a demonstrable return on investment (ROI) that is due exclusively to the clinical decision support intervention or feature?

BACKGROUND

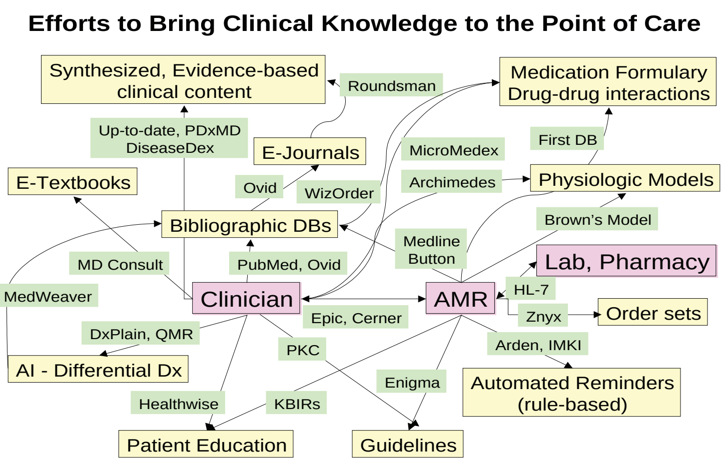
Where Do We Stand?

Numerous attempts have been made to bring various forms of clinical information to the clinician at the point of care. One way to solve this problem is to develop computer applications that stand alone, although often network accessible, and are available to the clinician upon his/her request. Another way is to incorporate the clinical knowledge directly into the clinical information system used by clinicians while giving care. Once it is there, the CIS system can automatically prompt the clinician or the clinician can request help. Recently, a third model for system development has been proposed that enables the clinician to request help from an outside source. Often several patient-specific data items are included in the request. Using such a system, a clinician is still completely in-charge of making the request for information and the information can be automatically configured based on a subset of patient information (Cimino, 1996). In addition, it is now possible to embed medication or procedure codes along with instructions that tell the CIS how to handle this information that allows a clinician to place an order directly into the patient's electronic medical record from the externally available information resource (Tang & Young, 2000).

The following diagram (Figure 1) is an attempt to illustrate both the key types of information or knowledge that investigators have focused on along with their mode of interaction [i.e., directly to the clinician or through the clinical information system (CIS)]. The boxes (yellow) represent the type of knowledge and the labels on the links (green) represent some of the key projects or concepts vendors have focused on this particular form of clinical decision support. The aim of this figure is to highlight the myriad attempts that have been made to develop clinical knowledge management applications and to help everyone understand how different clinical knowledge resources and applications are both related in terms of what they are trying to accomplish and different in the resources they utilize. Although this diagram is fairly complex, it is only a small, imperfect and incomplete representation of the entire clinical decision support landscape.

The following sections briefly review each of the projects or concepts. The first section focuses on "library-type" applications that enable a clinician to look up information in an electronic document. The second section describes a myriad of "real-time clinical decision support systems". These systems generally deliver clinical guidance to clinicians at the point of care within the CIS. The third section describes several "hybrid" systems, which combine aspects of knowledge-based

Figure 1. An overview of the numerous efforts researchers have made to bring pertinent clinical knowledge to clinicians at the point of care



clinical decision support systems with library-type information. Finally, section four looks at various attempts to bring clinical knowledge in the form of computable guidelines to the point of care.

LIBRARY-TYPE APPLICATIONS: FRONT-ENDS TO APPLICATIONS THAT DIRECTLY INTERACT WITH CLINICIANS

Bibliographic Databases (DBs)

Biomedical bibliographic databases contain on the order of millions of records, each representing a unique scientific journal article that has been published. Each record typically contains the title of the article, the authors, their affiliation(s), and the abstract of the article. In addition, several metadata tags may have been applied by human indexers to improve the likelihood that each document will be retrieved when and only when it is relevant to the user's query.

For the last 30 years, the National Library of Medicine has maintained the MedLine database, the most common bibliographic database used in clinical medicine. During that period, various attempts have been made to develop easy to use and reliable interfaces to this vast resource including Grateful Med (Cahan, 1989) and COACH (Kingsland, 1993). Currently, PubMed is the most widely used of all these interfaces. PubMed relies on a sophisticated free-text query processor to map freetext user queries to MeSH terms, when appropriate, and returns a highly relevant set of documents.

In addition to PubMed, several commercial vendors have created proprietary interfaces to the MedLine database in an attempt to improve either the recall or precision of the user's queries. For example, Ovid has developed an interesting MeSH mapper and query expander that has gathered outstanding reviews from highly trained librarians. Knowledge Finder has developed a fuzzy mapping algorithm that has also generated some good reviews. Unfortunately, none of these systems consistently enables clinicians to retrieve more than half of all the relevant articles on any particular topic (Hersh, 1998). In addition to these variations on a search interface, several projects have used automated differential diagnosis generators, such as DxPlain, as an interface to the bibliographic DBs. Finally, the Science Citation Index uses the reference list at the end of every scientific article published to generate linked lists of references. Such a scheme can also identify particularly noteworthy articles since these articles are referenced many more times. Interestingly, the Google search engine uses this same concept to generate its index of relevant web sites. It does this by keeping track of the number and quality of referring web sites rather than references at the end of the article (Brin, 1998).

Other Clinical Reference Materials

In addition to access to the bibliographic literature, clinicians could use other clinical reference information. For example, various systems exist to provide clinicians with access to the complete text and figures contained in textbooks (e.g., MD Consult) or journals (e.g., Ovid) in electronic form. In addition, several commercial companies have begun developing synthesized, evidence-based clinical summaries of the diagnosis and treatment of common clinical conditions [e.g., Up-to-Date, CliniAnswers, DiseaseDex, PDxMD (see http://www.informatics-review.com/KnowMan/Examples.html for other examples)].

REAL-TIME CLINICAL DECISION SUPPORT SYSTEMS

Real-time clinical decision support systems (RCDSS) are fundamentally different from "library" applications in that they interact with the clinician through the CIS. Rather than relying on the user to interpret the text and then make a decision, RCDSS combine the patient's clinical data with the clinical knowledge to help the clinician reach a decision. At least six distinct methods can provide this type of RCDSS:

1. Medication formularies and drug-drug interaction checkers provide routine checks of all orders entered into the EMR. They can recommend alternative medications based on the formulary and check for potential drug-drug interactions automatically following entry of a medication order assuming they have access to the patient's list of current medications. FirstDataBank, Micromedex, and Multum are examples of commercial providers of drug-drug interaction checking databases.
2. Automated rule-based reminders are used most commonly to remind clinicians to perform routine health maintenance procedures. The Arden syntax is currently the only "standard" means of encoding these rule-based systems (Arden, 2003). The Institute for Medical Knowledge Interchange (IMKI), another recent entry into this field (IMKI, 2003), recently scaled back their operations while they seek additional funding and support.
3. Condition-specific order sets and charting templates help clinicians remember and facilitate the entry of all related orders or answers to questions asked of a patient with a particular condition. Zynx, a recently formed company, is developing these ready-to-use, condition-specific order sets. In addition, most of the commercially available CISs either offer ready-to-use order sets or provide tools to help clinicians construct these order sets (Franklin, Sittig, Schmiz, Spurr, Thomas, O'Connell, & Teich, 1998).
4. Integrated clinical guidelines allow a clinician to compare a specific patient's clinical data against a guideline and automatically receive a recommendation. The most advanced example of this type of application is Enigma's PREDICT application. In an early attempt to develop this sort of functionality, Epic Systems created their "Active Guidelines" application. A major distinguishing feature between the Active Guidelines (AG) effort and that of Enigma is that the AG user is still required to search for, read, interpret, and then decide which portion of the large, text-based guideline document applies to a specific patient and then select the appropriate recommendation. Enigma's PREDICT automatically does all of this work and simply presents the clinician with the appropriate suggestions. Another recent entry into this market space is Theradoc, a small company located in Salt Lake City, Utah (www.theradoc.com).
5. Complex, physiologic models incorporate our best current understanding of human physiology along with evidence-based results of randomized clinical trials to generate patient- or population-specific predictions of future events. Archimedes, a complex model of diabetes and the entire healthcare system, is currently configured to allow specially trained clinicians to access this resource (Schlessinger, 2002). Jonathan Brown, from Kaiser Permanente's Center for Health Research in Portland, has recently begun working on a method to automatically link therapeutic suggestions generated by his Global Diabetes Model to a specific patient's clinical data from an EMR or clinical data warehouse (Brown, 2000).
6. Artificial Intelligence (AI) systems have been developed to help clinicians complete their differential diagnosis list. To date these systems have not had wide-spread acceptance by clinicians, even though they have been shown to be at least as good and often better than un-aided clinicians (Bankowitz, 1989; Berner, 1994). DxPlain, QMR,

and Weed's Problem Knowledge Couplers (Weed, 1986) are the best examples of these types of systems.

HYBRID APPLICATIONS

In addition to each of these fairly straightforward attempts to link clinical knowledge to the point of patient care, several companies have attempted to combine various clinical knowledge resources and applications to generate even more useful applications. MedWeaver (Detmer, 1997) was an early attempt to combine DxPlain, an AI-based differential diagnosis (DDx) system with a bibliographic database. The DDx in conjunction with the Unified medical Language system (UMLS) act as an automatic query expander and presents the user with information related to all the known items on the patient's differential diagnosis.

Geissbuhler and Miller (1998) developed WizOrder, a system to automatically generate potential drug-drug interaction reminders when clinicians order medication through the system. These reminders are based on the co-occurrence of clinical term frequencies[1] that are a part of the UMLS's Metathesaurus database. Rennels, Shortliffe, Stockdale, and Miller (1987) developed Roundsman, a prototype system, to automatically extract evidence-based recommendations from on-line structured journal articles.

Cimino (1996) developed the concept of the "InfoButton" which allows a clinician to automatically generate queries to a variety of clinical information resources to answer a set of generic concept-specific questions such as, "What's the treatment for X?", where X can be any clinical condition. Recently, the HL-7 Clinical Decision Support Task Force has begun efforts to develop a standard for such queries.

COMPUTABLE GUIDELINES

Current State of the Art

Various research and commercial efforts are underway in an attempt to create systems to achieve the vision outlined above. The distinguishing feature of computable guideline systems is that they are more than a simple application of "if-then-else" rules to generate alerts and reminders at the point-of-care. All of the systems described below are attempting to go one step further with multi-step clinical algorithms at the point-of-care. The more complex algorithms attempt to track a patient's treatments and progress over time, while the alerts and reminders serve only to look at what is best at each point in time. Therefore, I did not include the efforts of the Institute for Medical Knowledge Interchange (IMKI) or the Arden group to create libraries of rules and systems to facilitate their incorporation in various clinical information systems.

The following sections overview several attempts and describe which parts of the vision they either have or are currently focusing on. I also point out the strengths and weaknesses of each approach.

ACADEMIC RESEARCH EFFORTS

GLEE is a system for execution of guidelines encoded in the Guideline Interchange Format v3 (GLIF3) (Wang, 2002). GLEE provides an internal event-driven execution model that can be hooked up with the clinical event monitor of a local clinical information system environment. It was developed at Columbia University and is currently only a research prototype. No tools exist for knowledge maintenance other than those proposed by the GLIF3 organization.

GLARE (GuideLine Acquisition, Representation and Execution) is a domain-independent system for the acquisition, representation and execution of clinical guidelines (Terenziani, 2001). GLARE provides expert physicians with an "intelligent" guideline acquisition interface. The interface has different types of checks to provide a consistent guideline: syntactic and semantic tests verify whether or not a guideline is well-formed. GLARE technology has been successfully tested in different clinical domains (bladder cancer, reflux esophagitis, and heart failure), at the Laboratorio di Informatica Clinica, Azienda Ospedaliera S. Giovanni Battista, Torino, Italy. Currently, this application is not available for use by people outside of the research team. See: www.openclinical.org/gmm_glare.html#aimj01

COMMERCIAL VENDORS

Therapy Edge (http://www.therapyedge.com/) is a condition-specific, stand-alone, web-based, EMR with extensive clinical decision support for HIV patients. It is a stand-alone tool, that is, it has no integration with the existing CIS; it is only for HIV patients; and has no tools for knowledge based maintenance.

PRODIGY is a guideline-based decision support system in use by a large number of General Practitioners in the United Kingdom (Purves, 1999). The evaluation of PRODIGY 3 is currently underway and Phase 4 is being planned. PRODIGY I and PRODIGY II were implemented as extensions to proprietary UK electronic patient record systems. The PRODIGY system includes a proprietary guideline model, which in PRODIGY II was used to implement guidelines for the management of acute diseases. See: http://www.prodigy.nhs.uk/ClinicalGuidance/ReleasedGuidance/GuidanceBrowser.asp for a demo.

PREDICT is a decision support system for delivering evidence-based care directly to clinical practitioners desktops or handhelds. PREDICT integrates with existing clinical information systems using standard internet communication protocols, and allows clinicians to create new disease modules and manage their own guideline content. See: http://www.enigma.co.nz/framed_index.cfm?fuseaction=ourknowledge_products for more info.

eTG complete is an easy-to-use, HTML-based product, available for use on stand alone or networked PC or Mac computers. It covers over 2000 clinical topics and will be updated at regular intervals (three to four times per year). They do not even mention the possibility of integrating their work with any commercially available EMR products. See: http://www.tg.com.au/complete/tgc.htm for an online demonstration. This is just one of many examples of stand-alone, proprietary format, internet-based, guideline presentation systems. In my opinion, none of these are significantly better than what the USA's National Guideline Clearinghouse (www.guideline.gov) or the New Zealand Guidelines Group (www.nzgg.org.nz) are doing by simply creating a large database of clinical guidelines that are arranged by clinical condition.

AREZZO is a decision support technology (based on PROforma), for building and running clinical applications. AREZZO applications are designed to provide patient-specific advice, guiding the user through data collection, clinical actions and decision making. Applications can be quickly modeled and tested, using the AREZZO Composer, and instantly deployed on the Internet. AREZZO is capable of supporting the development of multi-disciplinary care pathways (customized for local circumstances); its safety-critical and task-based features support the rigor needed for clinical protocols. Currently the system is limited to supporting decisions on pain control in cancer, but this will be expanded in the future to include pain control in much more general terms (for example, arthritis, chronic pain, etc.) See: http://www.infermed.com/ds_arnocs.htm for more information.

SUMMARY AND CONCLUSIONS

Library-type applications that provide users with easy access to the latest medical knowledge in the form of journal articles or textbooks will not be going away in the near future. Even with our increasing understanding of how to build intelligent computer systems that can recognize the current patient context and suggest potentially relevant clinical information resources, the users must still have a means of getting directly to the knowledge resources so they can look up additional information. Therefore, work based on the Medline Button, as described by Cimino (1996) must continue. In fact, an effort is now underway within the HL-7 Clinical Decision Support Technical Committee to develop a standard interface that will facilitate further development of these types of applications.

The use and utility of real-time clinical decision support systems (RCDSS) will continue to grow exponentially. As more and more pressure is placed on clinicians to practice evidenced-based medicine and do it as inexpensively as possible, these RCDSSs will proliferate. A recent publication by the Health Information Management Systems Society (HIMSS) entitled, the "Clinical Decision Support Implementers' Workbook" is now freely available on the internet (Osheroff, 2004). This workbook and others should lead to increased use of various clinical decision support systems. Although little work is currently underway on the development of new hybrid systems, I believe that these types of systems will soon reappear and provide users with significant additional functionality that cannot be achieved using any other means. Finally, as healthcare delivery systems choose and implement commercially available clinical information systems, clinicians and researchers alike could gain valuable knowledge by beginning work on various clinical knowledge management projects with one or more of the vendors or research groups listed above. At the present time, this field has no clear-cut winner. In fact, no system is beyond the advanced research prototype stage. As such, there will be considerable changes in whichever solution is ultimately successful. I believe it is in everyone's best interest to be a part of the effort in defining and experimenting with these types of systems. Only then will we truly understand what is really useful and how we should proceed.

ACKNOWLEDGMENTS

I would like to thank Jerry Osheroff, M.D. for his comments on an early draft of this paper. A preliminary version of this chapter was presented at a health informatics conference in Auckland, New Zealand (2003).

REFERENCES

Arden Syntax for Medical Logic Systems. (2003). Home Page *http://cslxinfmtcs.csmc.edu/hl7/arden/*

Bankowitz, R.A., McNeil, M.A., Challinor, S.M., Parker, R.C., Kapoor, W.N. & Miller, R.A. (1989). A computer-assisted medical diagnostic consultation service. Implementation and prospective evaluation of a prototype. *Ann International Med.*, *110*(10), 824-32.

Berner, E.S., Webster, G.D., Shugerman, A.A., Jackson, J.R., Algina, J., Baker, A.L., Ball, E.V., Cobbs, C.G., Dennis, V.W., Frenkel, E.P., et al. (n.d.). Performance of four computer-based diagnostic systems. *New England Journal of Medicine*, *330*(25), 1792-6.

Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual {Web} search engine. *Computer Networks and ISDN Systems*, *30*(1-7), 107-117.

Brown, J.B., Russell, A., Chan, W., Pedula, K., & Aickin, M. (2000). The global diabetes model:

User friendly version 3.0. *Diabetes Res Clin Pract.*, *50*(Suppl 3), S15-46.

Cahan, M.A. (1989). GRATEFUL MED: A tool for studying searching behavior. *Med Ref Serv Q.*, *8*(4), 61-79

Cimino, J.J. (1996). Linking patient information systems to bibliographic resources. *Methods Inf Med.*, *35*(2), 122-6.

Detmer, W.M., Barnett, G.O., & Hersh, W.R. (1997). MedWeaver: Integrating decision support, literature searching, and Web exploration using the UMLS Metathesaurus. *Proceedings of AMIA Annual Fall Symposium* (pp. 490-494).

Franklin, M.J., Sittig, D.F., Schmiz, J.L., Spurr, C.D., Thomas, D., O'Connell, E.M. & Teich, J.M. (1998). Modifiable templates facilitate customization of physician order entry. *Proceedings of AMIA Symposium* (pp. 315-319).

Geissbuhler, A. & Miller, R.A. (1998). Clinical application of the UMLS in a computerized order entry and decision-support system. *Proceedings of AMIA Symposium* (pp. 320-324).

Hersh, W.R. & Hickam, D.H. (1998). How well do physicians use electronic information retrieval systems? A framework for investigation and systematic review. *JAMA*, *280*(15), 1347-1352.

Institute for Medical Knowledge Implementation (IMKI) (2003). Online *http://www. imki.org/*

Kingsland, L.C., Harbourt, A.M., Syed, E.J. & Schuyler, P.L. (1993). Coach: Applying UMLS knowledge sources in an expert searcher environment. *Bull Med Libr Assoc.*, *81*(2), 178-183.

Krall, M.A. (2002). Clinicians' assessments of outpatient electronic medical record alert and reminder usability and usefulness requirements: A qualitative study. Master's thesis. Online *http://www.ohsu.edu/biccinformatics/people/ms/theses/2002/krall.pdf 1-104*

Krall, M.A. & Sittig, D.F. (2001). Subjective assessment of usefulness and appropriate presentation mode of alerts and reminders in the outpatient setting. *AMIA Proceedings* (pp. 334-338).

Krall, M.A. & Sittig, D.F. (2002). Clinicians' assessments of outpatient electronic medical record alert and reminder usability and usefulness requirements. *AMIA Proceedings* (pp. 400-405).

Matheson, N.W. & Cooper, J.A. (1982). Academic information in the academic health sciences center: Roles for the library in information management. *Journal of Med. Educ.*, *57*(10 Pt 2), 1-93.

Osheroff, J.A., Pifer, E.A., Sittig, D.F., Jenders, R.A. & Teich, J.M. (2004). Clinical decision support implementers' workbook: Health information management and systems society. Available online *http://www.himss.org/CDSworkbook*

Purves, I.N., Sugden, B., Booth, N. & Sowerby, M. (1999). The PRODIGY project: The interactive development of the release one model. *Proceedings AMIA Symposium* (pp. 359-363).

Rennels, G.D., Shortliffe, E.H., Stockdale, F.E. & Miller, P.L. (1987). A computational model of reasoning from the clinical literature. *Comput. Methods Programs Biomed.*, *24*(2), 139-149.

Schlessinger, L. & Eddy, D.M. (2002). Archimedes: A new model for simulating health care systems: The mathematical formulation. *Journal of Biomed Inform.*, *35*(1), 37-50.

Shahar, Y. (2001). Automated support to clinical guidelines and care plans: The intention-oriented view. Online *http://www.openclinical.org/docs/int/briefingpapers/shahar.pdf*

Tang, P.C. & Young, C.Y. (2000). Free full text active guidelines: Integrating Web-based guidelines with computer-based patient records. *Proceedings of the AMIA Symposium* (pp. 843-847).

Terenziani, P., Molino, G. & Torchio, M. (2001). A modular approach for representing and executing clinical guidelines. *Artificial Intelligence Med., 23*(3), 249-276.

Wang, D. & Shortliffe, E.H. (2002). GLEE: A model-driven execution system for computer-based implementation of clinical practice guidelines. *Proceedings of the AMIA Symposium* (pp. 855-859).

Weed, L.L. (1986). Knowledge coupling, medical education and patient care. *Crit Rev Med Inform.*, *1*(1), 55-79.

ENDNOTE

1. The UMLS contains the co-occurrence relationships for the number of times concepts have co-occurred as key topics within the same articles, as evidenced by the Medical Subject Headings assigned to those articles in the MEDLINE database. Co-occurrence relationships have been also computed for different ICD-9-CM diagnosis codes assigned to the same patients as reflected in a discharge summary database. In contrast to the relationships asserted within source vocabularies, the statistical relationships in the Metathesaurus can connect very different concepts, such as diseases and drugs. There are specific Metathesaurus files for the co-occurrence relationships (MRCOC. RRF and MRCOC in ORF).

This work was previously published in Clinical Knowledge Management: Opportunities and Challenges, edited by B. Bali, pp. 285-296, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 1.20
Issues in Clinical Knowledge Management:
Revisiting Healthcare Management

Rajeev K. Bali
Coventry University, UK

Ashish Dwivedi
The University of Hull, UK

Raouf Naguib
Coventry University, UK

ABSTRACT

The objective of this chapter is to examine some of the key issues surrounding the incorporation of the Knowledge Management (KM) paradigm in healthcare. We discuss whether it would it be beneficial for healthcare organizations to adopt the KM paradigm so as to facilitate effective decision-making in the context of healthcare delivery. Alternative healthcare management concepts with respect to their ability in providing a solution to the above-mentioned issue are reviewed. This chapter concludes that the KM paradigm can transform the healthcare sector.

INTRODUCTION

In today's information age, data has become a major asset for healthcare institutions. Recent innovations in Information Technology (IT) have transformed the way that healthcare organizations function. Applications of concepts such as Data Warehousing and Data Mining have exponentially increased the amount of information to which a healthcare organization has access, thus creating the problem of "information explosion". This problem has been further accentuated by the advent of new disciplines such as Bioinformatics and Genetic Engineering, both of which hold very promising solutions which may significantly change the face of the entire healthcare process from diagnosis to delivery (Dwivedi, Bali, James, Naguib, & Johnston, 2002b).

Until the early 1980s, IT solutions for healthcare used to focus on such concepts as data warehousing. The emphasis was on storage of data in an electronic medium, the prime objective of which was to allow exploitation of this data at a later point in time. As such, most of the IT applications in healthcare were built to provide support for retrospective information retrieval needs and, in some cases, to analyze the decisions undertaken. This has changed healthcare institutions' perspectives towards the concept of utility of clinical data. Clinical data that was traditionally used in a supportive capacity for historical purposes has today become an opportunity that allows healthcare stakeholders to tackle problems before they arise.

Healthcare Management Concepts

Healthcare managers are being forced to examine costs associated with healthcare and are under increasing pressure to discover approaches that would help carry out activities better, faster and cheaper (Davis & Klein, 2000; Latamore, 1999). Workflow and associated Internet technologies are being seen as an instrument to cut administrative expenses. Specifically designed IT implementations such as workflow tools are being used to automate the electronic paper flow in a managed care operation, thereby cutting administrative expenses (Latamore, 1999).

One of the most challenging issues in healthcare relates to the transformation of raw clinical data into contextually relevant information. Advances in IT and telecommunications have made it possible for healthcare institutions to face the challenge of transforming large amounts of medical data into relevant clinical information (Dwivedi, Bali, James, & Naguib, 2001b). This can be achieved by integrating information using workflow, context management and collaboration tools, giving healthcare a mechanism for effectively transferring the acquired knowledge, as and when required (Dwivedi, Bali, James, & Naguib, 2002a).

Kennedy (1995, p. 85) quotes Keever (a healthcare management executive) who notes that "Healthcare is the most disjointed industry…in terms of information exchange... Every hospital, doctor, insurer and independent lab has its own set of information, and … no one does a very good job of sharing it." From a management perspective, these new challenges have forced healthcare stakeholders to look at different healthcare management concepts that could alleviate the problem of information explosion. The following are some of the new paradigms and concepts that have caught the attention of healthcare stakeholders.

EVIDENCE BASED MEDICINE (EBM)

EBM is defined as the "conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients" (Cowling, Newman, & Leigh, 1999, p. 149). A typical EBM process starts with an identification of knowledge-gaps in current healthcare treatment processes, followed by a search for the best evidence. This is then succeeded by a process to aid in the selection of appropriate electronic data/information sources and IT applications that focus on clinical competencies in the context of the best evidence generated.

The next step is to carry out a critical appraisal of the best evidence identified by carrying out checks for accuracy and diagnostic validity of the procedure/treatment identified by the best evidence generated. The costs and benefits of alternative procedures (i.e., the current best evidence procedure/treatment being recommended) are then considered. The last step is its application to patients' healthcare which calls for integration of the best evidence with the General Practitioners' (GP) clinical expertise so as to provide best treatment and care (Cowling et al., 1999).

MODEL OF INTEGRATED PATIENT PATHWAYS (MIPP/IPP)

Schmid and Conen (2000) have argued that the model of integrated patient pathways (MIPP/IPP) is a more comprehensive concept for healthcare institutions. As the acronym suggests, IPPs aim to enable better support for healthcare institutions by focusing on the creation of clinical guidelines for commonly accepted diagnostic and therapeutic procedures at a defined level of quality. This would lead to cost-efficient treatment. It could be argued that IPP calls for in-house development of standardized clinical treatment procedures for some pre-defined diagnoses and treatments.

Schmid & Conen (2000) elaborate that IPP aims to ensure that patients receive the right treatment which is based upon best practice guidelines that have sufficient evidence to warrant the label of "best practice" and which have been proven to be clinically adequate. They argue that when a hospital tries to implement IPP, it will automatically go through a circular chain process that calls for identifying sources of best practice, converting them to worldwide implementation practices and then, based upon their performance, converting them to benchmarks. Deliberation on current health reform is centered on two competing objectives: expanding access and containing costs. The challenge is to find an acceptable balance between providing increased access to healthcare services while at the same time conserving healthcare resources.

Pryga and Dyer (1992) have noted that, in the USA, hospitals receive a fixed amount per patient for each Medicare patient admission. As such, they have an objective of providing essential medical services whilst physicians are remunerated on the basis of the clinical service provided. The situation emerges where the physician and healthcare managers can have conflicting goals; such a dilemma is bound to affect formulation of best care practices particularly for preventive care.

CLINICAL GOVERNANCE (CG)

Clinical governance (CG) was first introduced in the UK by way of a National Health Service (NHS) white paper (Firth-Cozens, 1999) and calls for an integrated approach to quality, team development, clinical audit skills, risk management skills, and information systems. A typical CG process can be delineated into a sequential process that calls for (a) the means to disseminate knowledge about relevant evidence from research, (b) best treatments rather than focusing just on recognition of poor treatments, (c) better appreciation of what IT led solutions can do for clinical governance, and (d) knowing what data/information is available so as to provide baselines for best care and treatments.

Melvin, Wright, Harrison, Robinson, Connelly, and Williams (1999) have remarked that the NHS has witnessed the incorporation and development of many approaches that support and promote effective healthcare, but in practice, none of them have been successful. Research by Zairi and Whymark (1999) submits that the problem lies in the lack of proper systems to support the measurement of organizational effectiveness (i.e., clinical) in a healthcare delivery context.

According to Sewell (1997), one of the biggest challenges in having concise summaries of the most effective clinical practices is establishing what is meant by "quality in healthcare" (i.e., a measurement standard for clinical effectiveness). Sewell (1997) elaborates that measurement standards in clinical practice will change from each context and that this is attributed to the linkage between measurement standards and values and the expectations of the individual healthcare stakeholders (which, in turn, originate from the shared values and expectations to which all the healthcare stakeholders subscribe).

Melvin, et al. (1999) have noted that, in the UK, the NHS has started to support the concept of clinical governance by identifying individual best effective clinical practices. This process provides concise summaries of the most effective clinical practices in all key clinical areas. Summaries that are successfully substantiated are then disseminated throughout the NHS. Sewell (1997) has noted that the USA, Canada, Australia and New Zealand have adopted a formal accreditation system for the healthcare sector based upon the ISO 9000 approach.

COMMUNITY HEALTH INFORMATION NETWORKS (CHIN)

Modern day healthcare organizations have realized that in the future their survival would depend upon their ability to give the caregiver access to such information that would enable the caregiver to deliver personalized clinical diagnosis and treatment in real-time in very specific clinical contexts, a process termed Information Therapy (Dwivedi et al., 2002a). This vision has been translated into concepts such as Integrated Delivery System (IDS) and Community Health Information Networks (CHIN) (Lang, 1997; Mercer, 2001; Morrissey, 2000).

IDS refers to a Healthcare Information System (HIS), a business model based on computing technologies such as Object Orientation (OO) "to share key data, with partners and providers, that will allow faster and more accurate decision making … to deliver care to a broader population with fewer requirements for expensive and scarce resources" (Lang, 1997, p.18).

CHINs are integrated healthcare institutions based upon a combination of different technology platforms connected to enable support for data sharing amongst different healthcare providers (Mercer, 2001). Both IDS and CHIN are very similar in nature and both refer to an integrated network for allowing the delivery of personalized healthcare. CHINs were founded on the premise that patient information should be shared by competitors (Morrissey, 2000). The main aim of CHIN was to enable hospitals and other healthcare stakeholders to electronically exchange patient encounter summaries and medical records between emergency departments and related departments.

Another factor responsible for emphasis on CHIN was the perception in the healthcare industry that, for small-scale players to survive as individual entities, it was essential for them to form some sort of technological alliances (Huston & Huston, 2000). The original technological objective of CHIN was to enhance data-sharing capabilities amongst different healthcare stakeholders. The original technological infrastructure supported the creation of "point to point" connections. This did not succeed primarily due to limitations in technology coupled with the high amount of financial resources required to establish the "point to point" technological infrastructure (Morrissey, 2000).

The objective behind the incorporation of the CHIN concept is that it allows users to collect data which could be used to formulate "best practice protocols for effective treatment at a low-cost", that is, clinical best evidence practices for both healthcare diagnosis and delivery (Kennedy, 1995). It was anticipated that the advent of CHINs in conjunction with Internet technologies would empower healthcare stakeholders to provide healthcare to patients in real time whilst being in geographically distinct locations (Kennedy, 1995).

KM TAXONOMIES

KM has become an important focus area for organizations (Earl & Scott, 1999). It has been argued that KM evolved from the applications of expert systems and artificial intelligence (Liebowitz & Beckman, 1998; Sieloff, 1999). Almost all of the definitions of KM state that it is a multi-disciplinary paradigm (Gupta, Iyer & Aronson, 2000) that has further accentuated the controversy regarding the origins of KM. One of the main factors behind widespread interest in KM is its role as a possible source of competitive advantage (Havens & Knapp, 1999; Nonaka, 1991). A number of leading management researchers have affirmed that the Hungarian chemist, economist and philosopher Michael Polanyi was among the earliest theorists who popularized the concept of characterizing knowledge as "tacit or explicit" which is now recognized as the accepted knowledge categorization approach (Gupta et al., 2000; Hansen, Nohria & Tierney, 1999; Zack, 1999).

The cornerstone of any KM project is to transform tacit knowledge to explicit knowledge so as to allow its effective dissemination (Gupta et al., 2000). This can be best met by developing a KM framework. Authors such as Blackler (1995) have reiterated that the concept of knowledge is complex and, in an organizational context, its relevance to organization theory has not yet been sufficiently understood and documented. This is one of the fundamental reasons why KM does not have a widely accepted framework that can enable healthcare institutions in creating KM systems and a culture conducive to KM practices.

KM is underpinned by information technology paradigms such as Workflow, Intelligent Agents and Data Mining. According to Manchester (1999), a common point about software technologies such as (1) information retrieval, (2) document management and (3) workflow processing is that they blend well with the Internet and related technologies (i.e., technologies which focus on dissemination of information). Deveau (2000, p. 14) submits that: "KM is about mapping processes and exploiting the knowledge database. It's taking people's minds and applying technology." Deveau (2000) also noted that information technology puts the organization in a position to state the currently available information in the organizational knowledge base. At this point, the role of IT ends and the role of KM commences.

As KM deals with the tacit and contextual aspects of information, it allows an organization to know what is important for it in particular circumstances, in the process maximizing the value of that information and creating competitive advantages and wealth.

APPLICABILITY OF THE KM PARADIGM IN HEALTHCARE

A KM solution would allow healthcare institutions to give clinical data context, so as to allow knowledge derivation for more effective clinical diagnoses. In the future, healthcare systems would see increased interest in knowledge recycling of the collaborative learning process acquired from previous healthcare industry practices. This chapter puts forward the notion that this sector has been exclusively focused on IT to meet the challenges described above and reiterates that this challenge cannot be met by an IT led solution.

KM initiatives should be incorporated within the technological revolution that is speeding across healthcare industry. There has to be balance between organizational and technological aspects of the healthcare process, that is, one cannot exist without the other (Dwivedi et al., 2001a). This chapter emphasizes the importance of clinicians taking a holistic view of their organization. Clinicians therefore need to have an understanding of IT in a healthcare context and a shared vision of the organization. Clinicians and healthcare administrators thus need to acquire both organizational and technological insights if they are to have a holistic view of their organization.

The KM paradigm can enable the healthcare sector to successfully overcome the information and knowledge explosion, made possible by adopting a KM framework that is specially customized for healthcare institutions in light of their ICT implementation level. Adoption of KM is essential for healthcare institutions as it would enable them to identify, preserve and disseminate "best context" healthcare practices to different healthcare stakeholders.

PREFATORY ANALYSIS OF ALTERNATIVE HEALTHCARE CONCEPTS

The failure of some healthcare management concepts propelled a new stream of thought that advocated the incorporation of the KM paradigm in healthcare (Health Canada, 1999; Mercer, 2001). KM could allow healthcare organizations to truly take advantage of the driving forces behind the creation of the CHIN concept. However, very few organizations have adopted a comprehensive healthcare KM system. The main reason attributed is the failure of healthcare stakeholders in properly creating a conducive organizational culture. Based on a literature review above, a preliminary conceptual analysis of alternative healthcare management concepts is presented in Table 1. As can be seen from the table, healthcare stakeholders are searching for alternative paradigms that support collaboration in order to synergistically learn from others' experiences, training and knowledge within specific organizational cultures. Healthcare institutions have realized that existing concepts such as EBM and CG do not enable healthcare stakeholders to achieve this challenge as they do not holistically support effective integration of IT within specific organizational cultures and processes. Contemporary concepts such as EBM, CHIN, ICHDS and IPP focus on IT at the expense of having too little emphasis on people. This is further aggravated by the presence of dysfunctional organizational processes in the majority of healthcare institutions.

CONCLUSION

For any healthcare organization to succeed, it needs to excel in a number of key processes (i.e., patient diagnosis, care treatment, etc.) that are necessary for it to achieve its mission. If the processes are repetitive, automation is possible via the use of IT. Modern IT applications in healthcare are not sufficient in meeting the information needs of current healthcare institutions as they lack the ability to deliver precise, accurate and contextual information to the desired caregiver at the desired time.

This chapter has presented an analysis of alternative healthcare management concepts with respect to their ability in providing a solution to the issue of information management. Furthermore, this chapter has examined the feasibility of the KM paradigm in solving the problem of information explosion in healthcare and has found validation for the proposition that the current focus on technological solutions will aggravate the problem of explosion in clinical information systems for healthcare institutions.

Table 1. Prefatory analysis of alternative healthcare concepts

Concept	Support for People	Support for Process	Support for Technology	Limitations
CG	Present	Insufficient	Present	Policy initiative
EBM	Insufficient	Insufficient	Present	Tacit Processes?
CHIN	Insufficient	Absent	Present	Limited Trials
IHCDS	Insufficient	Insufficient	Present	Technology focus
IPP	Insufficient	Present	Present	Tacit Knowledge?
KM	Present	Present	Present	Not validated

The chapter has also presented the key requirements for creating a KM framework, which can act as a template in enabling healthcare institutions in their attempts to initiate KM projects. This chapter concludes that any potential solution has to come from a domain that synergistically combines people, organizational processes and technology, thereby enabling healthcare stakeholders to have a holistic view of the entire healthcare continuum. This chapter further concludes that KM is the only paradigm that combines the above-mentioned perspectives (i.e., people, organizational processes, and technology) into healthcare and as such, KM is the next indispensable step for integrated healthcare management.

REFERENCES

Blackler, F. (1995). Knowledge, knowledge work and organizations: An overview and interpretation. *Organization Studies*, *16*(6), 1021-1046.

Cowling, A., Newman, K. & Leigh, S. (1999). Developing a competency framework to support training in evidence-based healthcare. *International Journal of Health Care Quality Assurance*, *12*(4), 149-159.

Davis, M. & Klein, J. (2000). Net holds breakthrough solutions. *Modern Healthcare*, Supplement, Eye on Info. Feb 7, 2000, 14-15.

Deveau, D. (2000). Minds plus matter: Knowledge is business power. *Computing Canada*, *26*(8), 14 -15.

Dwivedi, A., Bali, R.K., James, A.E. & Naguib, R.N.G. (2001a). Telehealth systems, Considering knowledge management and ICT issues. *Proceedgins of the IEEE-EMBC 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, [CD-ROM], Istanbul, Turkey.

Dwivedi, A., Bali, R.K., James, A.E. & Naguib, R.N.G. (2001b). Workflow management systems, The healthcare technology of the future? *Proceedings of the IEEE EMBC-2001 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, [CD-ROM], Istanbul, Turkey.

Dwivedi, A., Bali, R.K., James, A.E. & Naguib, R.N.G. (2002a). The efficacy of using object oriented technologies to build collaborative applications in healthcare and medical information systems. *Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering (CCECE) 2002*, Winnipeg, Canada (Vol. 2, pp. 1188-1193).

Dwivedi, A., Bali, R.K., James, A.E., Naguib, R.N.G. & Johnston, D. (2002b). Merger of knowledge management and information technology in healthcare: Opportunities and challenges. *Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering (CCECE) 2002*, Winnipeg, Canada (Vol. 2, pp.1194-1199).

Earl, M. & Scott, I. (1999). What is a chief knowledge officer? *Sloan Management Review*, *40*(2), 29-38.

Firth-Cozens, J. (1999). Clinical governance development needs in health service staff. *Clinical Performance and Quality Health Care*, *7*(4), 155-160.

Gupta, B., Iyer, L.S. & Aronson, J.E. (2000). Knowledge management: Practices and challenges. *Industrial Management & Data Systems*, *100*(1), 17-21.

Hansen, M.T., Nohria, N. & Tierney, T. (1999). What's your strategy for managing knowledge? *Harvard Business Review*, *77*(2), 106-116.

Havens, C. & Knapp, E. (1999). Easing into knowledge management. *Strategy & Leadership*, *27*(2), 4-9.

Health Canada (1998). Vision and strategy for knowledge management and IM/IT for Health Canada [online]. Retrieved March 20, 2002, from *http://www.hc-sc.gc.ca/iacb-dgiac/km-gs/english/vsmenu_e.htm*

Huston, T.L. & Huston, J.L. (2000). Is telemedicine a practical reality? *Association for Computing Machinery, Communications of the ACM, 43*(6), 91-95.

Kennedy, M. (1995). Integration fever. *Computerworld, 29*(14), 81-83.

Lang, R.D. (1997). The promise of Object Technology. *Journal of Healthcare Resource Management, 15*(10), 18-21.

Latamore, G.B. (1999). Workflow tools cut costs for high quality care. *Health Management Technology, 20*(4), 32-33.

Liebowitz, J. & Beckman, T. (1998). *Knowledge organizations: What every manager should know.* USA: St. Lucie Press.

Manchester, P. (1999). Technologies form vital component, INFRASTRUCTURE, The IT infrastructure will capture, store and distribute the information that might be turned into knowledge. *Financial Times,* London, Nov 10, 1999, 8.

Melvin, K., Wright, J., Harrison, S.R., Robinson, M., Connelly, J. & Williams, D.R. (1999). Effective practice in the NHS and the contribution from public health, a qualitative study. *British Journal of Clinical Governance, 4*(3), 88-97.

Mercer, K. (2001). Examining the impact of health information networks on health system integration in Canada. *Leadership in Health Services, 14*(3), 1-30.

Morrissey, J. (2000). The evolution of a CHIN. *Modern Healthcare*, Supplement, Eye on Info, May 29, 2000, 42-43.

Nonaka, I. (1991). The knowledge-creating company. *Harvard Business Review, 69*(6), 96-104.

Pryga, E. & Dyer, C. (1992). The AHA's national health reform strategy: An overview. *Trustee, 45*(5), 10.

Schmid, K. & Conen, D. (2000). Integrated patient pathways, "mipp"—A tool for quality improvement and cost management in health care. Pilot study on the pathway "Acute Myocardial Infarction". *International Journal of Health Care Quality Assurance, 13*(2), 87-92.

Sewell, N. (1997). Continuous quality improvement in acute health care, creating a holistic and integrated approach. *International Journal of Health Care Quality Assurance, 10*(1), 20-26.

Sieloff, C. (1999). If only HP knew what HP knows: The roots of knowledge management at Hewlett-Packard. *Journal of Knowledge management, 3*(1), 47-53.

Zack, M.H. (1999). Managing codified knowledge. *Sloan Management Review, 40*(4), 45-58.

Zairi, M. & Whymark, J. (1999). Best practice organisational effectiveness in NHS Trusts Allington NHS Trust case study. *Journal of Management in Medicine, 13*(5), 298-307.

This work was previously published in Clinical Knowledge Management: Opportunities and Challenges, edited by B. Bali, pp. 1-10, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 1.21
Interactive Information Retrieval as a Step Towards Effective Knowledge Management in Healthcare

Jörg Ontrup
Bielefeld University, Germany

Helge Ritter
Bielefeld University, Germany

ABSTRACT

The chapter shows how modern information retrieval methodologies can open up new possibilities to support knowledge management in healthcare. Recent advances in hospital information systems lead to the acquisition of huge quantities of data, often characterized by a high proportion of free narrative text embedded in the electronic health record. We point out how text mining techniques augmented by novel algorithms that combine artificial neural networks for the semantic organization of non-crisp data and hyperbolic geometry for an intuitive navigation in huge data sets can offer efficient tools to make medical knowledge in such data collections more accessible to the medical expert by providing context information and links to knowledge buried in medical literature databases.

INTRODUCTION

During the last years the development of electronic products has lead to a steadily increasing penetration of computerized devices into our society. Personal computers, mobile computing platforms, and personal digital assistants are becoming omnipresent. With the advance of network aware medical devices and the wireless transmission of health monitoring systems pervasive computing has entered the healthcare domain (Bergeron, 2002).

Consequently, large amounts of electronic data are being gathered and become available online. In order to cope with the sheer quantity, data warehousing systems have been built to store and organize all imaginable kinds of clinical data (Smith & Nelson, 1999).

In addition to numerical or otherwise measurable information, clinical management involves administrative documents, technical documentation regarding diagnosis, surgical procedure, or care. Furthermore, vast amounts of documentation on drugs and their interactions, descriptions of clinical trials or practice guidelines plus a plethora of biomedical research literature are accumulated in various databases (Tange, Hasman, de Vries, Pieter, & Schouten, 1997; Mc Cray & Ide, 2000).

The storage of this exponentially growing amount of data is easy enough: From 1965—when Gordon Moore first formulated his law of exponential growth of transistors per integrated circuit—to today: "Moore's Law" is still valid. Since his law generalizes well to memory technologies, we up to now have been able to cope with the surge of data in terms of storage capacity quite well. The retrieval of data however, is inherently harder to solve. For structured data such as the computer-based patient record, tools for online analytical processing are becoming available. But when searching for medical information in freely formatted text documents, healthcare professionals easily drown in the wealth of information: When using standard search engine technologies to acquire new knowledge, thousands of "relevant" hits to the query string might turn up, whereas only a few ones are valuable within the individual context of the searcher. Efficient searching of literature is therefore a key skill for the practice of evidence-based medicine (Doig & Simpson, 2003).

Consequently, the main objective of this chapter is to discuss recent approaches and current trends of modern information retrieval, how to search very large databases effectively. To this end, we first take a look at the different sources of information a healthcare professional has access to. Since unstructured text documents introduce the most challenges for knowledge acquisition, we will go into more detail on properties of text databases considering MEDLINE as the premier example. We will show how machine learning techniques based on artificial neural networks with their inherent ability for dealing with vague data can be used to create structure on unstructured databases, therefore allowing a more natural way to interact with artificially context-enriched data.

SOURCES OF INFORMATION IN HEALTHCARE

The advance of affordable mass storage devices encourages the accumulation of a vast amount of healthcare related information. From a technical point of view, medical data can be coarsely divided into structured and unstructured data, which will be illustrated in the following sections.

Data Warehouses and Clinical Information Systems

One major driving force for the development of information processing systems in healthcare is the goal to establish the computer-based patient record (CPR). Associated with the CPR are patient registration information, such as name, gender, age or lab reports such as blood cell counts, just to name a few. This kind of data is commonly stored in relational databases that impose a high degree of structure on the stored information. Therefore, we call such data "structured." For evidence based medicine the access to clinical information is vital. Therefore, a lot of effort has been put into the goal to establish the computer-based patient record (CPR) as a standard technology in healthcare. In the USA, the Institute of Medicine (1991) defined the CPR as "an electronic patient record that resides in a system specifically designed to support users by providing accessibility to complete and accurate data, alerts, reminders, clinical decision support systems, links to medical knowledge, and other aids." However, as Sensmeier (2003) states, "this goal remains a vision [...] and the high expectations of these visionaries remain largely unfulfilled." One of the main reasons why "we are not there yet" is a general lack of integration.

A recent study by Microsoft[1] identified as much as 300 isolated data islands within a medical center. The unique requirements of different departments such as laboratory, cardiology, or oncology and the historical context within the hospital often results in the implementation of independent, specialized clinical information systems (CIS) which are difficult to connect (McDonald, 1997). Consequently, main research thrusts are strategies to build data warehouses and clinical data repositories (CDR) from disjoint "islands of information" (Smith & Nelson, 1999). The integration of the separated data sources involves a number of central issues:

1. Clinical departments are often physically separated. A network infrastructure is necessary to connect the disparate sites. Internet and Intranet tools are regarded as standard solutions.
2. Due to the varying needs of different departments, they store their data in various formats such as numerical laboratory values, nominal patient data or higher dimensional CT data. Middleware solutions are required to translate these into a consistent format that permits interchange and sharing.
3. In the real world, many data records are not suitable for further processing. Missing values or inconsistencies have to be treated: Some fields such as blood cell counts are not always available or data was entered by many different persons, and is thus often furnished with a "personal touch." Cleaning and consistency checks are most commonly handled by rule-based transformations.
4. Departmental workflow should not be strained by data copying tasks. Automated script utilities should take care of the data migration process and copy the data "silently in the background."

None of the steps above is trivial and certainly it requires a strong management to bring autonomous departments together and to motivate them to invest their working power for a centralized data warehouse. But once such a centralized repository of highly structured data is available it is a very valuable ground for retrospective analysis, aggregation and reporting. Online analytical processing (OLAP) tools can be used to perform customized queries on the data while providing "drill downs" at various levels of aggregation. A study with an interactive visualization framework has shown that it can massively support the identification of cost issues within a clinical environment (Bito, Kero, Matsuo, Shintani, & Silver, 2001). Silver, Sakata, Su, Herman, Dolins, and O'Shea (2001) have demonstrated that data mining "helped to turn data into knowledge." They applied a patient rule induction method and discovered a small subgroup of inpatients on the DRG level which was responsible for more than 50 percent of revenue loss for that specific DRG (diagnoses related group). The authors describe a "repeatable methodology" which they used to discover their findings: By using statistical tools they identified regions and patterns in the data, which significantly differentiated them from standard cases. In order to establish the *reason* for increased costs in those regions, they found it necessary that a human expert further analyzes the automatically encircled data patterns. Therefore, a closed loop between clinical data warehouse, OLAP tool and human expert is necessary to gain the maximum of knowledge from the available data.

A recently proposed data mart by Arnrich, Walter, Albert, Ennker, and Ritter (2004) targets in the same direction: Various data sources are integrated into a single data repository. Results from statistical procedures are then displayed in condensed form and can be further transformed into new medical knowledge by a human expert.

UNSTRUCTURED DATA IN BIOMEDICAL DATABASES

As the discussion on hospital information systems shows, interactive data processing tools form a large potential to extract hidden knowledge from large structured data repositories. A yet more challenging domain for knowledge acquisition tasks is the field of unstructured data. In the following we will briefly present the two premier sources of unstructured data available to the healthcare professional.

Narrative Clinical Data

Many hospitals do not only store numerical patient data such as laboratory test results, but also highly unstructured data such as narrative text, surgical reports, admission notes, or anamneses. Narrative text is able to capture human-interpretable nuances that numbers and codes cannot (Fisk, Mutalik, Levin, Erdos, Taylor, & Nadkarni, 2003). However, due to the absence of a simple machine-parsable structure of free text, it is a challenging task to design information systems allowing the exchange of such kind of information.

In order to address the interoperability within and between healthcare organizations widely accepted standards for Electronic Health Records (EHR) are important. Health Level 7 (HL7)—an ANSI accredited organization, and the European CEN organization both develop standards addressing this problem. The HL7 Clinical Document Architecture (CDA) framework "stems from the desire to unlock a considerable clinical content currently stored in free-text clinical notes and to enable comparison of content from documents created on information systems of widely varying characteristics." (Dolin, Alschuler, Beebe, Biron, Boyer, Essin, Kimber, Lincoln, & Mattison,, 2001). The emerging new standard based on HL7 version 3 is purely based on XML and therefore greatly enhances the interoperability of clinical information systems by embedding the unstructured narrative text in a highly structured XML framework. Machine-readable semantic content is added by the so-called HL7 Reference Information Model (RIM). The RIM is an all-encompassing look at the entire scope of healthcare containing more than 100 classes with more than 800 attributes. Similar to HL7 v3, the European ENV 13606 standard as defined by the TC 251 group of the CEN addresses document architecture, object models and information messages. For a more detailed discussion on healthcare information standards see Spyrou, Bamidis, Chouvarda, Gogou, Tryfon, and Maglaveras (2002).

Another approach to imprint structure on free text is the application of ontologies to explicitly formalize logical and semantic relationships. The W3C consortium has recently (as of May 2004) recommended the Web Ontology Language (OWL) as a standard to represent semantic content of web information and can therefore be seen as a major step towards the "semantic Web" (Berners-Lee, Hendler, & Lassila, 2001). In the medical domain, the Unified Medical Language System (UMLS) uses a similar technology to represent biomedical knowledge within a semantic network (Kashyap, 2003). Once an ontology for a certain domain is available, it can be used as a template to transfer the therein contained explicit knowledge to corresponding terms of free text from that domain. For example, the term "blepharitis" is defined as "inflammation of the eyelids" in the UMLS. If a medic professional would search for the term "eyelid" within the hospital's EHRs, documents containing the word "blepharitis" would not show up. However with knowledge from the UMLS, these documents could be marked as relevant to the query. However, the tagging of free text with matching counterparts from an ontology is either a cost intensive human labor or requires advanced natural language processing methods. We shall see below, how computational approaches could address the problem of adding semantics to free narrative text.

Biomedical Publications

For most of the time in human history, the majority of knowledge acquired by mankind was passed on by the means of the written word. It is a natural scientific process to formulate new insights and findings and pass them to the community by publishing. Consequently, scientific literature is constantly expanding. It is not only expanding at growing speed, but also increasing in diversification, resulting in highly specialized domains of expertise.

The MEDLINE database at the U.S. National Library of Medicine (NLM) has become the standard bibliographic database covering the fields of medicine, nursing, veterinary medicine, dentistry, pharmacology, the healthcare system, and the preclinical sciences. The majority of publications in MEDLINE are regular journals and a small number of newspapers and magazines. The database is updated on a regular daily basis from Tuesdays to Saturdays and on an irregular basis in November and December. At the time of this writing, MEDLINE contains approximately 12 million references to articles from over 37,000 international journals dating back to 1966. In 1980 there were about 100,000 articles added to the database, and in 2002 the number increased to over 470,000 newly added entries. That means, currently over 50 new biomedical articles are published on an hourly basis. Clearly, no human is able to digest one scientific article per minute 24 hours a day, seven days a week. But the access to the knowledge within is lively important nonetheless: "The race to a new gene or drug is now increasingly dependent on how quickly a scientist can keep track of the voluminous information online to capture the relevant picture … hidden within the latest research articles" (Ng & Wong, 1999).

Not surprisingly, a whole discipline has emerged which covers the field of "information retrieval" from unstructured document collections, which takes us to the next section.

COMPUTATIONAL APPROACHES TO HANDLE TEXT DATABASES

The area of information retrieval (IR) is as old as man uses libraries to store and retrieve books. Until recently, IR was seen as a narrow discipline mainly for librarians. With the advent of the World Wide Web, the desire to find relevant and useful information grew to a public need. Consequently, the difficulties to localize the desired information "have attracted renewed interest in IR and its techniques as promising solutions. As a result, almost overnight, IR has gained a place with other technologies at the center of the stage" (Baeza-Yates & Ribeiro-Neto, 1999).

Several approaches have shown that advanced information retrieval techniques can be used to significantly alleviate the cumbersome literature research and sometimes even discover previously unknown knowledge (Mack & Hehenberger, 2002). Swanson and Smalheiser (1997) for example designed the ARROWSMITH system which analyses journal titles on a keyword level to detect hidden links between literature from different specialized areas of research. Other approaches utilize MeSH headings (Srinivasan & Rindflesch, 2002) or natural language parsing (Libbus & Rindflesch, 2002) to extract knowledge from literature. Rzhetsky, Iossifov, Koike, Krauthammer, Kra, Morris, Yu, Duboue, Weng, Wilbur, Hatzivassiloglou, and Friedman (2004) have recently proposed an integrated system which visualizes molecular interaction pathways and therefore has the ability to make published knowledge literally visible.

In the following sections we go into more detail and look how computational methods can deal with text.

Representing Text

The most common way to represent text data in numerical form such that it can be further processed by computational means is the so-called "bag of words" or "vector space" model. To this end a standard practice composed of two complementary steps has emerged (Baeza-Yates & Ribeiro-Neto, 1999):

1. **Indexing**: Each document is represented by a high-dimensional feature vector in which each component corresponds to a term from a dictionary and holds the occurrence count of that term in the document. The dictionary is compiled from all documents in the database and contains all unique words appearing throughout the collection. In order to reduce the dimensionality of the problem, generally only word stems are considered (i.e., the words "computer", "compute", and "computing" are reduced to "comput"). Additionally, very frequently appearing words (like "the" & "and") are regarded as *stop words* and are removed from the dictionary, because they do not contribute to the distinguishability of the documents.
2. **Weighting**: Similarities between documents an then be measured by the Euclidean distance or the angle between their corresponding feature vectors. However, this ignores the fact that certain words carry more information content than others. The so-called *tf-idf* weighting scheme (Salton & Buckley, 1988) has been found a suitable means to deal with this problem: Here the *term frequencies (tf)* in the document vector are weighted with the *inverse document frequency idfi* = $\log(N/s^d_i)$, where N is the number of documents in the collection, and s^d_i denotes the number of times the word stem s_i occurs in document d.

Although the "bag of words" model completely discards the order of the words in a document, it performs surprisingly well to capture its content. A major drawback is the very high dimensionality of the feature vectors from which some computational algorithms suffer. Additionally, the distance measure typically does not take *polysemy* (one word has several meanings) and *synonymy* (several words have the same meaning) into account.

Evaluating Retrieval Performance

In order to evaluate an IR system, we need some sort of quality measure for its retrieval performance. The most common and intuitively accessible measures are *precision* and *recall*. Consider the following situation: We have a collection of *C* documents and a query *Q* requesting some information from that set. Let *R* be the set of relevant documents in the database to that query, and *A* be the set of documents delivered by the IR system as answers to the given query *Q*. The *precision* and *recall* are then defined by

$$precision = \frac{|R \cap A|}{|R|} \text{ and } recall = \frac{|R \cap A|}{|R|},$$

where $R \cap A$ is the intersection of the relevant and the answer set, that is, the set of found relevant documents. The ultimate goal is to achieve a precision and recall value of one, that is, to find only the relevant documents, and all of them. A user will generally not inspect the whole answer set *A* returned by an IR system—who would want to click through all of the thousands of hits typically returned by a web search? Therefore, a much better performance description is given by the so-called *precision-recall-curve*. Here, the documents in *A* are sorted according to some ranking criterion and the precision is plotted against the recall values obtained when truncating the sorted answer set after a given rank—which is varied between one and the full size of A. The particular point on the curve where the precision equals the recall is called *break even point* and is frequently used to characterize an IR system's performance with a single value.

Therefore, the quality of a search engine which generates hit lists is strongly dependent on its ranking algorithm. A famous example is PageRank, which is part of the algorithm used by Google. In essence, each document (or page) is weighted with the number of hyperlinks it receives from other web pages. At the same time the links themselves are weighed by the importance of the linking page (Brin & Page, 1998). The higher a page is weighted, the higher it gets within the hit list. The analogue to hyperlinks in the scientific literature are citations, that is, the more often an article is cited, the more important it is (Lawrence, Giles & Bollacker, 1999).

The Role of Context

So, we can handle unstructured text and measure how good an IR system works, but a critical question was not answered yet: How do we determine, which documents from a large collection are actually *relevant* to a given query? To answer this question, we have to address the role of *context.*

User Context

Consider the following example: A user enters the query "Give me all information about jaguars." The set of relevant documents will be heavily dependent on the context in which the user phrases this query: A motor sports enthusiast is probably interested in news about the "Jaguar Racing" Formula 1 racing stable, a biologist probably wants to know more about the animal. So, the context—in the sense of the situation in which the user is immersed—plays a major role (Johnson, 2003; Lawrence, 2000).

Budzik, Hammond and Birnbaum (2001) discuss two common approaches to handle different user contexts:

1. Relevance feedback: The user begins with a query and then evaluates the answer set. By providing positive or negative feedback to the IR system, this can modify the original query by adding positive or negative search terms to it. In an iterative dialogue with the IR system, the answer set is then gradually narrowed down to the relevant result set. However, as studies (Hearst, 1999) have shown, users are generally reluctant to give exhaustive feedback to the system.
2. Building user profiles: Similar to the relevance feedback, the IR system builds up a user profile across multiple retrieval sessions, that is, with each document the user selects for viewing, the profile is adapted. Unfortunately, such a system does not take account of "false positives", that is, when a user follows a link that turned out to be of no value when inspecting it closer. Additionally, such systems integrate short term user interests into accumulated context profiles, and tend to inhibit highly specialized queries which the user is currently interested in.

Budzik, Hammond & Birnbaum (2001) presented a system that tries to guess the user context from open documents currently edited or browsed on the work space. Their system constitutes an "Information Management Assistant [which] observes user interactions with everyday applications and anticipates information needs […] in the context of the user's task." Another system developed by Finkelstein and Gabrilovich (2002) analyzes the context in the immediate vicinity of a user-selected text, therefore making the context more focused. In an evaluation of their system both achieved consistently better results than standard search engines are able to achieve without context.

Document Context

Context can also be seen from another point of view: Instead of worrying about the user's intent and the context in which a query is embedded, an information retrieval system could make the context in which retrieved documents are embedded more explicit. If each document's context is immediately visible to the information seeker, the relevant context might be quickly picked out. Additionally, the user might discover a context he had not in mind when formulating the query and thus find links between his intended and an unanticipated context. In the following we describe several approaches that aim in this direction.

Augmenting Document Sets with Context

Text Categorization

One way to add context to a document is by assigning a meaningful label to it (Le & Thoma, 2003). This constitutes a task of text categorization and there exist numerous algorithms that can be applied. The general approach is to select a training set of documents that are already labelled. Based on the "bag of words" representation, machine learning methods learn the association of category labels to documents. For an in depth review of statistical approaches (such as naives Bayes or decision trees) see Yang (1999). Computationally more advanced methods utilize artificial neural network architectures such as the support vector machine which have achieved break even values close to 0.9 for the labelling of news wire articles (Joachims, 1998; Lodhi, 2001).

However, in the medical domain, 100 categories are seldom adequate to describe the context of a text. In case of the MEDLINE database, the National Library of Medicine has developed a highly standardized vocabulary, the Medical Subject Headings (MeSH) (Lowe & Barnett, 1994). They consist of more than 35,000 categories that are hierarchically organized and constitute the basis for searching the database. To guarantee satisfactory search results of constant quality, reproducible labels are an important prerequisite. However, the cost of human indexing of the biomedical literature is high: according to Humphrey (1992) it takes one year to train an expert the task of document labelling. Additionally, the labelling process lacks a high degree of reproducibility. Funk, Reid, and McGoogan (1983) have reported a mean agreement in index terms ranging from 74 percent down to as low as 33 percent for different experts. Because the improvement of index consistency is such demanding, assistance systems are considered to be a substantial benefit. Recently, Aronson, Bodenreider, Chang, Humphrey, Mork, Nelson, Rindflesh, and Wilbur (2000) have presented a highly tuned and sophisticated system which yields very promising results. Additionally to the bag of words model their system utilizes a semantic network describing a rich ontology of biomedical knowledge (Kashyap, 2003).

Unfortunately, the high complexity of the MeSH terms makes it hard to incorporate a MeSH-based categorization into a user interface. When navigating the results of a hierarchically ordered answer set it can be a time-consuming and frustrating process: Items which are hidden deep within the hierarchy can often only be obtained by descending a tree with numerous mouse-clicks. Selection of a wrong branch requires backing up and trying a neighboring branch. Since screen space is a limited resource only a small area of context is visible and requires internal "recalibration" each time a new branch is selected. To overcome this problem, *focus and context* techniques are considered to be of high value. These are discussed in more detail below.

The Cluster Hypothesis

An often cited statement is the *cluster hypothesis*, which states that documents which are similar in their bag of words feature space tend to be relevant to the same request (van Rijsbergen, 1979). Leuski (2001) has conducted an experiment where an agglomerative clustering method was used to group the documents returned by a search engine query. He presented the user not with a ranked list of retrieved documents, but with a list of clusters, where each cluster in turn was arranged as a list of documents. The experiment showed, that this procedure "can be much more helpful in locating the relevant information than the traditional ranked list." He could even show that the clustering can be as effective as the relevance feedback methods based on query expansion. Other experiments also validated the cluster hypothesis on several occasions (Hearst & Pedersen, 1996; Zamir & Etzioni, 1999). Recently the *vivisimo*[2] search engine has drawn attention by utilizing an online clustering of retrieval sets, which also includes an interface to PubMed / MEDLINE.

Visualizing Context

Another way to create context in line with the spirit of the cluster hypothesis is by embedding the document space in a visual display. By making the relationship between documents visually more explicit, such that the user can actually *see* inter-document similarities, the user gets *(i)* an overview of the whole collection, and *(ii)* once a relevant document has been found, it is easier to locate others, as these tend to be grouped within the surrounding context of already identified valuable items. Fabrikant and Buttenfield (2001) provide a more theoretical framework for the concept of "spatialization" with relation to cognitive aspects and knowledge acquisitions: Research on the cognition of geographic information has been identified as being important in decision making, planning and other areas involving human-related activities in space.

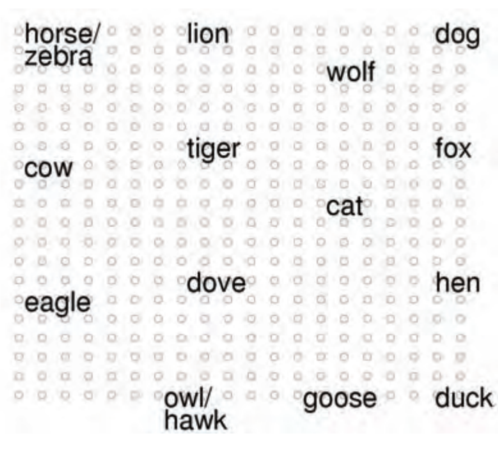
In order to make use of "spatialization", that is, to use cognitive concepts such as "nearness" we need to apply some sort of transformation to project the documents from their high-dimensional (typically several thousands) bag-of-words space onto a two dimensional canvas suitable for familiar inspection and interaction. The class of algorithms performing such a projection is called *multi dimensional scaling* (MDS). A large number of MDS methods such as Sammon mapping, spring models, projection pursuit or local linear embeddings have been proposed over the years. Skuping and Fabrikant (2003) provide an excellent discussion of the most common MDS variants in relation to spatialized information visualization.

Self-Organizing Maps

The notion of the self-organizing map (SOM) has been introduced by Kohonen (1982) more than 20 years ago. Since then is has become a well-accepted tool for exploratory data analysis and classification. While applications of the SOM are extremely wide spread—ranging from medical imaging, classification of power consumption profiles, or bank fraud detection—the majority of uses still follows its original motivation: to use a deformable template to translate *data similarities* into *spatial relations.*

The following example uses a simple toy dataset to demonstrate the properties of the SOM

Figure 1. A map of animals



algorithm. Consider a 13-dimensional dataset describing animals with properties such as s*mall, medium, big, two legs, hair, hooves, can fly, can swim, and* so on. When displaying the data as a large table, it is quite cumbersome to see the interrelationship between the items, that is, animals. The map shown in Figure 1 depicts a trained SOM with neurons placed on a 20x20 regular grid. After the training process, each animal is "presented" to the map, and the neuron with the highest activity gets labelled with the corresponding name. As can be seen in the figure the SOM achieves a semantically reasonable mapping of the data to a two-dimensional "landscape": birds and non-birds are well separated and animals with identical features get mapped to identical nodes.

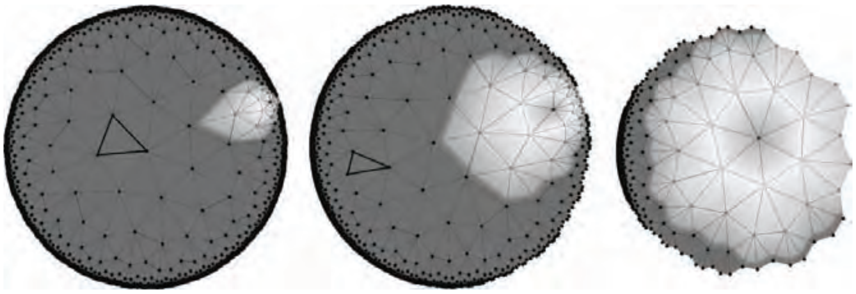
There have been several approaches to use the SOM to visualize large text databases and relations of documents therein. The most prominent example is the WEBSOM project. Kohonen, Kaski and Lagus, (2000) performed the mapping of 7 million patent abstracts and obtained a semantic landscape of the corresponding patent information. However, screen size is a very limited resource, and as the example with 400 neurons above suggests, a network with more than one million nodes—2500 times that size—becomes hard to visualize: When displaying the map as a whole, annotations will not be readable, and when displaying a legible subset, important surrounding context will be lost. Wise (1999) has impressively demonstrated the usefulness of compressed, map-like representations of large text collections: His *ThemeView* reflects major topics in a given area, and a zoom function provides a means to magnify selected portions of the map—unfortunately without a coarser view to the surrounding context.

Focus & Context Techniques

The limiting factor is the two dimensional Euclidean space we use as a data display: The neighborhood that "fits" around a point is rather restricted: namely by the square of the distance to that point. An interesting loophole is offered by *hyperbolic space*: it is characterized by uniform negative curvature and results in a geometry, where the neighborhood around a point increases *exponentially* with the distance. This exponential behavior was firstly exploited (and patented) by the "hyperbolic tree browser" from Lamandpng & Rao (1994), followed by a Web content viewer by Munzner (1998). Studies by Pirolli, Card and van der Wege (2001) showed that the particular *focus & context* property offered by hyperbolic space can significantly accelerate "information foraging".

Naturally, it becomes apparent to combine the SOM algorithm with the favorable properties of hyperbolic space (Ritter, 1999). The core idea of the hyperbolic self-organizing map (HSOM) is to employ a grid of nodes in the hyperbolic plane H^2. For H^2 there exists an infinite number of tessellations with congruent polygons such that each grid point is surrounded by the same number of neighbors (Magnus, 1974). As stated above, an intuitive navigation and interaction methodology is a crucial element for a well-fitted visualization framework. By utilizing the *Poincaré* projection and the set of *Möbius* transformations (Coxeter, 1957) a "fish-eye" fovea can be positioned on the HSOM grid allowing an intuitive interaction methodology. Nodes within the fovea are displayed with high resolution, whereas the surrounding context is still visible in a coarser view. For further technical details of the HSOM see (Ritter, 1999; Ontrup & Ritter, 2001). An example for the application of the fish-eye view is given in Figure 2. It shows a navigation sequence where the focus was moved towards the highlighted region of interest. Note, that from the left to the right details in the target area get increasingly magnified, as the highlighted region occupies more and more display space. In contrast to standard zoom operations, the current surrounding context is not clipped, but remains gradually compressed at the periphery of the field of view. Since all operations

Figure 2. Navigation snapshot showing isometric transformation of HSOM tessellation. The three images were acquired while moving the focus from the center of the map to the highlighted region at the outer perimeter. Note the "fish-eye" effect: All triangles are congruent, but appear smaller as further they are away from the focus.



are continuous, the focus can be positioned in a smooth and natural way.

Browsing MEDLINE in Hyperbolic Space

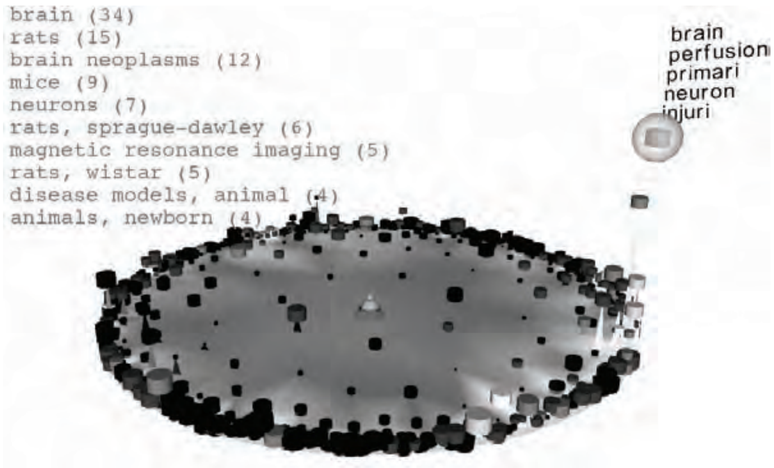
The following example presents a HSOM framework that combines the aspects of information retrieval, context enrichment and intuitive navigation into a single application. The following images show a two dimensional HSOM, where the nodes are arranged on a regular grid consisting of triangles similar to those in Figure 2. The HSOM's neurons can be regarded as "containers" holding those documents for which the corresponding neuron is "firing" with the highest rate. These containers are visualized as cylinders embedded in the hyperbolic plane. We can then use visual attributes of these containers and the plane to reflect document properties and make context information explicit:

- The container sizes reflect the number of documents situated in the corresponding neuron.
- A color scale reflects the number of documents which are labelled with terms from the *Anatomy* hierarchy of MeSH terms, i.e. the brighter a container is rendered, the more articles labelled with *Anatomy* MeSH entries reside within that node.
- The color of the ground plane reflects the average distance to neighboring nodes in the semantic bag-of-words space. This allows the identification of thematic clusters within the map.

Figure 3 shows a HSOM that was trained with approximately 25,000 abstracts from MEDLINE. The interface is split into two components: the graphical map depicted below and a user interface for formulating queries (not shown here). In the image below, the user has entered the search string "brain". Subsequent to the query submission, the system highlights all nodes belonging to abstracts containing the word "brain" by elevating their corresponding nodes. Additionally the HSOM prototype vectors—which were autonomously organized by the system during a training phase—are used to generate a key word list which annotate and semantically describe the node with the highest hit rate. To this end, the words corresponding to the five largest components of the reference vector are selected. In our example these are: *brain, perfusion, primari, neuron,* and *injuri*. In the top left the most prominent MeSH terms are displayed, that is, 34 articles are tagged with the medical subject heading "brain", 15 are tagged with "rats", and so on.

The map immediately shows that there are several clusters of brain-related articles in the

Figure 3. MEDLINE articles mapped with the HSOM. The user submitted the query string "brain" and the corresponding results are highlighted by the system.



right part of the map. Moving the focus into this area leads to the situation shown in Figure 4. By selecting single neurons with the mouse, the user can selectively retrieve key words of individual containers. Additionally, the MeSH terms of the selected node are displayed. As can be seen from the keyword distribution, all nodes share semantically related characteristics. The two leftmost labelled nodes do not contain key words like "brain" or "neuron", but the labelling with "cord", "spinal", "eeg", and "gamma", respectively, indicates a close relation to the selected "brain-area".

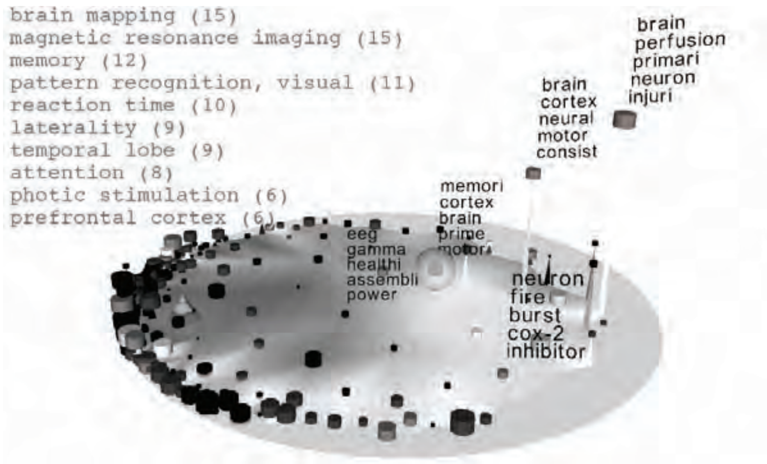
The drill-down process to the document level combines the strengths of the graphical and the textual display: After zooming-in on a promising region, the titles of documents pertaining to a particular node can be displayed by a double mouse click. Note, that due to the ordering process of the SOM algorithm these documents are typically very similar to each other—and in the sense of the clustering hypothesis therefore might all provide an answer to the same query. A further double click on a document title then displays the selected abstract on the screen. In this way, the hyperbolic self-organizing map can provide a seamless interface offering rapid overview of large document collections while allowing at the same time the drill down to single texts.

BENEFITS FOR A HOSPITAL INFORMATION SYSTEM

In the framework of a hospital information system, such a context enriched interactive document retrieval system might be of great benefit to the clinical personal. According to Chu and Cesnik (2001) a hospital may generate up to five terabytes of data a year, where 20 to 30 percent of such data are stored as free text reports such as medical history, assessment and progress notes, surgical reports or discharge letters. Free text fields in medical records are considered of invaluable importance for medical relevance (Stein, Nadkarni, Erdos, & Miller, 2000). By using document standards like the CDA as briefly discussed in Section 2.2, the clinician is able to browse and retrieve the data according to the structure provided by the document framework. However, many existing information systems do not yet take advantage of the complex CDA or the even more substantial Reference Information Model—because it is a very time consuming task to transfer documents to the new standards and to fit the enormous mass of model objects to free text data.

Nevertheless, these systems store valuable information—just in a format which cannot be

Figure 4. The user has moved the focus into the target area and can now inspect the context enriched data more closely.



directly analyzed by structured methods. In order to provide access by content, the aforementioned Information Retrieval methodologies can significantly contribute to facilitate these tasks: a self-organizing map creates structure and context purely on the statistical word distributions of free text and thus allows a semantic browsing of clinical documents. This in turn offers the possibility to gain a problem orientated perspective on health records stored in information systems, that is, as Lovis et al (2000) have noted: "intelligent browsing of documents, together with natural language emerging techniques, are regarded as key points, as it appears to be the only pertinent way to link internal knowledge of the patient to general knowledge in medicine."

CONCLUSION

Recent advances in hospital information systems have shown that the creation of organization-wide clinical data repositories (CDR) plays a key role for the support of knowledge acquisition from raw data. Currently, many case studies are published which emphasize that data mining techniques provide efficient means to detect significant patterns in medical records. Interactive OLAP tools support best practice and allow for informed real time decision-making and therefore build the solid foundation for further improvements in patient care.

The next step to further broaden the knowledge base in healthcare environments will be the integration of unstructured data, predominantly free narrative text documents. The nature of this type of data will demand for new approaches to deal with the inherently vague information contained in these documents. We have shown how artificial neural networks with their ability for self-organization of non-crisp data provide an almost natural link between the hard computational world and the soft-computing of the human brain. We believe that interactive information retrieval methodologies that account for the rich context of natural language will significantly contribute to knowledge acquisition through text analysis.

REFERENCES

Arnrich, B., Walter, J., Albert, A., Ennker, J. & Ritter, H. (2004). Data mart based research in heart surgery: Challenges and benefit. *Proceedings of the MEDINFO 2004*, San Francisco.

Aronson, A.R., Bodenreider, O., Chang, H.F., Humphrey, S.M., Mork, J.G., Nelson, S.J., Rindflesh, T.C. & Wilbur, W.J. (2000). The NLM indexing initiative. *Proceedings of the Annual AMIA Symposium* (pp. 17-21).

Baeza-Yates, R. & Ribeiro-Neto, B. (1999). *Modern information retrieval*. New York: ACM Press/Addison-Wesley.

Bergeron, B.P. (2002). Enterprise digital assistants: The progression of wireless clinical computing. *Journal of Medical Practice Management, 17*(5), 229-233.

Berners-Lee, T., Hendler, J. & Lassila, O. (2001). The Semantic Web. *Scientific American* (May), 35-43.

Bito, Y., Kero, R., Matsuo, H., Shintani, Y. & Silver, M. (2001). Interactively visualizing data warehouses. *Journal of Healthcare Information Management, 15*(2), 133-142.

Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Proceedings of the 7th International WWW Conference*, Brisbane, Australia.

Budzik, J., Hammond, K.J. & Birnbaum, L. (2001). Information access in context. *Knowledge-Based Systems, 14*(1-2), 37-53.

Chu, S. & Cesnik, B. (2001). Knowledge representation and retrieval using conceptual graphs and free text document self-organisation techniques. *International Journal of Medical Informatics, 62*(2-3), 121-133.

Coxeter, H.S.M. (1957). *Non Euclidean geometry*. Toronto: University of Toronto Press.

Doig, G.S. & Simpson, F. (2003). Efficient literature searching: A core skill for the practice of evidence-based medicine. *Intensive Care Medicine, 29*(12), 2109-11.

Dolin, R.H., Alschuler, L., Beebe, C., Biron, P.V., Boyer, S.L., Essin, D., Kimber, E., Lincoln, T. & Mattison, J.E. (2001). The HL7 clinical document architecture. *Journal of the American Medical Informatics Association, 8*(6), 552-69.

Fabrikant, S.I. & Buttenfield, B.P. (2001). Formalizing semantic spaces for information access. *Annals of the Association of American Geographers, 91*, 263-280.

Finkelstein, L. & Gabrilovich, E. (2002). Placing search in context: the concept revisited. *ACM Transactions on Information Systems, 20*(1), 116-131.

Fisk, J.M., Mutalik, P.G., Levin, F.W., Erdos, J., Taylor, C. & Nadkarni, P.M. (2003). Integrating query of relational and textual data in clinical databases: A case study. *Journal of the American Medical Informatics Association, 10*(1), 21-38.

Funk, M.E., Reid, C.A. & McGoogan, L.S. (1983). Indexing consistency in MEDLINE. *Bulletin of the Medical Library Association, 2*(71), 176-183.

Hearst, M.A. (1999). User interfaces and visualization. In R. Beaza-Yates & B. Ribeiro-Neto. *Modern information retrieval*. New York: ACM Press/Addison-Wesley.

Hearst, M.A. & Pedersen, J.O. (1996). Reexamining the cluster hypothesis: scatter/gather on retrieval results. *Proceedings of 19th ACM International Conference on Research and Development in Information Retrieval* (pp. 76-84).

Humphrey, S.M. (1992). Indexing biomedical documents: From thesaural to knowledge-based retrieval systems. *Artificial Intelligence in Medicine, 4*(5), 343-371.

Institute of Medicine. (1991). *The computer-based patient record: An essential technology for healthcare*. Washington, DC: National Academy Press.

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. *Proceedings of 10th European Conference on Machine Learning (ECML)* (pp. 137-142).

Johnson, J.D. (2003). On contexts of information seeking. *Information Processing and Management, 39*(5), 735-760.

Kashyap, V. (2003). The UMLS semantic network and the Semantic Web. *Annual Conference of the American Medical Informatics Association (AMIA)* (pp. 351-355).

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics, 43,* 59-69.

Kohonen, T., Kaski, S. & Lagus, K. (2000). Organization of a massive document collection. *IEEE Transactions on Neural Networks. Special Issue on Neural Networks for Data Mining and Knowledge Discovery,* 574-585.

Lamping, J. & Rao, R. (1994). Laying out and visualizing large trees using a hyperbolic space. *ACM Symposium on User Interface Software and Technology,* 13-14.

Lawrence, S. (2000). Context in Web search. *IEEE Data Engineering Bulletin, 23*(2), 25-32.

Lawrence, S., Giles, C.L. & Bollacker, K. (1999). Digital libraries and autonomous citation indexing. *IEEE Computer, 32*(6), 67-71.

Le, D.X. & Thoma, G.R. (2003). Automated document labeling for Web-based online medical journals. *Proceedings of the 7th World Multiconference on Systemics, Cybernetics and Informatics,* July (pp. 411-415).

Leuski, A. (2001). Evaluating document clustering for interactive information retrieval. *Conference on Information and Knowledge Management (CIKM)* (pp. 33-40).

Libbus, B. & Rindflesch, T.C. (2002). NLP-based information extraction for managing the molecular biology literature. *Annual Conference of the American Medical Informatics Association (AMIA),* 445-449.

Lodhi, H., Shawe-Taylor, J., Christianini, N. & Watkins, C. (2001). Text classification using string kernels. *Advances in Neural Information Processing Systems (NIPS), 13,* 563-569.

Lovis, C., Baud, R.H., & Planche, P. (2000). Power of expression in the electronic patient record: Structured data or narrative text? *International Journal of Medical Informatics,* 101-110.

Lowe, H.J. & Barnett, G.O. (1994). Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. *Journal of the American Medical Association, 271*(4), 1103-1108.

Mack, R. & Hehenberger, M. (2002). Text-based knowledge discovery: search and mining of life-sciences documents. *Drug Discovery Today, 7*(11), 89-98.

Magnus, W. (1974). *Noneuclidean tesselations and their groups.* London: Academic Press.

McCray, A.T. & Ide, N.C. (2000). Design and implementation of a national clinical trials registry. *Journal of the American Medical Informatics Association, 7,* 313-23.

McDonald, C. (1997). The barriers to electronic medical record systems and how to overcome them. *Journal of the American Medical Informatics Association, 4,* 213-221.

Munzner, T. (1998). Exploring large graphs in 3D hyperbolic space. *IEEE Computer Graphics and Applications, 18*(4), 18-23.

Ng, S.K. & Wong, M. (1999). Towards routine automatic pathway discovery from on-line scientific text abstracts. *Genome Informatics, 10,* 104-112.

Ontrup, J. & Ritter, H. (2001). Hyperbolic self-organizing maps for semantic navigation. *Advances in Neural Information Processing Systems (NIPS), 14,* 1417-1424.

Pirolli, P., Card, S.K. & van der Wege, M.M. (2001). Visual information foraging in a focus + context visualization. *ACM Conference on Human Factors in Computing Systems* (pp. 506-513).

Ritter, H. (1999). Self-organizing maps in non-euclidean spaces. In E. Oja & S. Kaski (Eds.), *Kohonen maps* (pp. 97-110). Amer Elsevir.

Rzhetsky, A., Iossifov, I., Koike, T., Krauthammer, M., Kra, P., Morris, M., Yu, H., Duboue, P.A., Weng, W., Wilbur, W.J., Hatzivassiloglou, V. & Friedman, C. (2004). GeneWays: A system for extracting, analyzing, visualizing, and integrating molecular pathway data. *Journal of Biomedical Informatics, 37*(1), 43-53.

Salton, G. & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing and Management, 24*(5), 513-523.

Sensmeier, J. (2003). Advancing the state of data integration in healthcare. *Journal of Healthcare Information Management, 17*(4), 58-61.

Silver, M., Sakata, T., Su, H.C., Herman, C., Dolins, S.B. & O'Shea, M.J. (2001). Case study: How to apply data mining techniques in a healthcare data warehouse. *Journal of Healthcare Information Management, 15*(2), 155-164.

Skupin, A. & Fabrikant, S.I. (2003). Spatialization methods: A cartographic research agenda for non-geographic information visualization. *Cartography and Geographic Information Science, 30*(2), 95-119.

Smith, A. & Nelson, M. (1999). Data warehouses and clinical data repositories. In M.J. Ball, J.V. Douglas & D. Garets (Eds.). *Strategies and technologies for healthcare information* (pp. 17-31). New York: Springer Verlag.

Spyrou, S.S., Bamidis, P., Chouvarda, I., Gogou, G., Tryfon, S.M. & Maglaveras, N. (2002). Healthcare information standards. *Health Informatics Journal, 8*(1), 14-19.

Srinivasan, P. & Rindflesch, T.C. (2002). Exploring text mining from MEDLINE. *Annual Conference of the American Medical Informatics Association (AMIA).*

Stein, H.D., Nadkarni, P., Erdos, J. & Miller, P.L. (2000). Exploring the degree of concordance of coded and textual data in answering clinical queries from a clinical data repository. *Journal of the American Medical Informatics Association, 7,* 42-54.

Swanson, D.R., Smalheiser. N.R. & Sleeman, D.H. (1997). An interactive system for finding complementary literatures: A stimulus to scientific discovery. *Artificial Intelligence, 91*(2), 183-203.

Tange, H.J, Hasman, A., de Vries, R., Pieter, F. & Schouten, H. C. (1997). Medical narratives in electronic medical records. *International Journal of Medical Informatics, 46*(1), 7-29.

Van Rijsbergen, C.M. (1979). *Information retrieval*. London: Butterworths.

Wise, J. (1999). The ecological approach to text visualization. *Journal of the American Society for Information Science, 50*(13), 1224-1233.

Yang, Y. (1999). An evaluation of statistical approaches to text categorization. *Information Retrieval, 1*(2), 69-90.

Zamir, O. & Etzioni, O. (1999). Grouper: A dynamic clustering interface to Web search results. *Computer Networks, 31,* 1361-1374.

ENDNOTES

1 *http://www.microsoft.com/resources/casestudies/CaseStudy.asp? CaseStudy-ID=14967*

2 *http://vivisimo.com*

This work was previously published in Clinical Knowledge Management: Opportunities and Challenges, edited by B. Bali, pp. 52-71, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 1.22
Key Performance Indicators and Information Flow:
The Cornerstones of Effective Knowledge Management for Managed Care

Alexander Berler
National Technical University of Athens, Greece

Sotiris Pavlopoulos
National Technical University of Athens, Greece

Dimitris Koutsouris
National Technical University of Athens, Greece

ABSTRACT

It is paradoxical that, although several major technological discoveries such as Magnetic Resonance Imaging and Nuclear Medicine and Digital Radiology, which facilitate improvement in patient care, have been satisfactorily embraced by the medical community, this has not been the case with Healthcare Informatics. Thus, it can be argued that issues such as Data Management, Data Modeling, and Knowledge Management have a long way to go before reaching the maturity level that other technologies have achieved in the medical sector. This chapter proposes to explore trends and best practices regarding knowledge management from the viewpoint of performance management, based upon the use of Key Performance Indicators in healthcare systems. By assessing both balanced scorecards and quality assurance techniques in healthcare, it is possible to foresee an electronic healthcare record centered approach which drives information flow at all levels of the day-to-day process of delivering effective and managed care, and which finally moves towards information assessment and knowledge discovery.

INTRODUCTION

The advantages of the introduction of Information and Communication Technologies (ICT) in the complex Healthcare sector have already been depicted and analyzed in the Healthcare Informatics bibliography (Eder, 2000; Englebardt & Nelson, 2002; Harmoni, 2002; Norris, Fuller, Goldberg, & Tarczy-Hornoch, 2002; Shortliffe, Perreault, Wiederhold, & Fagan, 2001; Stegwee & Spil, 2001). It is nevertheless paradoxical that, although several major technological discoveries such as Magnetic Resonance Imaging, Nuclear Medicine and Digital Radiology, which facilitate improvement in patient care, have been satisfactorily embraced by the medical community, this has not been the case with Healthcare Informatics. Thus, it can be argued that issues such as Data Management, Data Modeling, and Knowledge Management have a long way to go before reaching the maturity level that other technologies have achieved in the medical sector.

A variety of reasons could be proposed for this issue, though with a short analysis it becomes rather clear that modern ICT present integration problems within the healthcare sector because of the way the latter is organized. Healthcare is a strongly people-centered sector in which ICT has been considered more as an intruder, as a "spy" to the healthcare professionals' way of doing things and as a competitor to this people-centered model. Thus, if ICT intend to prove its advantages towards establishing an information society, or even more a knowledge society, it has to focus on providing service-oriented solutions. In other words, it has to focus on people and this has not been the case in most of the circumstances. It is common knowledge that in order to install any type of information system in healthcare, especially if it involves knowledge management, six main groups of issues have to be dealt with (Iakovidis, 1998, 2000):

1. The organizational and cultural matters related to healthcare. This issue is rather important, regardless of any information system, since organizational models and culture do not endorse the continuity of care or any type of structured data collection. Issues such as mistrust between different specialists, between the different healthcare structures or between doctors and nurses prevent in many cases the effective sharing of information. Health reforms are currently under way in many countries stressing the will to deal with this problem.
2. The technological gap between healthcare professionals and information science experts. Doctors are often reluctant to use information systems that they believe are not designed for them. From another point of view, Healthcare Informatics have been introduced in healthcare institutions mostly on pilot-based projects aiming at addressing specific issues and have proposed solutions addressing a small number of healthcare practitioners, resulting in establishing a complex map of information niches. This approach is the consequence of applying information technology to procedures that where not designed for it, thus creating a panspermia of information models which are neither compatible nor interoperable, even within a single institution's environment. Efforts in creating interoperability standards and protocols such as HL7 are proposing solutions to address this issue, thus enabling data manipulation and knowledge management.
3. The legal requirements on the confidentiality of personal and patient related data and on data privacy. It is clear that if this issue is not addressed at a managerial and procedural level by imposing suitable policies to meet these requirements, there is little chance that medical data will be kept digitally in a structured manner (thus allowing the transition from digital islands of clinical data towards a structured electronic healthcare record). The implementation of an information system, where the electronic healthcare record is considered to be the core of the system (patient-centered model), is the only way

to drive data management towards creating new knowledge. The complexity of the problem can be explained if one just observes the course of implementation of both the Health Information Privacy and Accountability Act (HIPAA) in the US and Directive 95/46/EC in the EU. The issues seem to have been dealt with at the strategic level, but still a lot has to be done in the implementation and setup of those strategies.

4. The industrial and market position of Healthcare Informatics. In general, the healthcare market is seen by the industry as large in size but not highly profitable, mainly due to the lack of standards in implementing and interoperating healthcare informatics products. As a consequence, the industry has focused on creating mostly small-scale products (i.e., Laboratory Information Systems, Radiology Information Systems, Clinical Information Systems) and not on evangelizing the production of information system that are dealing with healthcare as a whole. The lack of end-to-end solutions is dealt with by interconnecting heterogeneous information systems (a rather complex task with constant change management issues) and by introducing solutions from other business sectors (i.e., ERP, SCM, CRM) that have often been rejected by "key users" as non-compliant with their job description. Nevertheless, the new Web technology approaches (Web services, XML, etc.) and the new information technology strategies (i.e., service oriented architecture) could be the drivers towards merging information technology and healthcare services and thus enabling the establishment of knowledge management products.
5. The lack of vision and leadership of healthcare managers and health authorities, and the lack of willingness to re-engineer healthcare processes for the benefits of efficiency and quality of care delivery. Some countries are in the process of introducing or implementing such Business Process Reengineering projects in order to address healthcare delivery in a more information flow conformant way. This is a key point in reaching knowledge management, knowledge re-use and sharing, and finally proposing a solution for the knowledge-based society of tomorrow. This issue should be dealt with by proposing strategies that focus on processes and by establishing key performance indicators, balanced scorecards, or other metrics that are the upper level of a structured information flow-based model.
6. User acceptability and usability of the proposed information systems. This issue is the one most strongly related to the problem of dealing with the people-centered approach of the healthcare sector. This issue deals with information systems' user friendliness, with usability issues such as the time to reach a data entry point, the speed of information retrieval, the quality of information retrieval, the complex security procedures, and so on. In order to implement information systems and knowledge management systems, education and training must be addressed with high priority since user acceptability is strongly related to them. Service oriented models and patient-centered information systems have a higher chance of passing the user acceptability test. A system that is not accepted by the user is often a system with poor data quality (or no data at all) and knowledge management, business intelligence or data warehousing solutions are consequently inoperable and unsuccessful.

Taking all of the above issues into consideration, this chapter proposes to explore trends and best practices regarding knowledge management from the viewpoint of performance management, based upon the use of Key Performance Indicators (KPI) in healthcare systems. By assessing both balanced scorecards (Kaplan/Norton) and quality assurance techniques in healthcare (Donabedian), it is possible to foresee an electronic healthcare record centered approach which drives information flow at all levels of the day-to-day process of delivering effective and managed care, and which finally moves towards information assessment and knowledge discovery (both with administrative and medical data). KPIs should be regarded as the strategic assessment tool, for both the executives and the clinical decision-makers, that will lead healthcare delivery to excellence and to knowledge discovery and assessment.

BACKGROUND

Knowledge Management as a Transformation Driver in Healthcare

Today, Knowledge Management (KM) is on everyone's mind. Healthcare organizations are no exception and are accepting the challenge to more effectively share knowledge both internally and externally (Strawser, 2000). The growth of KM projects (i.e., decision support systems, data mining tools, business intelligence solutions) signals a growing conviction that managing institutional knowledge is crucial to business success and possibly business survival. When the hype and confusion are stripped away, it is apparent that KM initiatives can profoundly change a healthcare enterprise for the better, and bring numerous advantages to Healthcare Information Management (HIM) professionals. For HIM professionals, KM is worthy of special attention because it informs them not only on how to do things, but also on how they might do them better. In order for this to happen, data should be provided in specific patterns and should be based upon a strategy that will empower a healthcare system by gaining knowledge of its processes, its outcomes, and its structures.

Despite the obvious advantages, many healthcare decision makers view the idea of a KM initiative with scepticism, possibly because of an incomplete or incorrect understanding of the tools needed to achieve it. Many of the tools and strategies associated with implementing KM are not new; what is new is a cohesive approach to KM design and implementation. Certainly there are pitfalls and limitations in using information technology for KM—trying to force fluid knowledge into rigid data structures, for example, or focusing too much on the tools and not enough on the content. But networks and computers, with their ability to connect people and store and retrieve virtually unlimited amounts of information, can dramatically improve departmental efficiencies. Some examples of knowledge management applications are listed below:

- **Data Mining tools** enable decision makers to search and analyze large sets of data by using specific querying methods and tools (Standard Query Language, Rough Data set, On Line Analytical Processing).
- **Document and Content Management systems** are widely used to store and archive files of any type (text, images, video, etc.) and correlated them with keywords that have a business meaning to the end user.
- **Knowledge Maps** are graphical or other representations of how and by whom a specific set of information is created, distributed and assessed. Knowledge Maps are very important tools in Total Quality Management projects.
- **Intelligent Agents** use a combination of profiling techniques, search tools, and recognition algorithms to provide up to date specific information to the end user. For example, intelligent agents could be used to forward completed test results to the corresponding physicians of a patient.
- **Web Browsers** are the most commonly used tools for searching information in an intranet

or the Internet. As such, Web browsers are increasingly becoming the most common graphical user interface, even for specific software products such as financial accounting and patient order entry systems.

- **Business Intelligence tools and Data Warehouses** enable the decision maker to have predefined access to specific information of interest regardless of the physical location of the data. Such systems are ideal for performance management and executive reporting and serve as the technological base for supporting the idea of a digital dashboard of indicators.
- **Workflow applications** play a very important role in KM since knowledge is created during the process-based operations that take place in a healthcare institution. A computerised patient order entry system is a classic example of a process-based operation in healthcare that requires the constant monitoring of the workflow status.
- **E-learning and collaboration tools** are part of the knowledge distribution process, which is extremely important in healthcare, since continuous education is a key factor in effective practice of care.

The essence of effective knowledge management does not rely on the use of one or more existing or forthcoming information technology tools. It is mostly about people, about processes and about capturing the results of people following processes, about transforming information into knowledge (explicit or tacit) and reusing it within a healthcare framework.

Performance Management: Monitor and Manage Healthcare

In order to persuade healthcare decision makers to assess the added value of KM tools, the latter should initially be used to propose new performance measurement and performance management techniques at all levels of a healthcare system (Hurst & Jee-Hughes, 2001). In that sense, performance management has long been considered as a tool for controlling spending and for increasing the efficiency of healthcare systems (Oxley & MacFarlan, 1994). There are three broad goals that governments generally pursue in the healthcare area:

- Equity: where citizens should have access to some incompressible minimum level of healthcare and treatment based on the need for care rather that solely on income.
- Micro-economic efficiency: where quality of care and consumer satisfaction should be maximized at minimum cost.
- Macroeconomic cost control: where the healthcare sector should consume an "appropriate" share of GDP.

In addition, healthcare systems are often facing factors that put pressure on the system. As a consequence, an effective performance management framework is the only solution towards controlling factors such as:

- Population aging
- Increased income and higher demand for healthcare services;
- Increased access to healthcare services; and
- Increase of high technology usage which in turn increases the healthcare services usage creating sometime unnecessary demand (from a medical point of view).

Most existing policies for controlling the performance of healthcare systems were based upon financial assessment of past results (macroeconomic control of spending and micro-efficiency improvements) by giving incentives to payers and providers of a healthcare system. By its very nature the financial measurements are not forward looking and are exclusionary to non-financial measures. In addition, the emergence of the Information Society in the late 90s rendered many of the fundamental performance management assumptions obsolete. Information Society has brought a new set of assumptions that institutions have to include into their strategy. Amongst others we could refer to:

- The cross functional aspects of processes based upon specialisation, increased skills and high technology
- The integration of processes in the healthcare sector from the suppliers to the patient and the ability to manage and monitor materials upon requests and needs
- The ability to offer specific services to patients in accordance to their needs while being able to constraint costs of this customised care. In fact this is the essence of managed care: providing satisfactory and high quality of care at a reasonable cost.
- The global scale of healthcare: This sector is no exception to any other global marketplace, thus making healthcare delivery more comparable at a national or regional level and within accepted standards (e.g., clinical procedure guidelines)
- Innovation, which has been a key driver towards quality of care and quality of life for many years in the healthcare sector
- Knowledge workers: the increasing complexity of medicine and technology has created a need for highly skilled personnel at every level of a healthcare institution. Employee empowerment is driven by knowledge as it is created in the daily process of delivering healthcare services.

Nowadays, performance measurement is moving towards the adoption of a set of objectives for a healthcare system. To our knowledge, there is no complete agreement on what is meant by "performance" of health systems and many sets of objectives are generally proposed. The use of Key Performance Indicators (KPI) helps to establish this set of objectives more thoroughly by focusing on the real needs.

Using Balanced Scorecards

In order to address the issue of creating a set of KPIs , the Balanced Scorecard (BSC) framework (Kaplan, 2001; Norton & Kaplan, 1996) initially proposed by Kaplan and Norton for the strategic management of financial organizations in the mid 1990s, is one of the most suited approaches in that direction. This model has now proven its value and since it is a generic framework it is applicable to the healthcare sector. The BSC concept involves creating a set of measurements and objectives for four strategic perspectives:

- Financial: financial performance measures typically indicate whether a proposed strategy implementation and execution is contributing to bottom line improvements, based upon an accurate summary of economic consequences of actions already taken.
- Customer (i.e., the Patient): this perspective is a set of objectives that focuses on identifying the patient's needs, the targeted market (this could be the case of a specialised institution) and on measuring the performance of each specific business unit (i.e., a department) that has some influence on how the patient sees the healthcare organization.
- Internal Business Process. This perspective should gather all objectives related to processes and the way these are monitored and fine-tuned in order to achieve both excellence in financial accomplishments and patient satisfaction. In that sense, the BSC approach is a constant business process reengineering process based upon specific goals to meet and not on improving de facto established processes.
- Learning and Growth (innovation and vision). This is an important perspective of

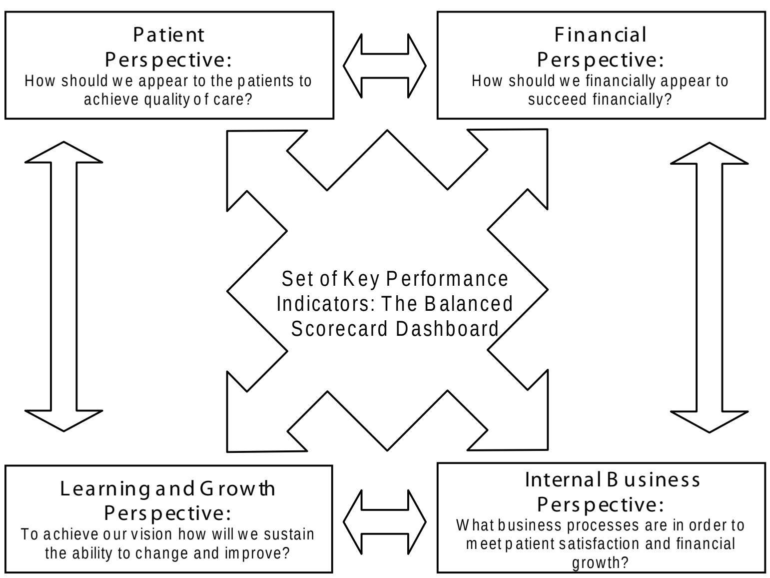
a BSC implementation because it focuses on the objectives and goals to achieve the incorporation of business innovation (e.g., installing a Positron Emission Tomography Device) and the continuous education of medical and administrative staff. In that sense, this perspective identifies measurable tasks in order to build long-term growth and improvement. In the healthcare sector, improvement is also measured by assessing the outcome of treatment and care.

These four perspectives comprise a framework, meaning that they must be assessed and populated in accordance to each business case. In order to achieve this, one has to set for each category a list of objectives that can be feasibly measured. To each objective, a specific target should be set and the initiative to reach that target well described.

The BSC approach retains the "traditional" financial perspective, which mostly focuses on reporting about past events. Financial figures, even if prospected in the future by statistical means, are inadequate for guiding and evaluating the journey that healthcare institutions must make to create future value through innovation, knowledge, patient satisfaction, employee empowerment, process refinement, technology assessment and material management. In that sense, the three other cited perspectives are completing the puzzle in order to create a more valid and future-oriented performance management strategy. In recent years, there have been some implementations of BSC frameworks in the healthcare sector mostly on a single institution base (i.e., a hospital, a clinic, an information technology component). Readers are encouraged to read and assess the best practices, cited as references to this chapter (Aidermark, 2001; Castaneda-Mendez, Mangan, & Lavery, 1998; Forgione, 1997; Freeman, 2002; Gordon & Geiger, 1999; Inamdar, Kaplan, & Bower, 2002; Oliveira J, 2001; Pink, Mc Killop, Schraa, Preyra, Montgomery, & Baker, 2001; Tarantino, 2003; Weber, 1999)

The model in Figure 1 enables decision makers to also value how the latter three perspectives have worked and thus it enables the measurement of intangible aspects of a system. The measurements focus on a single strategy that can be easily broken

Figure 1. A healthcare oriented BSC framework



down to the various levels of depth, depending of the organization type. Thus, by using BSC one can create a top-to-bottom design of a system, starting up from the needed strategy (what is the market, who are the customers, what are the critical processes, what is required by the stakeholders) going down to design metrics, processes, structures and finally the needed technical and functional specification to create an information system and a knowledge management framework capable of producing the right data and serve the strategy.

QUALITY ASSURANCE IN HEALTHCARE AND THE ROLE OF INFORMATION TECHNOLOGY

Known Problems and Issues of the Healthcare Sector

It is rather of common knowledge that the healthcare sector is not a sector without problems. Especially in Europe, the great majority of healthcare organizations are state funded. This means that institutions have in some cases very restricted budgets to satisfy their needs thus making performance management a critical issue in their daily routine. Other common problems are increased bed coverage (some institutions are almost at 100 percent), long length of stay, increased waiting lists, poor facilities, and so on. Taking all this into account and adding the life-related processes, one can clearly see that healthcare organizations are difficult to manage even for experienced managers of other sectors. One could list some of the major issues that healthcare stakeholders are confronted with:

- Diversity of cases
- Need of high technology medical devices
- Public Policy Restrictions (e.g., payment by per day quotas that do not cover the inpatient treatment costs)
- High lengths of stay
- Increased waiting times
- Obsolete facilities
- Restricted funds for training
- Restricted funds for maintenance
- Increased number of medical errors (sometimes fatal)
- Erroneous drug prescription and intake (sometime fatal)
- Geographical issues (too many patients, too few nurses and medical staff)
- Large lists of materials to be managed (more than 3000 on average)
- Excessive waste management (10 m^3 per day on average)

Quality Assurance and Performance Management

The application of a BSC framework will not by itself solve any of the aforementioned problems and issues. One could even say that BSC projects often fail as a consequence of misunderstanding or of not using a BSC strategy. BSC projects also fail because the variables of the scorecard are incorrectly identified as the primary drivers and because the improvement goals (the targeted objectives) are negotiated across an institution instead of being based upon stakeholders' requirements, fundamental processes and improvement process capabilities. They also fail because there is no deployment system installed to disseminate, maintain and promote the BSC framework, or because some very important KPIs are not used, or metrics are poorly defined.

Furthermore in order to create a structure that can be monitored the KPIs should not be more that 10 for each perspective, while non financial metrics should overcome the financial metrics by approximately six to one (Schneiderman, 1999). In order to support any BSC framework, a deployment and maintenance system based upon quality assurance specially designed for Healthcare should be established. Traditionally in healthcare, Quality Assurance (QA) has been meant to apply predominantly to healthcare itself as provided directly to patients by legitimate healthcare practitioners. We also include other services that directly affect the ability of practitioners to perform well, meaning services such as radiology, pharmaceutical, laboratory, and patient admission. The basic quality assurance terms (Donabedian, 2003) are:

- Efficacy: the ability of the science and technology of healthcare to bring about improvements in health when used under the most favorable circumstances.
- Effectiveness: the degree to which attainable improvements in health are in fact attained.
- Efficiency: the ability to lower the cost of care without diminishing attainable improvements in health.
- Optimality: the balancing of improvements in health against the costs of such improvements.
- Acceptability: conformity to wishes, desires, and expectations of patients and their families.
- Legitimacy: Conformity to social preferences as expressed in ethical principles, values, norms, mores, laws, and regulations.
- Equity: Conformity to a principle that determines what is just and fair in the distribution of healthcare and its benefits among the members of the population.

One can clearly see the benefit of applying quality assurance components in the development of a BSC strategy. A BSC framework that meets quality assurance in healthcare is most probable that will meet patient needs, practitioners' feelings, patient-practitioner relationship, amenities of care (e.g., confidentiality, privacy, comfort, cleanliness, convenience), as well as financial and organizational aspects required.

The most popular quality assurance model in healthcare is based, to the best of our knowledge, upon the Donadebian approach (Donabedian, 1980, 1982, 1985, 1993, 2003) where a healthcare organization (i.e., a hospital) is a system formed by the interaction of structures, processes, and outcomes. Structures are used to establish processes in order to create healthcare outcomes that have an effect on structures that need to change or adjust processes to meet the required outcomes. Strongly believing that healthcare outcomes are more important than financial outcomes in a healthcare system, we are confronted with a model where intangible assets are more important than tangible assets. This last statement makes a healthcare system very difficult to manage and a straightforward strategy hard to define.

The Need for a Specific Implementation Plan

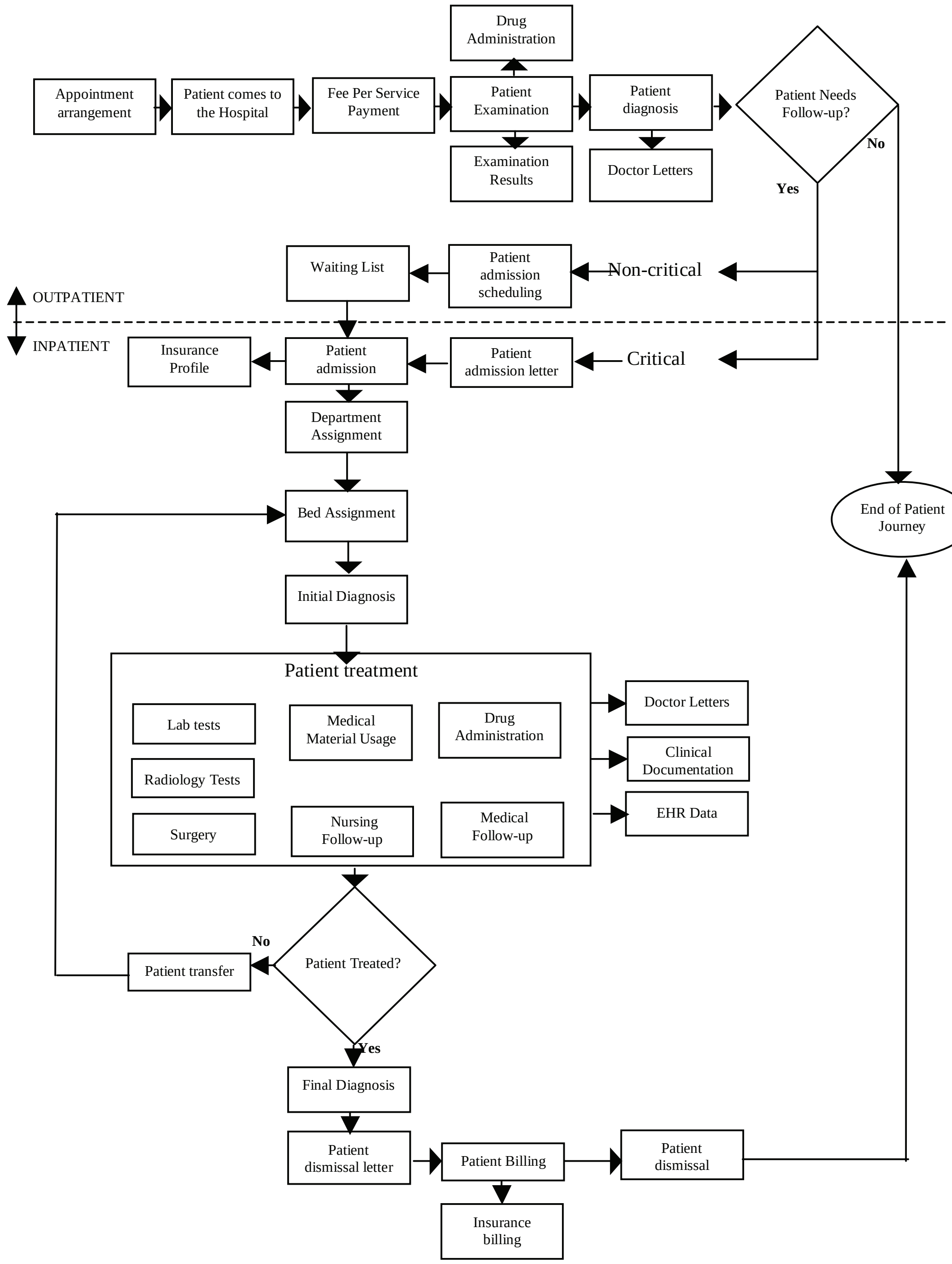
In order to implement a viable performance management strategy (i.e., a BSC framework) the steps one needs to take include:

- Determining what to monitor
- Determining priorities in monitoring
- Selecting an approach to assessing performance
- Formulating criteria and standards (i.e., Key Performance Indicators)
- Obtaining the necessary information
- Determining how and when to monitor
- Constructing a monitoring system
- Managing changes and improvements

Building a Patient-Centered Healthcare Model

Finally, the proposed KPIs are directly or indirectly driven from healthcare processes. As an example, we propose to analyze a standard patient journey of a citizen in a healthcare institution. Figure 2 shows how the patient journey for a hospital is

Figure 2. The patient journey centered model for a hospital institution



conceived (the patient journey in a primary care setting may be simpler). This workflow is the heart of the healthcare system and a prerequisite for any patient-centered information system to properly manage the information flow. This workflow is nowadays extended to include new processes such as emergency pre-hospital care and home care monitoring in order to create the hospital without walls of the 21st century.

Based on Figure 2 one can create a table where quality assurance and balanced scorecard features are confronted, analyzed, assessed and finally set. Table 1 is an example (non exhaustive) of the initial process.

From the above, it becomes apparent that the design and proposal of KPIs is not an easy task. KPI selection can vary upon specific measurement needs, upon goal set, and so on. In order to manage and validate the proposed KPIs by each BSC strategy, a set of KPI dashboards for each management entity, department or any other region of interest should be created.

USE CASE: REGIONAL HEALTHCARE AUTHORITIES IN GREECE

In 2001, a reform of the Greek National Healthcare System was introduced in order to enhance the performance and control of healthcare provision in Greece (Greek National Healthcare System Reform Act, 2001; Vagelatos & Sarivougioukas, 2003). One of the main changes was the division of the country in 17 autonomous healthcare regions where the Regional Healthcare Authority (RHA) is responsible for the regional healthcare strategy. This reform introduced the need to establish a three-level decision-making and performance management mechanism as described in Figure 3.

The proposed methodology was used to reach an initial set of KPIs by assessing existing knowledge and future needs (Decision No 1400/97/EC, 1997; McKee & Healy, 2002; Polyzos, 1999). Those KPIs were processed especially to serve the new strategy introduced in Greece.

As described in Figure 3, the regional healthcare system is comprised of a series of information systems covering the whole structures existing at any level, the processes required to meet the administrative and medical needs and finally, the outcomes that must come out from the implementation of such a complex interpolation of informatics infrastructure. The above information model was introduced to establish a community of networked healthcare organizations (hospitals, primary care) that are interoperating in order to support and implement the new healthcare strategy: to provide integrated and high quality healthcare services to the citizens based upon equal access to the resources (Information Society SA, 2003). In order to achieve this goal, two main issues were raised:

- How and when will information systems interoperate?
- What is the minimum required dataset to achieve the proposed strategy?

The first issue can be answered by using standards and protocols such as HL7 to meet with interoperability issues in healthcare (Spyrou, Berler & Bamidis, 2003). The second issue is partially addressed by the proposed initial set of KPIs presented below. The proposed KPIs are forming a complete set of metrics that enable the performance management of a regional healthcare system. In addition, the performance framework established is technically applied by the use of state-of-the art knowledge management tools such as data warehouses and business intelligence information systems.

The proposed KPIs are categorised into the four perspective stated by Norton & Kaplan, having also taken into account the raised quality

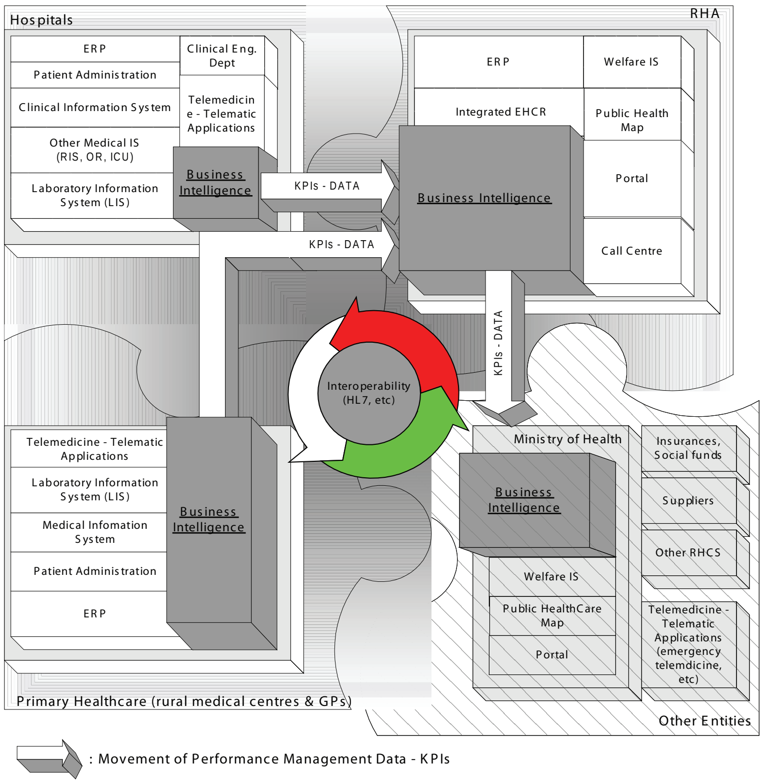
Table 1. Defining KPIs from processes, an example

Workflow process	BSC Perspective	KPI	QA approach	Source of information
Appointment arrangement	Patient	Appointments/day	Process	Hospital Information System (HIS), Scheduling S/W
Patient comes to the Hospital	Process	Number of outpatients	Process	Hospital Information System
Fee Per Service Payment	Financial	Mean cost per examination	Process	Billing, ERP S/W
Drug Administration	Process	Number of prescription/ drug	Outcome	Patient Order Entry S/W, HIS
Patient Examination	Patient	Patient satisfaction	Process	HIS, surveys
Examination Results	Growth	Number of patient with re-examination	Outcome	HIS, Electronic Healthcare Record (EHR)
Doctor Letters	Process	Number examinations/practitioner	Outcome	HIS, Electronic Healthcare Record (EHR)
Patient diagnosis	Growth	Visits/ICD codes	Outcome	HIS, Electronic Healthcare Record (EHR)
Patient Needs Follow-up?	Process	Number of inpatient from outpatient clinic	Process	HIS
Waiting List	Patient	Waiting time in days	Structure	Hospital Information System (HIS), Scheduling S/W
Patient admission scheduling	Patient	Equity of delivered care	Process	Hospital Information System (HIS), Scheduling S/W
Patient admission letter	Process	Number of emergency cases/day	Process	HIS
Patient admission	Process	Number of inpatients	Process	HIS
Insurance Profile	Financial	Net cash flow per insurance company	Process	HIS, ERP
Department Assignment	Financial	Mean operational cost per department	Structure	HIS
Bed Assignment	Process	Bed coverage rate	Structure	HIS
Initial Diagnosis	Process	Admission per case type (ICD 10)	Outcome	HIS, Electronic Healthcare Record (EHR)
Lab tests	Process	Mean value of lab test per doctor, per patient	Outcome	Laboratory Information System (LIS)
Radiology Tests	Process	Mean value of radiology test per doctor, per patient	Outcome	Radiology Information System (RIS)
Surgery	Financial	Mean cost of surgical procedure	Structure	HIS, Electronic Healthcare Record (EHR)
Medical Material Usage	Financial	Mean cost of medical material consumption	Outcome	HIS, ERP
Nursing Follow-up	Growth	Number of Nurses per bed	Process	HIS, Electronic Healthcare Record (EHR)
Medical Follow-up	Growth	Number of practitioners per bed	Process	HIS, Electronic Healthcare Record (EHR)
Clinical Documentation	Process	Number of medical procedures per day	Outcome	HIS, Electronic Healthcare Record (EHR)
EHR Data	Patient	Number of cases with EHR	Outcome	HIS, Electronic Healthcare Record (EHR)
Patient transfer	Process	Number of patient transfers/ patient or /day	Process	HIS
Patient Treated?	Growth	Number of patients treated under a specific critical pathway	Process	HIS, Electronic Healthcare Record (EHR)
Final Diagnosis	Growth	Cases per final diagnosis	Outcome	HIS, Electronic Healthcare Record (EHR)
Patient dismissal letter	Patient	Inpatient Satisfaction	Process	HIS, surveys
Patient Billing	Financial	Mean treatment cost per day	Process	Billing, ERP S/W
Patient dismissal	Process	Mean length of stay /per dept. per ICD code	Process	HIS
Insurance billing	Financial	Return of Capital Employed (ROCE)	Process	Billing, ERP S/W

assurance issues stated earlier. The performance indicators in a regional healthcare setting are depicted in Table 2.

Each of the above KPIs is the result of analysis based upon the needs of a standard regional healthcare authority. The proposition of a set of KPIs is nevertheless not the complete solution of the problem. Implementing KPIs is a constant process based upon specific metrics that each regional healthcare authority and each department or institution under its control must periodically assess and reengineer. Assessment should be based

Figure 3. Regional healthcare information systems framework and interoperability



upon specific goals met and reengineering is often required due to administrative, demographic or other important changes that must occur. Focusing for example in the KPI marked as "mean treatment cost per day" one should notice that the KPI does not mean much without a metric. In Greece, most of inpatient treatments are based upon fixed prices per day and do not follow the pay per service model which is financially more viable.

Healthcare financials are part of a national policy aiming at providing high quality healthcare services to all citizens regardless of their income, social status or other characteristic. As a consequence, the use of the fixed price model (per day quotas) in Greece serves that purpose albeit with its advantages and disadvantages. Current treatment cost per day (for an inpatient) has been fixed to about 135 Euros and this value could be used as an initial metric. If this value is exceeded this would mean that the RHA budget will cover the difference or transfer the cost to insurance companies and social welfare. In addition, a regional healthcare authority will then have credible proof that national standards are outdated and require revision in order to support the system. As a result, this KPI has now a specific meaning linked to regional strategy and budgetary needs. Following that example, all financial KPIs are therefore an important perspective of BSC since they are the measurement of the financial viability of the regional healthcare authority. In fact, all KPIs should be associated with adequate metrics in order to be assessed, thus driving the RHA towards the right strategic decisions.

In order to meet and populate the above-mentioned KPIs a regional healthcare authority has to implement a complex information technology system in order to gather up all needed informa-

Table 2. KPIs in a regional healthcare setting

No	Financial KPI description	No	Process KPI description
F1	Mean treatment cost per day	P1	Length of stay
F2	Mean cost of medical treatment per patient	P2	Patient admission rate per medical unit
F3	Mean cost of drugs consumption	P3	Percentage of bed coverage
F4	Mean cost of radiology testing	P4	Vaccination rate
F5	Mean cost of laboratory testing	P5	Mean value of performed test per patient, per doctor
F6	Mean cost of material consumption	P6	Number of inpatients
F7	Mean cost of surgical procedure	P7	Number of outpatients
F8	Mean operational cost per dept./clinic	P8	Number of drug prescription
F9	Mean cost of vaccination procedures	P9	Number of laboratory tests
F10	Mean cost per medical examination	P10	Number of surgery procedures
F11	Return of capital employed (ROCE)	P11	Number of radiology tests
F12	Net Cash flow	P12	Number of visit in outpatient clinics
F13	Income per employee	P13	Number of visits in primary care institutions
F14	Payroll rate versus operational costs.	P14	Number of dental care processes
		P15	Number of processed emergency cases
		P16	Number of unprocessed order entries on the same day
		P17	Number of preventive care visits
		P18	Number of home care monitored patients
		P19	Assessment of patient satisfaction

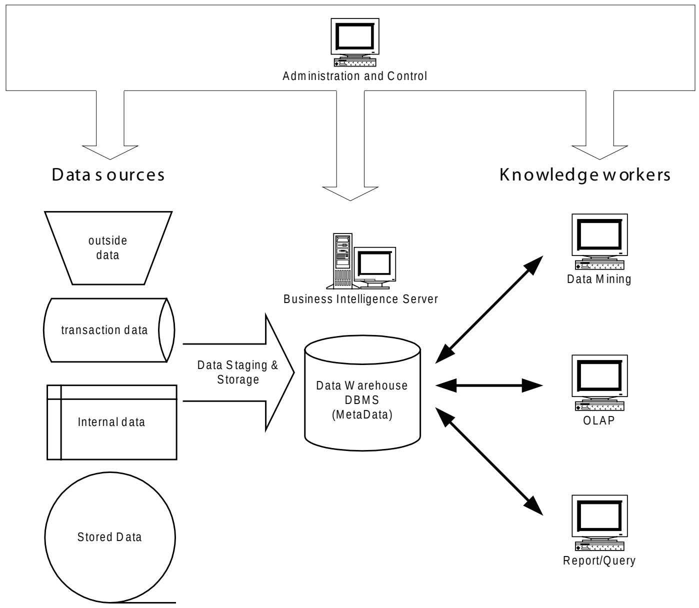
No	Customer (patient) KPI description	No	Learning and Growth KPI description
C1	Mortality rate	L1	Medical device usage growth
C2	Morbidity rate	L2	Healthcare professionals training rate
C3	Number of medical staff per 1000 inhabitants	L3	Employee Satisfaction rate
C4	Number of beds per 1000 inhabitants	L4	Number of doctors per bed
C5	Accessibility of patients to the medical units	L5	Number of nurses per bed
C6	Time in a waiting list	L6	Number of existing healthcare professionals versus expected job positions
C7	Equity of delivered care	L7	Personnel productivity rate
C8	Number of readmission per patient	L8	Number of medical interventions per doctor
C9	Mean length of stay	L9	Number of patient with re-examinations
C10	Patient Satisfaction rate	L10	Visits/ICD codes
		L11	Admissions per case type (ICD 10)
		L12	Dismissals per case type (ICD 10)

tion. Then the information collected through the use of enterprise resource planning software, hospital information systems, clinical information systems, radiology information system, and laboratory information systems has to be processed and interpolated to produce the final metadata set from which the KPIs are driven. In other words, the model is complemented by a business intelligence solution similar to the one depicted in Figure 4.

Figure 4 shows how data are collected from the various data sources, cleansed and homogenised, and finally redistributed to the knowledge workers and decision makers of the regional healthcare authority (Extraction, Transformation and Load—ETL).

The data collection process is extremely important since it is a basic feature of successfully populating the KPIs. In that sense both organizational and technological issues to achieve data quality should be considered. In the proposed setting, the regional healthcare authority has imposed on its healthcare units the use of specific classifications, codifications and taxonomies such as the 10th edition of the International Classification of diseases (ICD10). In addition, the proposed KPIs can be seen as attributes of structure, process or outcome (based upon the Donabedian approach) so that they can be used to draw an inference about quality. As such the KPIs are proposed, designed, tested and assessed by a panel of experts (executive officers of the RHA, practitioners). By implementing this organizational structure the quality level of the proposed KPIs is such that technological issues are greatly reduced.

Figure 4. Regional healthcare authority business intelligence framework



FUTURE TRENDS

Up to this point, this chapter has mostly dealt with organizational and strategic features of knowledge management in healthcare. In the proposed use case we have shown an ongoing implementation of a Balanced Scorecard Framework in a regional healthcare authority environment. This has been done intentionally wishing to state that the technological part on such an implementation is probably the less important issue. If one regards the future trends in knowledge management, one can see that a multitude of new tools are already proposed for use. This chapter will briefly set the focus on the trends that to our knowledge are the most promising and present more opportunities to healthcare organizations in creating effective performance management facilities.

Service Oriented Architecture and Patient-Centered Architecture (Based on the Electronic Health Record)

The term of patient-centered architecture has been already in the literature for some years. Many techniques have been used in the past such as Corbamed (Object Management Group, 2001) and Distributed Healthcare environment (DHE). The introduction of web technologies such as the Extensible Markup Language (XML), the Simple Object Access Protocol (SOAP), the Web Services Description Language (WDSL), and more precisely the concept of Web Services (Deitel, Deitel, DuWaldt, & Trees, 2003; W3C, 2003) are driving information technologies towards a Service Oriented Architecture (SOA).

A service is a software component that is suitable for cross-application access. A service is never a complete application or transaction. It is always a building block. SOA is the architecture of an application that uses services. Services define reusable business functions; SOA binds services into applications. Logically, services are defined by their interfaces. Technically, services are defined by their implementations (sometimes complex integrated flows, other times a single simple program). SOA is a logical concept, and its design is focused on the definition of service interfaces and interactions between service interfaces. Fundamental to SOA is the loose coupling between its components. At the logical level, this translates to the ability to add a new service for the end-user unobtrusively to the service provider. At the technical level, this translates to the ability of software developers to deploy a new application that calls a service without the need to redeploy or change the service. The use of SOA will allow the creation of process-based components of applications that will manipulate knowledge and information based upon the processes and the required or designed outcomes.

The Semantic Web

The Semantic Web goes beyond the World Wide Web by encoding knowledge using a structured, logically connected representation, and by providing sets of inference rules that can be used to conduct automated reasoning. Whilst the idea of knowledge representation languages is not new, existing languages generally use their own set of ontologies and inference rules to identify and eliminate logical contradictions and inconsistencies. The Resource Description Framework—RDF (W3C, 2003) and XML Topic Maps (TopicMaps.Org, 2001) are the most promising tools towards the implementation of the Semantic Web in practice. Nevertheless, a long way towards maturity has still to be covered since issues such as specific metadata frameworks and data quality are not yet solved. In any case, the Semantic Web should enhance the promotion of clinical practice guidelines and evidence based medicine. They can be seen as taxonomies of medical cases that could be both used for performance monitoring (in respect to commonly agreed levels of delivered care) and decision support for the healthcare practitioners.

Critical Pathways

Critical Pathways (Wall & Proyect, 1998) are mechanisms for transforming a reactive bureaucratic ritual to a dynamic, indispensable, clinical improvement process. A critical pathway when established is a mechanism for:

- Integrating continuous quality improvements with traditional patients' care review
- Managing and impacting of clinical and financial outcomes for a specific treatment procedure
- Proactively addressing economic and regulatory changes
- Improving clinical outcomes through reduction in variation
- Controlling unnecessary cost and resource usage without jeopardising quality of care
- Fostering multi-disciplinary approach to patient care
- Linking quality management to staff education
- Managing limited financial resources
- Making efficient use of scarce organizational resources
- Increasing readiness for anticipated changed in healthcare
- Applying and using clinical practice guidelines and other taxonomies set up by different professional societies

Critical Pathways can be seen as "specialized" performance management tools that would provide a BSC framework with very specific performance indicators for each treatment or clinical process.

CONCLUSIONS

Performance management is a key issue in the continuous process of delivering high quality healthcare services. The use of KPIs has proved that the design of a Balance Scorecard acts as the "cockpit" of a regional (or national) healthcare authority where all metrics are the flight instruments enabling the provision of healthcare based upon equity, financial control, continuous process and structure refinement, and outcome measurements. In that sense, the proposed infrastructure is, technologically speaking, an important knowledge management tool that enables knowledge sharing amongst various healthcare stakeholders and between different healthcare groups. The use of BSC is an enabling framework towards a knowledge management strategy in healthcare since KM is about discovering knowledge from existing information, about creating new knowledge and about implementing processes and taxonomies that enable the reuse and assessment of information as part and bits of knowledge.

Knowledge can be seen as a performance management tool both for administrative purposes and clinical improvements.

During the implementation process of deploying a technological platform for performance management or any other type of knowledge management infrastructure, one must have in mind that:

- The six issues described in the introduction section must being taken into strong consideration from day one.
- It is important to focus on people, processes and outcomes, and to set-up a straightforward strategy to plan, manage, assess, educate, disseminate, and maintain the developed BSC framework.
- Any type of knowledge management project bases its success on continuous improvement and assessment. Metrics and processes are meant to change in order to reflect improvements towards quality of healthcare.
- The knowledge and technology is there, but still very limited best practices have been successfully implemented.
- The implementation of BSC framework is a time consuming process that has to involve all stakeholders' representatives.

REFERENCES

Aidermark, L.G. (2001). The meaning of balanced scorecards in the health care organization. *Financial Accountability & Management, 17*(1), 23-40.

Castaneda-Mendez, K., Mangan, K. & Lavery A.M. (1998, January / February). The role and application of the balanced scorecard in healthcare quality management. *Journal of Healthcare Quality, 20*(1), 10-3.

Decision No 1400/97/EC of the European Parliament and of the Council of 30th June 1997 adopting a program of Community action on health monitoring within the framework for action in the field of public health, *Official Journal L193*, 22/07/1997, 0001-0010.

Deitel, H.M., Deitel, P.J., DuWaldt, B. & Trees, L.K. (2003). *Web services: A technical introduction*. Prentice Hall Editions.

Donabedian, A. (1980). *Explorations in quality assessment and monitoring, Volume I: The definition of quality approaches to its measurement.* Ann Arbor, MI: Health Administration Press.

Donabedian, A. (1982). *Explorations in quality assessment and monitoring, Volume II: The criteria and standards of quality.* Ann Arbor, MI: Health Administration Press.

Donabedian, A. (1985). *Explorations in quality assessment and monitoring, Volume III: the methods and findings of quality assessment and monitoring – An illustrated analysis.* Ann Arbor, MI: Health Administration Press.

Donabedian, A. (1993). Continuity and change in the quest for quality. *Clinical Performance and Quality in Healthcare, 1*, 9-16.

Donabedian, A. & Bashshur, R. (2003). *An introduction to quality assurance in health care.* Oxford: University Press.

Eder, L. (2000). *Managing healthcare information systems with Web enabled technologies*. Hershey, PA: Idea Group Publishing.

Englebardt, S. & Nelson, R. (2002). *Healthcare Informatics: An interdisciplinary approach*. Mosby Edition.

Forgione, D.A. (1997). Health care financial and quality measures: International call for a balanced scorecard approach. *Journal of Health Care Finance*, *24*(1), 55-58.

Freeman, T. (2002). Using performance indicators to improve health care quality in the public sector, a review of the literature. *Health Services Management Research*, *15*(2), 126-137.

Gordon, D. & Geiger, G. (1999). Strategic management of an electronic patient record project using the balanced scorecard. *Journal of Healthcare Information Management*, *13*(3), 113-123.

Greek National Healthcare System Reform Act (2001), *Greek Government Printing Office*, N2889/2001 / FEK-/37/02.03.2001 [In Greek].

Harmoni, A. (2002). *Effective healthcare information system*. Hershey, PA: IRM Press.

Hurst J. & Jee-Hughes, M. (2001). Performance measurement and performance management in OECD health systems. *Labour Market and Social Policy, Occasional Papers No. 47*, DEELSA/ELSA/WD(2000)8, OECD.

Iakovidis, I. (1998). Towards personal health record: Current situation, obstacles and trends in implementation of electronic healthcare records in Europe. *Internernational Journal of Medical Informatics*, *52*(123), 105-117.

Iakovidis, I. (2000). Towards a health telematics infrastructure in the European Union. In *Information technology strategies from US and the European Union: Transferring research to practice for healthcare improvement*. Amsterdam: IOS Press.

Inamdar, N., Kaplan R.S. & Bower, M. (2002). Applying the balanced scorecard in healthcare provider organizations. *Journal of Healthcare Management*, *47*(3), 179-195.

Information Society SA (2003, May 15th). Healthcare information system for the 2nd Regional Healthcare Authority of Central Macedonia, *Request For Proposal co-funded by the 3rd CSF under the EU decision C(2001)551/14-3-2001*, Greece.

Kaplan, R.S. (2001). Strategic performance measurement and management in non-profit organizations. *Non-Profit Management and Leadership*, *11*(3), 353-371.

Kaplan, R.S. & Norton, D.P. (1996). *The balanced scorecard: Translating strategy into action*. Harvard Business School Press.

McKee, M. & Healy, J. (2002). *Hospital in a changing Europe*. European Observatory on Health Care Systems Series, Open University Press.

Norris, T.E., Fuller, S.S., Goldberg, H.I. & Tarczy-Hornoch P. (2002). *Informatics in primary care, strategies in information management for the healthcare provider.* Springer Editions.

Object Management Group (2001). *The Common Object Request Broker Architecture (CORBA) specification.* OMG Document Number 2001-09-01. Information retrieved on December 15, 2003 from *http://www.omg.org/technology/documents/formal/corba_iiop.htm*

Oliveira, J. (2001). The balanced scorecard: An integrative approach to performance evaluation. *Healthcare Financial Management Journal*, *55*(5), 42-46.

Oxley, H. & MacFarlan, M. (1994). Healthcare reform controlling spending and increasing efficiency. *Economics Department Working Papers No. 149*, OCDE/GD(94)101, OECD.

Pink, G.H., Mc Killop, I., Schraa, E.G., Preyra, C., Montgomery, C. & Baker, G.R. (2001). Creating a balanced scorecard for a hospital system. *Journal of Health Care Finance*, *27*(3), 1-20.

Polyzos, N. (1999). *Hospital efficiency based upon the use of DRGs.* TYPET Edition [in Greek].

Schneiderman, A.M. (1999). Why balanced scorecard fail. *Journal of Strategic Performance Measurement*, January(4), 6-11.

Shortliffe, E.H., Perreault L.E., Wiederhold, G. & Fagan, L.M. (2001). *Medical informatics, computer applications, healthcare and biomedicine* (2nd edition). Springer Editions.

Spyrou, S., Berler, A. & Bamidis, P. (2003). Information system interoperability in a regional healthcare system infrastructure: A pilot study using healthcare information standards. *Proceedings of MIE 2003,* Saint Malo, France (pp. 364-369). IOS Press.

Stegwee, R. & Spil, T. (2001). *Strategies for healthcare information systems.* Hershey, PA: Idea Group Publishing.

Strawser, C.L. (2000). Building effective knowledge management solutions. *Journal of Healthcare Information Management*, *14*(1), 73-80.

Tarantino, D.P. (2003). Using the balanced scorecard as a performance management tool. *Physician Exec*, *29*(5), 69-72.

The TopicMap.Org (2001). Retrieved August 6, 2001, from *http://www.topicmaps.org/xtm/1.0/*

Vagelatos, A. & Sarivougioukas, J. (2003). Regional healthcare authorities delivering application services to primary health care units. *MIE 2003*, Saint Malo, France.

Wall, D.K. & Proyect, M.M. (1998). *Critical pathways development guide*. Chicago: Precept Press Editions.

Weber, D. (1999). Performance management – The balanced scorecard: A framework for managing complex and rapid changes. *Strategies for Healthcare Excellence*, *12*(11), 1-7.

The World Wide Web Consortium –W3C Web site (2003). Information retrieved March 23, 2003, from *http://www.w3.org*

Chapter 1.23
Medical Decision Support Systems and Knowledge Sharing Standards

Srinivasa Raghavan
Krea Corporation, USA

ABSTRACT

This chapter discusses about the concept of medical decision support system and the knowledge sharing standards among medical decision support systems. The author discusses the evolution of decision support in the healthcare arena, the characteristics and components of a medical decision support system, the medical decision support problem domains, and the popular medical decision support systems. Furthermore, a unique challenge in the healthcare arena—sharing of knowledge among medical decision support systems is discussed. The author discusses about the need for knowledge sharing among medical decision support systems, the evolution of various knowledge sharing standards, and the application of the knowledge sharing standards by the medical decision support systems. Finally, interesting aspects about the future trends in the medical decision support systems, its awareness, its usage and its reach to various stakeholders are discussed.

INTRODUCTION

The healthcare industry has been a pioneer in the application of decision support or expert systems capabilities. Even though the area of medical informatics and decision support has been around for more than four decades, there is no formal definition for a medical decision support system. Wyatt & Spiegelhalter (1991) describes a medical decision support system as a computer-based system using gathered explicit knowledge to generate patient specific advice or interpretation. The healthcare industry has witnessed a phenomenal growth and advances both in the areas of practice and research. This rapid growth of medical science has made the practice of medicine both challenging and complex. To address these challenges, the recognized medical standards organizations developed medical practice guidelines to simplify the research findings to practical applications in order to improve the overall healthcare quality and delivery. Despite such initiatives, it has been difficult for the physicians to keep up with the guidelines and to tune it to their practice settings. A clear gap began to develop between developing of practice guidelines and the implementation of the same. Grimshaw and Russell (1993) identified that the knowledge dissemination and implementation strategies are critical to the impact the developed guidelines will have on the physician behavior.

The natural evolution was to develop paper-based decision support workflow or protocols. Paper-based decision support models were developed and are widely used at most physician practices to apply the guidelines into practice. Paper-based decision support models, while effective, are also error prone. For example, dose recommendation models could require tedious calculations and can be error prone when performed manually. Such errors could defeat the purpose of using a decision support model. Also, paper-based models require education and training to the personnel performing those calculation and require an additional layer to review the calculations. Most paper-based models are used by the medical personnel as a batch work flow process.

Thus computer-based decision support systems were developed to provide accurate guideline compliance and to enhance physician performance (Hunt, Haynes, Hanna, & Smith, 1998). Computerized decision support system can be extremely valuable for treatment or diagnosis support and compliance accuracy when used at the point of care (Lobach & Hammond, 1997). This feature of computerized medical decision support system is a key differentiator that makes the paper-based decision support models inferior. A well designed computerized medical decision support system can be used to provide patient specific support at the desired time and location with the adequate content and pace. When decision support systems are blended into the day-to-day practice workflow, these systems have the potential to function as a valuable assistant and also as an educational tool (Thomas, Dayton, & Peterson, 1999).

The computerized decision support systems make decisions based on the clinical practice guidelines. Clinical practice guidelines are the rule-based knowledge that guides the decision makers in a medical setting. These guidelines have been developed over the years to reduce the variations among medical practices with a common goal to provide cost effective and high quality healthcare services (Field & Lohr, 1992). The availability of several decision support systems and the use of common knowledge and rules triggered the need for a common method of sharing the knowledge. This technique can save the cost of developing the same medical practice guidelines for multiple decision support systems.

While it is common among organizations to share the data and information between computer systems, sharing the knowledge was a challenge. The medical arena rose to the occasion and various associations were involved in development of a standard way of sharing the medical knowledge and clinical practice guidelines among systems. After reading this chapter, the reader should be able to:

- Understand the complexity involved in medical decision making process
- Identify the commonly used decision support techniques in medical decision support systems
- Identify the characteristics of medical decision support systems
- Identify the various phases of the decision making process
- Identify the components involved in the development of a medical decision support system
- Analyze the various decision support functions addressed by medical decision support systems

- Know about the popular medical decision support systems in use around the world
- Understand the importance of knowledge sharing standards in the medical decision support arena
- Know about the popular medical knowledge sharing standards
- Understand the future trends in decision support in the field of medicine

BACKGROUND

The invention of computers resulted from the dream of creating an electronic brain. This quest to artificial intelligence has driven several technological advances in various industries. The potential of intelligence and decision support in the area of medicine was set forth more than four decades ago by Lusted and Ledley (1959). While several computer systems have been developed to improve the administrative efficiency and medical records access, the significant challenges have always been in the development of medical decision support systems. Knowledge-based decision support systems acquire, formalize, and store the expert knowledge in a computer system and infer the represented knowledge in a problem area by modeling the decision process of experts.

One of the first expert systems developed is in the field of medicine. This expert system called MYCIN, provides decision support to physicians on treatment related advices for bacterial infections of the blood and meningitis. The basic structure of the MYCIN expert system is to present a question and answer dialog to the physician. After collecting the basic information about the patient, it asks about the suspected bacterial organisms, suspected sites, presence of relevant symptoms, laboratory results and recommends the course of antibiotics (Buchanan & Shortliffe, 1984). Even though this was a primitive expert system, it brought out the power of such decision support tool in the area of medicine and the potential for more growth in the area of artificial intelligence in medicine. After the introduction of MYCIN in the field of medicine, various decision support systems were introduced and have been successfully used. The commonly used decision support techniques in medical decision support systems are intelligent agents, rule-based engines, heuristics, and decision algorithms.

The use of an intelligent agent or softbot (intelligent software robot) is a popular decision support technique that has the potential to become one of the most important in the next decade (Turban & Aronson, 2000). There are several definitions of an intelligent agent, but the general concept is that the agent carries out a set of operations with some degree of autonomy (Murch & Johnson, 1999). To perform these tasks, an intelligent agent should contain some knowledge. The key characteristics of an intelligent agent are that it should be autonomous, goal-oriented, collaborative, and flexible (Brenner, Zarnekow, & Wittig, 1998). For example, an appropriate task for an intelligent agent would be to use a patient's history and problem list to assist the physician with diagnostic coding during the order entry process.

A heuristic is a rule of thumb that is used primarily to arrive at a "good enough" solution to a complex problem. This technique is used primarily when input data are limited and incomplete. When the problem or the reality is extremely complex and optimization techniques are not available, a heuristic may be used. Heuristics do not yield definitive solutions, but assist decision makers in arriving at provisional solutions (Camms & Evans, 1996). For example, the determination of the choice of initial doses of both medication and dialysis treatment for a renal patient relies on population-based pharmacokinetic and solute kinetic heuristics, respectively. Subsequent dosing does not require heuristic reasoning, but may be calculated on the basis of measurements of the individual patient's response to the initial dose (Raghavan, Ladik, & Meyer, 2005).

Rule-based decision support systems primarily store knowledge in the form of rules and problem solving procedures. In expert systems, the knowledge base is clearly separated from processing. Inputs may come from the user, or may be collected from various other programs. Rules are applied on the data collected using an inference engine, decisions are suggested, and explanations supplied (Turban & Aronson, 2000). For example, DARWIN, a renal decision support system, uses rule-based reasoning to advise dialysis technical staff regarding the response to bacteriologic monitoring of fluids used during hemodialysis. Another example of a rule-based or expert system is the use of clinical protocols to generate suggestions regarding anemia treatment using erythropoietin and iron protocols. On the basis of knowledge of new and historical laboratory results and medications, the expert system can suggest the modification of drug doses or the discontinuation of a medication. The inference engine can also provide an explanation of the basis of the recommendation in the protocol.

A decision algorithm is a set of instructions that is repeated to solve a problem. Highly intelligent algorithms include the capability to learn and to perform several iterations of the algorithm or of parts of the algorithm until an optimal solution is reached (Turban & Aronson, 2000). A very simple decision algorithm may alert physicians on receipt of out-of-range laboratory results. If the decision algorithm verifies the patient's history and decides if the standard abnormal range is indeed abnormal for the patient, the algorithm can be classified as an intelligent decision algorithm.

In the next sections, we will discuss about the characteristics of a medical decision support systems, generic components, problem areas in which decision support systems are used and knowledge sharing standards.

CHARACTERISTICS OF MEDICAL DECISION SUPPORT SYSTEMS

To understand the characteristics of a medical decision support system, it is important to analyze the medical decision making process. A clear understanding of the medical decision making process is essential to appreciate the value and the characteristics of a medical decision support system. Decision making in medicine is not a single point event. It is an extremely complex sequence of inter related and differentiated activities that occur over a period of time. The vastness of the knowledge area presents itself infinite paths to reach a decision. Thus, the importance of finding an optimal path is extremely important in the decision making process. Any assistance in this critical step can be of immense value to the medical decision makers. Hastie (2001) identifies three components to a decision making process. The first component is the choice of options and courses of actions. The second is the beliefs and opinions about the objective states, and the processes including input and outcomes states. The third is the values and consequences attached to each outcome of the event-choice-action combination.

The decision support literature classifies problems into three major categories: structured, semi-structured, and unstructured (Gorry & Morton, 1971). Structured problems are routine and repetitive: solutions exist, are standard, and pre-defined. Unstructured problems are complex and fuzzy; they lack clear and straight-forward solutions. Semi-structured problems combine the features of the two previous categories; their solution requires human judgment as well as the application of standard procedures (Simon, 1971).

Thus the complexity involved in a decision making process can be grouped as a knowledge gathering process, knowledge storage, knowledge retrieval, and information processing. The literature categorizes the decision making process into four distinct phases: intelligence, design, choice, and review (Stohr & Konsynski, 1992).

During the intelligence phase of the decision making process, the need for a decision making (or the trigger) event is recognized and the clinical problem or opportunity is properly identified and defined. This is done by eliciting two different types of knowledge domains: public and private. Private knowledge generally is heuristic and experience-based knowledge that usually comes from the clinical practitioners. Public knowledge can be gathered from standards, guidelines, text books and journals. The office of technology assessment (1995) suggests usage of the following questions in identifying the problems and opportunities for medical decision support.

1. Can the solution to the problem/opportunity assist in diagnosing a patient's condition?
2. Can the solution to the problem/opportunity assist in determining what the proper drug dosage level should be?
3. Can the solution to the problem/opportunity remind the appropriate care giver about the preventative services to be administered to a patient or to patient care related function?
4. Can the solution to the problem/opportunity assist in carrying out diagnostic procedure by recommending specific treatments or tests?
5. Can the solution to the problem/opportunity assist in carrying out medical procedures by alerts regarding potential adverse events?
6. Can the solution to the problem/opportunity assist in providing cost effective medical care by reminding previous orders, results, frequency rule checks, and schedule of treatment or procedure?

During the design phase, the problems or opportunities identified during the intelligence phase are further analyzed to develop possible courses of actions to construct decision models. A thorough search for ready made solutions, customized off the shelf solution, or a decision to custom solution development is made. Modeling of a decision problem involves identification of variables, identification of the relationship between those variables, and developing an abstraction into quantitative or qualitative forms.

During the choice phase, the actual decision to follow a certain course of action is made. If a single option results from the design phase, the choice phase is a simple acceptance or rejection of the option. If multiple courses of actions are identified in the design phase, then the decision makers are challenged in choosing from multiple conflicting objectives. Some decision rules for multivariable problems have applied strategies like holistic evaluation methods, heuristic elimination, and holistic judgment.

During the review phase, the outcomes of the decision making process are reviewed for validity and applicability for the case in hand. Any deviation and changes to these outcomes could very well become an expert knowledge that can be fed back into the knowledgebase to help future decision making cases.

Several studies were performed to identify the characteristics and functions of a medical decision support system. Turban and Aronson (2000) and Beatty (1999) have performed extensive research in this area of medical decision support system. Turban and Aronson (2000) summarizes the characteristics into few categories as listed below.

1. Support decision makers in semi structured and unstructured situations by bringing medical expert judgment and computer knowledgebase together.
2. Support decision making when the problem area consists of various inter-dependent and/or sequential paths.
3. Should be adaptive and flexible in nature. This allows new learning, adjustments to current knowledge and knowledge pattern recognition.

4. User friendliness is a key for success. Even a great decision support system can be useless if the users are challenged by its interface. This is particularly true in the case of healthcare arena. The system should be easy, fast, and require few mouse clicks by physicians.
5. Allow decision makers to construct simple decision constructs.
6. Should provide both automated and interactive decision support depending on the nature of the problem/opportunity.
7. Should provide complete control to the medical decision maker and should not attempt to replace the clinician's judgment.

Beatty (1999) conducted an empirical study about the characteristics and functions of a medical decision support system. This study provided several interesting findings on what support do medical decision makers expect during a decision making process. Similar to Turban and Aronson (2000) study, Beatty's (1999) finding also indicates that the decision makers favored characteristics in a decision support system that advised or guided rather than controlled. Thus, medical decision makers prefer to be in total control during the decision making process and would use the computerized system as a support or assistance during the process.

Beatty's (1999) finding also differentiated between the expectations of a medical user from a non medical user in medical decision making. This is an important variation in a medical setting as several non medical people perform various routine functions in a medical setting and medical decision support systems could assist their functions as well. The interesting finding is that the non medical users prefer to give more control and trust to the decision support system while the medical users prefer to keep control.

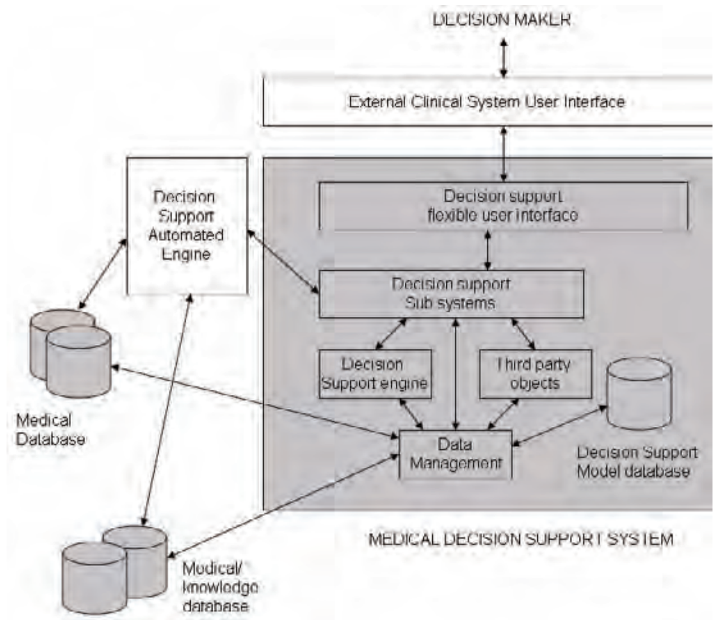
Medical decision support systems should be capable to operate in both an offline mode and a real time mode. This allows the system to be used as an educational tool during offline mode and as an advisor during the real time mode. Beatty (1999) indicates that the users prefer a detailed advice about a medical problem, treatment option, and so on, preferably through flow diagrams or case studies but a real time support should be clear, precise, alarms, and very sophisticated. The knowledge base for the decision support systems should be gathered from literature, case studies, and wide group of experts.

Turban & Aronson (2000) used a schematic view to describe the components of a decision support system. This view was used as a guideline to develop a schematic view to identify the components of a medical decision support system. As depicted in Figure 1, a typical medical decision support system consists of a medical decision support model database, a medical data management component, a medical decision support engine, third party objects including medical databases and rules, decision support sub systems, and a decision support user interface. The decision support model database contains all the tables and data needed to support the decision support models. The data needed for development of rules engine, prediction models, and protocols are stored in the model database.

The medical data management component is the critical part of the medical decision support system. This interfaces with the databases external to the decision support system to allow the decision support engine and third party interfaces to apply model database rule on the medical data to arrive at intelligent decisions.

The medical decision support engine is a set of programs that constitute the model for decision support. It was important to develop a flexible engine that could be data-driven. Several day-to-day decisions in patient care are based on changing regulations and practice guidelines. These guidelines are periodically reviewed and changed. In order to expedite such decision rule changes, it is important to maintain the rules themselves in databases, thus making the model database a meta-database.

Figure 1. Components of a medical decision support system



There are a few decision support processes that are available commercially. These third party objects may or may not include the model database bundled within the object. If they are bundled, it would become a direct interface with the decision support subsystem and the data management system. If they are not bundled, then the data is carefully placed either in the model database or in the medical database as appropriate.

Decision support subsystems are independent decision support systems that can interface with any external subsystem directly or through the decision support user interface. The medical decision support system generally consists of both automated and interactive subsystems. The automated decision support subsystems may use the automated engine to integrate the automated decision making capabilities with existing automated functions. For decision support functions that require user interaction, the decision support system flexible interface is called from the user interface layer and presented to the decision maker.

In the past four decades, numerous medical decision support systems were developed using a wide range of techniques such as automated reasoning, natural language processing, game playing, automatic programming, robotics, machine learning, and expert systems. These decision support systems generally address the medical decision support functions listed in Table 1 (Raghavan, Ladik, & Meyer, 2005). These functions are generic in nature, while the knowledge base of the decision support systems is generally specific to a medical domain area like oncology, nephrology, pediatrics, and so on.

The most common decision support function found in medical decision support systems is alerts and reminders. In a real time environment, these decision support functions are attached to the monitoring devices to provide immediate alerts as and when the trigger condition occurs. For example, the oxygen and blood pressure monitors in an acute setting can alert the nurses if the patient's condition goes beyond the set threshold value. In a chronic setting, a simple scan of laboratory results and an email or pager alert to the corresponding

Table 1. Decision support functions and clinical problems

Function	Sample Clinical Problems/Opportunities
Alert	Based on laboratory results with the range customizable at various levels
Diagnosis	Identify the possible diagnosis based on the history, physical, results, and evaluation inputs
Reminder	Reminding practitioners in order approvals, and schedules
Notification	Non conformance, risks, abnormal events, and episodes of care
Suggestion	Drug adjustments based on the recent lab values, trends, and current drug levels
Interpretation	Guidelines as applicable to the current situation – lab test schedule, protocol development
Prediction	Predicting results based on some independent variables
Assistance	Providing drug to drug interaction and drug-allergy interaction
Critique	The usage of diagnostic codes for medical procedure justification based on the applicable guidelines and the patient's medical history

decision maker are valuable decision support functions. A few medical decision support systems can provide image recognition and interpretation functions. These are extremely helpful in large hospitals where various radiology reports can be interpreted and alerts can be generated to gain the attention of the experts.

Diagnostic support is a key function that several medical decision support systems attempt to offer to help the physicians in detecting the problem based on symptoms and etiology. Such systems are commonly used to detect rare diseases and also as an aid for inexperienced practitioners. A few medical decision support systems offer care plan critiques by looking at data inconsistencies, errors, and omissions against practice guidelines.

In the next section, the problem domain, and the decision support functions provided by various popular medical decision support systems will be discussed. While there are close to hundred known medical decision support systems in use, the most popular medical decision support systems based on the existing literature and usage will be discussed in this chapter.

MEDICAL DECISION SUPPORT SYSTEMS

HELP (Health Evaluation through Logical Processes)

This is the most successful and popular medical decision support system in the United States. The HELP system of the Latter Day Saints Hospital in Salt Lake City is a hospital-based medical information system that gives practitioners a comprehensive patient record with decision support capabilities. HELP is a comprehensive medical information system with full fledged medical records, physician order entry, charges, radiology, pharmacy, ICU monitoring, laboratory, and robust decision support functions (MedExpert-HELP, 2004).

The success of HELP information system was due to the fact that the decision support functions were integrated with the hospital information system. The availability of required data and the knowledge base makes this a powerful system. The decision support capabilities made available to the decision makers at the point of care made this an effective product. This system has also been implemented at more than twenty hospitals.

The HELP information system supports the following decision support functions: alerts and reminders, decision critiquing, patient diagnosis, care suggestions, and protocols. The integrated laboratory information system allows automatic monitoring and alerts for abnormalities or out of range values. The critiquing feature is integrated within the transfusion ordering module to critique the reason provided against the strict guidelines. The system also includes modules such as automated surveillance system that uses various data elements to diagnose nosocomial infections. As an extension of this module, the antibiotic assistant recommends the antibiotics that can produce optimal benefit to the patient.

DXplain

Dxplain is a diagnostic decision support system that is owned by Massachusetts General Hospital and can be licensed and accessed over the internet as well. DXplain is a powerful diagnosis decision support system used in general medicine. The power of DXplain is its knowledge base that can diagnose close to 2000 diseases emphasizing the signs and symptoms, etiology, pathology and prognosis. This system was developed during the early 1990s by Massachusetts General Hospital and has been used by thousands of physicians and medical students.

DXplain uses a set of signs, symptoms, and laboratory results to produce a ranked list of diagnoses which might explain the clinical manifestations. DXplain uses Bayesian logic to derive the clinical diagnosis interpretation (MedExpert-DxPlain, 2004). It is important to note that the system only provides suggestion and not definitive conclusions. Since this system is in production and licensed as a product, the knowledge base should be continually updated.

DXplain is routinely used at medical schools for clinical education. DXplain diagnosis decision system has characteristics of an electronic medical textbook and a reference system. After receiving the inputs, the system develops a list of ranked diagnosis based on the input and provides the justification. This is one of the main reasons why this system is used widely for education purposes.

DXplain system provides about 10 appropriate references for each suggested diagnosis. The user interface for the DXplain system is very intuitive with graphical user interface requiring only clicks from the lists to drill down to the recommendations. This is one of the main reasons suggested for the success of this product.

RMRS

Regenstrief Medical Records System (RMRS) is one of the oldest, popular, and commonly cited decision support system in the United States. RMRS, developed at Indiana University, is a hospital-based medical information system with decision support capabilities like preventive care reminders and advice regarding cost- effectiveness details during physician order entry.

As early as 1974, RMRS began to deliver paper generated automatic reminders to the physicians the night before a patient's visit. This intelligent system reviewed the patient's medical record against the set of pre-defined protocols to generate reminders about the patient's condition and to suggest corrective actions. In 1984, this report-based decision support system was programmed into an interactive Gopher-based system to break the long list of reminders into logical groups, allowing physicians to step through the reminders during the physician order entry process. The success of RMRS system was due to the fact that the decision support functions were integrated with the hospital information system. The availability of required data and the knowledge base makes this a powerful system. The decision support capabilities made available to the decision makers at the point of care made this an effective product.

In the past decade, RMRS went through a lot of developments and has included various decision support techniques such as pattern recognition and intelligent agents to help physicians assess the patterns of care and identify patients with particular risk factors and adverse outcomes. This system is used at more than 40 inpatient and outpatient facilities and the commercial version is marketed internationally.

PRODIGY

PRODIGY (PROject prescribing rationally with Decision support In General practice study) is a prescribing decision support system that offers evidence-based, cost effective general practice prescribing. The Sowerby Center of Health Informatics at Newcastle University was approached by the English National Health Service Executive during 1995 to develop a medical decision support system to guide UK general practitioners on therapeutic actions covering prescribing. While developing PRODIGY, the research team realized the importance of integration with already existing general practice software used by the general practitioners. Hence integration with all the five major general practice software in the UK was performed to make PRODIGY a powerful medical decision support system.

While several medical decision support systems assist in diagnosing the problem, PRODIGY starts after the diagnosis is made by providing medical advice and therapeutic recommendations (MedExpert-Prodigy, 2004). The system also stores patient information leaflets containing information about the disease in simple terms. The trigger for decision support within PRODIGY is the input of diagnosis code. Once the physician enters the diagnosis code in their general practice system, it activates PRODIGY decision support to check if suitable medical recommendations are available. The decision maker at that point has the option to explore the recommendations or ignore it. When explored, the decision maker will be presented with therapy scenarios, evidence-based prescribing, and availability of patient information leaflets along with justification and references for the recommendations. PRODIGY medical decision support system has been successfully implemented nation wide at all general practitioners in the UK.

CADIAG – II (Computer Assisted DIAGnosis)

CADIAG – II is a hospital-based medical decision support system developed by the Department of Medical Computer Sciences at University of Vienna. CADIAG decision support system has been integrated with the medical information system of the Vienna General Hospital.

CADIAG decision support system focuses on the colon diseases, rheumatic diseases, gall bladder, pancreatic diseases, and bile duct diseases (MedExpert-Cadiag, 2004). This medical decision support system provides online consultation support for the physician to assist in diagnosis of the diseases given the signs, symptoms, etiology, and laboratory results. This system contains a very strong knowledge database in the specialty areas of colon diseases and rheumatic diseases.

NeoGanesh

NeoGanesh is a widely cited knowledge-based decision support system that is used primarily to manage the mechanical ventilation in Intensive Care Units. This decision support system is one of the very few automated and controlled system with very limited human intervention. This real time medical decision support system checks for the real time data and controls the mechanical assistance provided to the patients suffering from lung diseases in a pressure supported ventilation mode (Dojat, Pachet, Guessoum, Touchard, Harf, & Brochard, 1997). The other interesting feature of the system is to develop a therapeutic strategy to respond to the patients and evaluate the capacity to breathe. This system is used at Henri Mondor Hospital in France. This system is being planned to be released for commercial purpose as well.

MEDICAL DECISION SUPPORT KNOWLEDGE SHARING

Overview

The field of medicine has been a pioneer in the introduction of decision support systems. One of the first expert systems to be developed was in the field of medicine. The complexity of decision making in the field of medicine and the challenge of keeping up to date with the new findings and research has been a key motivator for the usage of medical decision support systems. As discussed in the previous sections, numerous medical decision support systems exist in the market. The knowledge for most of these medical decision support systems are acquired from the clinical practice guidelines issued by various government and medical associations. Clinical practice guidelines are the rule-based knowledge that guides the decision makers in a medical setting. These guidelines have been developed over the years to reduce the variations among medical practices with a common goal to provide cost effective and high quality healthcare services (Field & Lohr, 1992).

Healthcare organizations historically have focused more on the development of the clinical practice guidelines compared to the implementation of the guidelines (Grimshaw & Russell, 1993). Past research studies have proved that the computerized medical decision support systems when integrated with the clinical workflow can improve the practitioners' compliance with clinical practice guidelines and outcomes (Johnsons, Langton, Haynes & Mathieu, 1994). Thus, it was clear that the medical decision support systems are best vehicles to promote compliance. The availability of several decision support systems and the use of common knowledge and rules triggered the need for a common method of sharing the knowledge. This technique can save the cost of developing the same medical practice guidelines for multiple decision support systems.

While it is common among organizations to share the data and information between computer systems, sharing the knowledge was a challenge. While knowledge sharing could be considered as losing competitive advantage, the healthcare industry was one of the few industries where several guidelines are common as they are generally developed by government agencies or medical associations. The medical arena rose to the occasion and various associations were involved in development of a standard way of sharing the medical knowledge and clinical practice guidelines among systems. In the following sections, the various prominent clinical practice guideline models and knowledge sharing standards are discussed.

ARDEN Syntax

ARDEN Syntax is an ASTM (American Society for Testing and Materials) and ANSI (American National Standard Code for Information Interchange) standard a hybrid knowledge representation format that was created with the main aim to address the ability to share, reuse, and understand the medical knowledge base. It is first introduced in 1989 at the Columbia University's Arden homestead conference. The initial idea for the conference was to create a knowledge representation format that can facilitate the definition of knowledge bases and thus could result in an effective implementation of new knowledge consistently (Hripcsak, Pryor & Wigertz, 1995). The development of knowledge bases is the biggest challenge faced by the ever changing medical field and introducing a knowledge representation format was thought to be of significant help to allow various institutions to share their knowledge among medical decision support systems.

The ARDEN syntax was largely derived from the logical modules used by the HELP and RMRS medical decision support systems. The concept of the ARDEN syntax is to develop rule-based modular logical rules called medical logical modules (MLM). The medical decision support system that is based on Arden Syntax polls for the occurring events and executes the MLM's out of its knowledge base that has defined the occurring event as the triggering condition.

In ARDEN syntax, the MLM's are composed of slots that are grouped into three categories called maintenance, library, and knowledge. Each category has a set of predefined slots. The slots are broadly classified as textual slots, textual list slots, coded slots, and structured slots. The maintenance category contains the slots that specify general information about the MLM like title, mlmname, arden syntax version, mlm version, institution, author, specialist, date, and validation. The library category contains slots about the knowledge base maintenance that is related to the module's knowledge. This category contains the relevant literature, explanation, and links that were used in defining the MLM. The library category slots are purpose, explanation, keywords, citations and links. The knowledge category contains the slots that specify the real knowledge of the MLM. The knowledge category dictates the triggering event of the MLM and the logic of the MLM. The knowledge category slots are type, data, priority, evoke, logic, action, and urgency.

Example

Let us look at a simple anemia management protocol for a renal patient as an example. On receipt of new hemoglobin laboratory result, if the value is greater than 13.3, and if the patient is on Erythropoietin dose, then recommend discontinuation of the dosage. The MLM for this simple rule is given in Box 1.

ARDEN syntax-based decision support systems are widely in use. Since this syntax became an ANSI standard, it was embraced by several medical information system vendors to provide decision support capabilities. Vendors who have developed ARDEN compliant applications include: Cerner Corporation, Healthvision, McKesson, SMS, and Micromedex.

Asbru

Asbru is a knowledge representation language to capture the clinical guideline and protocols as time oriented skeletal plans. The development of Asbru was part of the Asgaard project developed by the Technical University of Vienna, Stanford Medical Informatics, University of Newcastle, and University of Vienna. The goal of the Asgaard project is to develop task specific problem solving methods that perform medical decision support and critiquing tasks (Seyfang, Miksch, & Marcos, 2002).

Asbru is a difficult language to be understood by a physician and hence a graphical knowledge acquisition tool is used to gather the knowledge or guideline and then output that into Asbru syntax behind the scenes for computer interpretation. The graphical tool used is called Protégé. The Asbru output created by Protégé can be shared with other medical decision support systems that are compliant with Asbru format.

Box 1.

```
maintenance:

    title:  Screen for anemia management (triggered by hemoglobin storage);;

    filename:  renal_anemia_Hemo_Epo;;
    version:  1.00;;
    institution:  Sample Institute; Sample Medical Center;;
    author:  Sri Raghavan, PhD.;;
    specialist:  ;;
    date:  2004-03-01;;
    validation:  testing;;

library:

    purpose:
      Warn the patient care personnel of the level of hemoglobin and EPO dosage.;;

    explanation:
      Whenever a hemoglobin blood result is stored, it is checked for anemia
      management. This simple protocol checks if the value is greater than 13.3,
      and if EPO is administered to the patient, then an alert is generated to
      recommend the practitioner to discontinue the dosage.;;

    keywords:  anemia; hemoglobin; renal; Erythropoietin;;

    citations:  ;;

knowledge:

    type:  data_driven;;

    data:

      /* evoke on storage of a hemoglobin result*/
      storage_of_hgb := EVENT {storage of hemoglobin};

      /* read the potassium that evoked the MLM */
      hgb:= READ LAST {hemoglobin level};

      /* get the last active Erythropoietin order */
      epo_order := read last epo order};

      ;;

    evoke:
      /* evoke on storage of a hemoglobin */
      storage_of_hemoglobin;;

    logic:

      /* exit if the hemoglobin value is invalid */
      if hemoglobin is not number then
          conclude false;
      endif;

      /* exit if there hemoglobin is <=13.3 */
      if hemoglobin <= 13.3 then
          conclude false;
      endif;

      /* exit if Erythropoietin order cannot be found */
```

continued on following page

Box 1. continued

```
        if (epo_order is null) then
            conclude false;
        endif;

        /* send an alert */
        conclude true;

        ;;

      action:

        write "The patient's hemoglobin level on (" ||time of potassium|| ") is" ||hemoglobin|| ". The patient is currently on
Erythropoietin dosage. As per clinical guideline, it is recommended to discontinue the Erythropoietin dosage.";

        ;;

      urgency:  90;;

    end:
```

EON

EON is a clinical guideline modeling system developed by Stanford University for creating guideline-based decision modules. The EON guideline model system was developed to address the problem of modeling clinical guidelines and protocols to provide patient specific decision support. This modeling system allows creating a knowledge engineering environment for easy encoding of clinical guidelines and protocols. EON guideline model allows association of conditional goals with guidelines. It provides three criteria languages to allow the medical practitioners to use and express their decision making complexities: a simple object oriented language that medical practitioners can use to encode the decision criteria, a temporal query type language, and a predicate logic.

EON also uses Protégé as the medical knowledge acquisition tool to help medical practitioners provide medical decision criteria. The EON language is generated at the back end by the Protégé tool. This code is platform independent and can be shared between decision support systems that are EON compliant.

EON is widely used in European countries and one of the major decision support system that uses EON is PRODIGY. The decision support system ATHENA, which provides hypertension related advices, also uses the EON guideline model as its decision support architecture.

GEM (Guideline Elements Model)

GEM, an XML- (Extensible Markup Language) based guideline document model was introduced in 2000 by Yale center of medical informatics. This initiative is intended to facilitate the translation of medical guideline documents to a standard computer interpretable format. The key difference between the other techniques and GEM is that it uses the industry standard XML for knowledge transfer purposes. GEM's XML hierarchy is made up of 100 discrete tags and 9 major branches. These branches are identity, developer, intended audience, target population, method of development, testing, review plan, and knowledge components (Shiffman, Karras, Agrawal, Chen, Marenco, & Nath, 2000). Several of these branches appear identical to the slots of ARDEN syntax. While the goal of all of these knowledge transfer techniques is the same, they differ in their implementation strategies.

The strength of this model is that it encodes both the recommendations and adequate information about the guideline recommendations, including its reason and quality of evidence (Shiffman et al., 2000). GEM was accepted as an ASTM standard in 2002. Knowledge extraction in GEM is a simple document markup rather than a programming task. GEM uses a knowledge acquisition tool called GEM Cutter to interactively gather he knowledge data and store it in a XML format.

GLARE (Guide Line Acquisition Representation and Execution)

GLARE is a medical domain independent guideline model system for acquiring, representing, and executing clinical guidelines. GLARE was introduced in 1997 by the Dipartimento di Informatica, Universita del Piemonto Orientale , Alexandria, Italy. This system boasts a modular architecture with acquisition and execution modules. The GLARE Knowledge representation language is designed to satisfy both the complexity and expressive rules. The format is limited but focused set of primitives. It consists of plans and atomic actions that can be queries, decisions, work actions, and conclusions (Open Clinical-Glare, 2004).

GLARE knowledge acquisition system is an easy to use graphical user interface with syntactic and semantic verifications to check the formulated guidelines. Artificial intelligence temporal reasoning techniques are applied to weed out inconsistencies. The GLARE system goes one step ahead and also incorporates a decision support system for knowledge execution purposes. GLARE system has been tested for applicability in medical domains like bladder cancer, reflux esophagitis, and heart failure at the Laboratorio di Informatica Clinica, Torino, Italy. The commercial version of this product is still under development.

FUTURE TRENDS

While medical decision support systems have been around for the past four decades, its usage among clinicians is still questionable. Several variables play a role in a success of a medical decision support system and it is important to address these to move ahead from where we stand now. Issues like physician reluctance, intimidating systems, proprietary interests, local practice, technological factors, and outdated knowledgebase are all limiting factors for the growth and acceptance of computer-based decision support systems. In addition to these limitations, the decision making functions are no longer limited to the physicians. The patients, healthcare administrators, insurance carriers, healthcare professionals, policy makers, risk analysts, and the government at large are all stakeholders in this medical decision making process. It will be extremely limiting to tie down these requirements to local proprietary computer systems.

While the field of medicine pioneered the knowledge sharing techniques among decision support systems, the introduction of several sharing standards segmented market as a net result defeated the purpose of such standard. The growth, availability, and accessibility of the Internet opened the door for development of a new model for sharing medical knowledge and decision support systems across networks. Internet-based decision support systems with focus on a particular medical domain will be prevalent in the future.

The various knowledge sharing standards are naturally meeting a common merger. For example, several medical guideline modeling languages are using a common interface for knowledge acquisition purposes. Protégé is the most commonly used tool to acquire the knowledge and then formulate the output in various modeling languages. This converging trend will begin to happen at the knowledge representation and knowledge execution modules as well. This will give raise to a true acceptance of a common knowledge sharing technique.

The next key area of medical decision support systems is the unique characteristics of its users. Great medical decision support systems when seldom used solves no purpose. Most users get intimidated by the complexities and challenges posed by the user interface. While medical decision making is not a simple process, the decision support systems should help reducing the complexity and not increasing them. Past studies indicate that the complexity of customizing the knowledge or a simple rule forces the physicians to refrain form using the system. Training clinicians in usage of complex decision support process is a worthy investment. The benefits of the usage of medical decision support systems outweigh the costs incurred in development, implementation, and maintenance of such products.

CHAPTER SUMMARY

Medical decision support systems, when used along with the clinical information systems at the point of care can improve the quality of care provided. Despite evidence of the benefits of decision support, and despite endorsement by a national patient safety panel (Kohn, Corrigan & Donaldson, 2000), computerized clinical decision support has not achieved wide diffusion. A recent survey found that such systems were used in less than five percent of all healthcare facilities (Wong, Legnini, Whitmore, & Taylor, 2000). When the user interface, availability, and accessibility issues of the medical decision support systems are properly addressed, the usage of the systems will grow multi fold! When such medical decision support systems are blended into the day-to-day practice workflow, then these systems are destined to succeed!

REFERENCES

Beatty, P.C.W. (1999). User attitudes to computer-based decision support in anesthesia and critical care: A preliminary survey. *The Internet Journal of Anesthesiology*, *3*(1).

Brenner, W., Zarnekow, R. & Wittig, H. (1998). *Intelligent software agents: Foundations and applications.* New York: Springer-Verlag.

Buchanan, B.G. & Shortliffe, E.H. (1984). *Rule-based expert systems: The MYCIN experiments of the Stanford heuristic programming project.* Reading, MA: Addison-Wesley.

Camms, J.D. & Evans, J.R. (1996). *Management science: Modeling, analysis and interpretation*, Cincinnati, OH: South-Western.

Dojat, M., Pachet, F., Guessoum, Z., Touchard, D., Harf, A., & Brochard, L. (1997). NeoGanesh: A working system for the automated control of assisted ventilation in ICUs. *Artificial Intelligence in Medicine*, *11*, 97-117.

Field, M.J. & Lohr, K.B. (1992). *Guidelines for clinical practice: From development to use*. Washington, DC: National Academy Press.

Frenster, J.H. (1989). *Expert systems and open systems in medical artificial intelligence*. Presented at the Congress on Medical Informatics. May 11-13.

Gorry, G.A. & Morton, S. (1971). A framework for management information systems, *Sloan Management Review*, *13*(1).

Grimshaw, J.M. & Russell, I.T. (1993). Effect of clinical guidelines on a medical practice: A systematic review of rigourous evaluations. *Lancet*, 342(8883).

Hastie, R. (2001). Problems for judgment and decision making. *Ann Rev Psychol*, *52*, 653-683.

Hripcsak, G., Pryor, A.T. & Wigertz, O.B. (1995). Transferring medical knowledge bases between different HIS environments. In H.U. Prokosch & J. Dudeck (Eds.), *Hospital information systems: Design and development characteristics, impact and future architectures*. Amsterdam: Elsevier Science B.V.

Hunt, D.L., Haynes, R.B., Hanna, S.E. & Smith, K. (1998). Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *JAMA*, *280*(15).

Johnsons, M.E., Langton, K.B., Haynes, R.B. & Mathieu, A. (1994). Effects of computer based clinical decision support systems on clinician performance and patient outcome. *Ann Intern Med.*, *120*(2).

Kohn, L.T., Corrigan, J. & Donaldson. M.S. (2000). *To err is human: Building a safer health system*. Washington, DC: National Academic Press.

Lobach, D.F. & Hammond, E. (1997). Computerized decision support based on a clinical practice guideline improves compliance with care standards. *Am J Med.*, *102*(1).

Ludley, R.S. & Lusted, L.B. (1959). Reasoning foundations of medical diagnosis. *Science*.

MedExpert-Cadiag (2004). MedExpert WWW - A medical knowledge base server - HELP. Retrieved March 31 2004, from *http://medexpert.imc.akh-wien.ac.at/cadiag_info.html*

MedExpert-DxPlain (2004). MedExpert WWW - A medical knowledge base server - HELP. Retrieved March 31 2004, from *http://medexpert.imc.akh-wien.ac.at/dxplain_info.html*

MedExpert-HELP (2004). MedExpert WWW - A medical knowledge base server - HELP. Retrieved March 31, 2004, from *http://medexpert.imc.akh-wien.ac.at/help_info.html*

MedExpert-Prodigy (2004). MedExpert WWW - A medical knowledge base server - HELP. Retrieved March 31, 2004, from *http://medexpert.imc.akh-wien.ac.at/prodigy_info.html*

Murch, R. & Johnson, T. (1999). *Intelligent software agents*. Upper Saddle River, NJ: Prentice Hall.

Office of Technology Assessment. (1995). *Bringing health care online: The role of information technologies*. Washington, DC: US Government Printing Office.

Open Clinical-Glare (2004). Open clinical knowledge management for medical care - Guideline modelling methods and technologies. Retrieved March 31, 2004, from *http://www.openclinical.com/gmm_glare.html*

Patel, V.L, Kaufman, D.R. & Arocha, J.F. (2000). Emerging paradigms of cognition in medical decision making. *Journal of Biomedical Informatics*.

Rahavan, S., Ladik, V. & Meyer, K.B. (2005). Developing decision support for dialysis treatment of chronic kidney failure. *IEEE Transactions on Information Technology in BioMedicine*.

Seyfang, A., Miksch, S. & Marcos, M. (2002). Combining diagnosis and treatment using Asbru. *International Journal of Medical Informatics*, *68*, 1-3.

Shiffman, R.N, Karras, B.T, Agrawal, A., Chen, R., Marenco, L. & Nath, S. (2000). GEM: A proposal for a more comprehensive guideline document model using XML. *JAMIA 2000*, *7*(5).

Simon, H. (1971). *The new science of management*. Englewood Cliffs, NJ: Prentice Hall.

Stohr, E.A & Konsynski, B.R. (1992). *Information systems and decision processes*. IEEE Computer Society Press.

Thomas, K.W., Dayton, C.S. & Peterson, M.W. (1999). Evaluation of Internet based clinical decision support systems. *Jounal of Medical Internet Research, 1*(2), 6.

Turban, E. & Aronson, J.E. (2000). *Decision support systems and intelligent systems.* Upper Saddle River, NJ: Prentice Hall.

Wong, H.J., Legnini, M.W., Whitmore, H.H. & Taylor, R.S. (2000). The diffusion of decision support systems in healthcare: Are we there yet? *Journal of Healthcare Management,* 45(4).

Wyatt, J. & Spiegelhalter, D. (1991). Field trials of medical decision-aids: Potential problems and solutions. In P.D. Clayton (Ed.) *Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care*. Washington: American Medical Informatics Association. 3-7.

This work was previously published in Clinical Knowledge Management: Opportunities and Challenges, edited by B. Bali, pp. 196-218, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 1.24
Kernel Methods in Genomics and Computational Biology

Jean-Philippe Vert
Ecole des Mines de Paris, France

ABSTRACT

Support vector machines and kernel methods are increasingly popular in genomics and computational biology due to their good performance in real-world applications and strong modularity that makes them suitable to a wide range of problems, from the classification of tumors to the automatic annotation of proteins. Their ability to work in a high dimension and process nonvectorial data, and the natural framework they provide to integrate heterogeneous data are particularly relevant to various problems arising in computational biology. In this chapter, we survey some of the most prominent applications published so far, highlighting the particular developments in kernel methods triggered by problems in biology, and mention a few promising research directions likely to expand in the future.

INTRODUCTION

Recent years have witnessed a dramatic evolution in many fields of life science with the apparition and rapid spread of so-called high-throughput technologies, which generate huge amounts of data to characterize various aspects of biological samples or phenomena. To name just a few, DNA sequencing technologies have already provided the whole genome of several hundreds of species, including the human genome (International Human Genome Sequencing Consortium, 2001; Venter, 2001). DNA microarrays (Schena, Shalon, Davis, & Brown, 1995), that allow the monitoring of the expression level of tens of thousands of transcripts simultaneously, opened the door to functional genomics, the elucidation of the functions of the genes found in the genomes (DeRisi, Iyer, & Brown, 1997). Recent advances in ionization technology have boosted large-scale capabilities in mass spectrometry and the rapidly growing field of proteomics, focusing on the systematic, large-scale analysis of proteins (Aebersold & Mann, 2003). As biology suddenly entered this new era characterized by the relatively cheap and easy generation of huge amounts of data, the urgent need for efficient methods to represent, store, process, analyze, and finally make sense out of these data triggered the parallel development of numerous data-analysis algorithms in computational biology. Among them, kernel methods in general and support vector machines (SVMs) in particular have quickly gained popularity for problems involving the classification and analysis of high-dimensional or complex data. Half a decade after the first pioneering papers (Haussler, 1999; T. S. Jaakkola, Diekhans, & Haussler, 1999; Mukherjee, Tamayo, Mesirov, Slonim, Verri, & Poggio, 1998), these methods have been applied to a variety of problems in computational biology, with more than 100 research papers published in 2004 alone. The main reasons behind this fast development involve, beyond the generally good performances of SVM on real-world problems and the ease of use provided by current implementations, (a) the particular capability of SVM to resist high-dimensional and noisy data, typically produced by various high-throughput technologies, (b) the possibility to model linear as well as nonlinear relationships between variables of interest, and (c) the possibility to process nonvectorial data, such as biological sequences, protein structures, or gene networks, and to easily fuse heterogeneous data thanks to the use of kernels. More than a mere application of well-established methods to new data sets, the use of kernel methods in computational biology has been accompanied by new developments to match the specificities and the needs of the field, such as methods for feature selection in combination with the classification of high-dimensional data, the invention of string kernels to process biological sequences, or the development of methods to learn from several kernels simultaneously. In order to illustrate some of the most prominent applications of kernel methods in computational biology and the specific developments they triggered, this chapter focuses on selected applications related to the manipulation of high-dimensional data, the classification of biological sequences, and a few less developed but promising applications. This chapter is therefore not intended to be an exhaustive survey, but rather to illustrate with some examples why and how kernel methods have invaded the field of computational biology so rapidly. The interested reader will find more references in the book by Schölkopf, Tsuda, and Vert (2004) dedicated to the topic. Several kernels for structured data, such as sequences or trees, widely developed and used in computational biology, are also presented in detail in the book by Shawe-Taylor and Cristianini (2004).

CLASSIFICATION OF HIGH-DIMENSIONAL DATA

Several recent technologies, such as DNA microarrays, mass spectrometry, or various miniaturized assays, provide thousands of quantitative parameters to characterize biological samples or phenomena. Mathematically speaking, the results of such experiments can be represented by high-dimensional vectors, and many applications involve the supervised classification of such data. Classifying data in high dimension with a limited number of training examples is a challenging task that most statistical procedures have difficulties in dealing with, due in particular to the risk of overfitting the training data. The theoretical foundations of SVM and related methods, however, suggest that their use of regularization allows them to better resist to the curse of dimension than other methods. SVMs were therefore naturally tested on a variety of data sets involving the classification of high-dimensional data, in particular, for the analysis of tumor samples from gene expression data, and novel algorithms were developed in the framework of kernel methods to select a few relevant features for a given high-dimensional classification problem.

Tumor Classification from Gene Expression Data

The early detection of cancer and prediction of cancer types from gene expression data have been among the first applications of kernel methods in computational biology (Furey, Cristianini, Duffy, Bednarski, Schummer, & Haussler, 2000; Mukherjee et al., 1998) and remain prominent. These applications indeed have potentially important impacts on the treatment of cancers, providing clinicians with objective and possibly highly accurate information to choose the most appropriate form of treatment. In this context, SVMs were widely applied and compared with other algorithms for the supervised classification of tumor samples from expression data of typically several thousands of genes for each tumor. Examples include the discrimination between acute myeloid and acute lymphoblastic leukemia (Mukherjee et al.); colon cancer and normal colon tissues (Moler, Chow, & Mian, 2000); normal ovarian, normal nonovarian, and ovarian cancer tissues (Furey et al.); melanoma; soft-tissue sarcoma and clear-cell sarcoma (Segal, Pavlidis, Noble, et al., 2003); different types of soft-tissue sarcomas (Segal, Pavlidis, Antonescu, et al., 2003); and normal and gastric tumor tissues (Meireles et al., 2003), to name just a few. Another typical application is the prediction of the future evolution of a tumor, such as the discrimination between relapsing and nonrelapsing Wilms tumors (Williams et al., 2004), the prediction of metastatic or nonmetastatic squamous cell carcinoma of the oral cavity (O'Donnell et al., 2005), or the discrimination between diffuse large B-cell lymphoma with positive or negative treatment outcome (Shipp et al., 2002).

The SVMs used in these studies are usually linear hard-margin SVMs, or linear soft-margin SVMs with a default C parameter value. Concerning the choice of the kernel, several studies observe that nonlinear kernels tend to decrease performance (Ben-Dor, Bruhn, Friedman, Nachman, Schummer, & Yakhini, 2000; Valentini, 2002) compared to the simplest linear kernel. In spite of contradictory opinions (Pochet, De Smet, Suykens, & De Moor, 2004), this is coherent with the intuition that the complexity of learning nonlinear functions in very high dimension does not play in their favor. On the other hand, the choice of hard-margin SVM, sometimes advocated as a default method when data are linearly separable, is certainly worth questioning in more detail. Indeed, the theoretical foundations of SVM suggest that in order to learn in high dimension, one should rather increase the importance of regularization as opposed to fitting the data, which corresponds to decreasing the C parameter of the soft-margin formulation. A few recent papers highlight indeed the fact that the choice of C has an important effect on the generalization performance of SVM for the classification of gene expression data (Huang & Kecman, 2005).

A general conclusion of these numerous studies is that SVMs generally provide good classification accuracy in spite of the large dimension of the data. For example, in a comparative study of several algorithms for multiclass supervised classification, including naive Bayes, *k*-nearest neighbors, and decision trees, Li, Zhang, and Ogihara (2004, p. 2434) note that "[SVMs] achieve better performance than any other classifiers on almost all the datasets." However, it is fair to mention that other studies conclude that most algorithms that take into account the problem of large dimension either through regularization or through feature selection reach roughly similar accuracy on most data sets (Ben-Dor et al.). From a practical point of view, the use of the simplest linear kernel and the soft-margin formulation of SVM seems to be a reasonable default strategy for this application.

Feature Selection

In the classification of microarray data, it is often important, for classification performance, biomarker identification, and the interpretation of results, to select only a few discriminative genes among the thousands of candidates available on a typical microarray. While the literature on feature selection is older and goes beyond the field of kernel methods, several interesting developments with kernel methods have been proposed in the recent years, explicitly motivated by the problem of gene selection from microarray data.

For example, Su, Murali, Pavlovic, Schaffer, and Kasif (2003) propose to evaluate the predictive power of each single gene for a given classification task by the value of the functional minimized by a one-dimensional SVM, trained to classify samples from the expression of only the single gene of interest. This criterion can then be used to rank genes and select only a few with important predictive power. This procedure therefore belongs to the so-called filter approach to feature selection, where a criterion (here using SVM) to measure the relevance of each feature is defined, and only relevant features according to this criterion are kept.

A second general strategy for feature selection is the so-called wrapper approach, where feature selection alternates with the training of a classifier. The now widely used recursive feature elimination (RFE) procedure of Guyon, Weston, Bamhill, and Vapnik (2002), which iteratively selects smaller and smaller sets of genes and trains SVMs, follows this strategy. RFE can only be applied with linear SVMs, which is nevertheless not a limitation as long as many features remain, and works as follows. Starting from the full set of genes, a linear SVM is trained and the genes with the smallest weights in the resulting linear discrimination function are eliminated. The procedure is then repeated iteratively starting from the set of remaining genes and stops when a desired number of genes is reached.

Finally, a third strategy for feature selection, called the embedded approach, combines the learning of a classifier and the selection of features in a single step. A kernel method following this strategy has been implemented in the joint classifier and feature optimization (JCFO) procedure of Krishnapuram, Carin, and Hartemink (2004). JCFO is, roughly speaking, a variant of SVM with a Bayesian formulation, in which sparseness is obtained both for the features and the classifier expansion in terms of a kernel by appropriate choices of prior probabilities. The precise description of the complete procedure to train this algorithm, involving an expectation-maximization (EM) iteration, would go beyond the scope of this chapter, and the interested reader is referred to the original publication for further practical details.

Generally speaking, and in spite of these efforts to develop clever algorithms, the effect of feature selection on the classification accuracy of SVM is still debated. Although very good results are sometimes reported, for example, for the JCFO procedure (Krishnapuram et al., 2004), several studies conclude that feature selection, for example, with procedures like RFE, do not actually improve the accuracy of SVM trained on all genes (Ambroise & McLachlan, 2002; Ramaswamy et al., 2001). The relevance of feature-selection algorithms for gene expression data is therefore currently still a research topic that practitioners should test and assess case by case.

Other High-Dimensional Data in Computational Biology

While early applications of kernel methods to high-dimensional data in genomics and bioinformatics mainly focused on gene expression data, a number of other applications have flourished more recently, some being likely to expand quickly as major applications of machine learning algorithms. For example, studies focusing on tissue classification from data obtained by other technologies, such as methylation assays to monitor the patterns of cytosine methylation in the upstream regions of genes (Model, Adorjan, Olek, & Piepenbrock, 2001), or array comparative genomic hybridization (CGH) to measure gene copy number changes in hundreds of genes simultaneously (Aliferis, Hardin, & Massion, 2002), are starting to accumulate. A huge field of application that barely caught the interest of the machine learning community is proteomics, that is, the quantitative study of the protein content of cells and tissues. Technologies such as tandem mass spectrometry to monitor the protein content (proteins or characteristic fragments of proteins) of a biological sample are now well developed, and the classification of tissues from these data is a future potential application of SVM (Wagner, Naik, & Pothen, 2003; Wu et al., 2003). Applications in toxicogenomics (Steiner et al., 2004), chemogenomics (Bao & Sun, 2002; Bock & Gough, 2002), and the analysis of single-nucleotide polymorphisms (Listgarten et al., 2004; Yoon, Song, Hong, & Kim, 2003) are also promising applications for which the capacity of SVM to classify high-dimensional data has only started to be exploited.

SEQUENCE CLASSIFICATION

The various genome sequencing projects have produced huge amounts of sequence data that need to be analyzed. In particular, the urgent need for methods to automatically process, segment, annotate, and classify various sequence data has triggered the fast development of numerous algorithms for strings. In this context, the possibility offered by kernel methods to process any type of data as soon as a kernel for the data to be processed is available has been quickly exploited to offer the power of state-of-the-art machine learning algorithms to sequence processing.

Problems that arise in computational biology consist of processing either sets of sequences of a fixed length, or sets of sequences with variable lengths. From a technical point of view, the two problems differ significantly: While there are natural ways to encode fixed-length sequences as fixed-length vectors, making them amenable to processing by most learning algorithms, manipulating variable-length sequences is less obvious. In both cases, many successful applications of SVM have been reported, combining ingenious developments of string kernels, sometimes specifically adapted to a given classification task, with the power of SVM.

Kernels for Fixed-Length Sequences

Problems involving the classification of fixed-length sequences appear typically when one wants to predict a property along a sequence, such as the local structure or solvent accessibility along a protein sequence. In that case, indeed, a common approach is to use a moving window, that is, to predict the property at each position independently from the others, and to base the prediction only on the nearby sequence contained in a small window around the site of interest. More formally, this requires the construction of predictive models that take a sequence of fixed length as input to predict the property of interest, the length of the sequences being exactly the width of the window.

To fix notations, let us denote by p the common length of the sequences, and by a typical sequence, where each x_i is a letter from the alphabet, for example, an amino acid. The most natural way to transform such a sequence into a vector of fixed length is to first encode each letter itself into a vector of fixed length , and then to concatenate the codes of the successive letters to obtain a vector of size $n = pl$ for the whole sequence. A simple code for letters is the following so-called sparse encoding: The size of the alphabet is denoted by α, and the i the letter of the alphabet is encoded as a vector of dimension α containing only zeros, except for the i the dimension that is set to l. For example, in the case of nucleotide sequences with alphabet (A,C,G,T), the codes for A, C, G, and T would respectively be (1,0,0,0), (0,1,0,0), (0,0,1,0), and (0,0,0,1), while the code for the sequence of length 3 AGT would be (1,0,0,0,0,0,1,0,0,0,0,1). Several more evolved codes for single letters have also been proposed. For example, if one has a prior matrix of pairwise similarities between letters, such as widely used similarity matrices between amino acids, it is possible to replace the 0/1 sparse encoding of a given letter by the vector of similarity with other letters; hence, the A in the previous example could, for instance, be represented by the vector (1,0,0.5,0) to emphasize one's belief that A and G share some similarity. This is particularly relevant for biological sequences where mutations of single letters to similar letters are very common. Alternatively, instead of using a prior matrix of similarity, one can automatically align the sequence of interest to similar sequences in a large sequence database, and encode each position by the frequency of each letter in the alignment. As a trivial example, if our previous sequence AGT was found to be aligned to the following sequences, AGA, AGC, CGT, and ATT, then it could be encoded by the vector (0.8,0.2,0,0,0,0,0.8,0.2,0.2,0.2,0,0.6), corresponding to the respective frequencies of each letter at each position.

In terms of a kernel, it is easy to see that the inner product between sparsely encoded sequences is the number of positions with identical letters. In this representation, any linear classifier, such as that learned by a linear SVM, associates a weight to each feature, that is, to each letter at each position, and the score of a sequence is the sum of the scores of its letters. Such a classifier is usually referred to as a position-specific score matrix in bioinformatics. Similar interpretations can be given for other letter encodings. An interesting extension of these linear kernels for sequences is to raise them to some small power d; in that case, the dimension of the feature space used by kernel methods increases, and the new features correspond to all products of d original features. This is particularly appealing for sparse encoding because a product of d binary factors is a binary variable equal to 1 if and only if all factors are 1, meaning that the features created by the sparse encoding to the power d exactly indicate the simultaneous presence of up to d particular letters at d particular positions. The trick to take a linear kernel to some power is therefore a convenient way to create a classifier for problems that involve the presence of several particular letters at particular positions.

A first limitation of these kernels is that they do not contain any information about the order of the letters: They are, for example, left unchanged if the letters in all sequences are shuffled according to any given permutation. Several attempts to include ordering information have been proposed. For example, Rätsch, Sonnenburg, and Schölkopf (2005) replace the local encoding of single letters by a local encoding of several consecutive letters; Zien, Rätsch, Mika, Schölkopf, Lengauer, and Müller (2000) propose an ingenious variant to the polynomial kernel in order to restrict the feature space to products of features at nearby positions only.

A second limitation of these kernels is that the comparison of two sequences only involves the comparison of features at identical positions. This can be problematic in the case of biological sequences where the insertion of deletions of letters is common, resulting in possible shifts within a window. This problem led Meinicke, Tech, Morgenstem, and Merkl (2004) to propose a kernel that incorporates a comparison of features at nearby positions using the following trick: If a feature f (e.g., binary or continuous) appears at position i in the first sequence, and a feature g appears at position j in the second sequence, then the kernel between the two sequences is increased by , where is a basic kernel between the features f and g such as the simple product, and σ is a parameter that controls the range at which features are compared. When σ is chosen very large, then one recovers the classical kernels obtained by comparing only identical positions($i = j$); the important point here is that for smaller values of σ, features can contribute positively even though they might be located at different positions on the sequences.

The applications of kernels for fixed-length sequences to solve problems in computational biology are already numerous. For example, they have been widely used to predict local properties along protein sequences using a moving window, such as secondary structure (Guermeur, Lifschitz, & Vert, 2004; Hua & Sun, 2001a), disulfide bridges involving cysteines (Chen, in, Lin, & Hwang, 2004; Passerini & Frasconi, 2004), phosphorylation sites (Kim, Lee, oh, Kimm, & Koh, 2004), interface residues (Res, Mihalek, & Lichtarge, 2005; Yan, Dobbs, & Honavar, 2004), and solvent accessibility (Yuan, Burrage, & Mattick, 2002). Another important field of application is the annotation of DNA using fixed-length windows centered on a candidate point of interest as an input to a classifier to detect translation initiation sites (Meinicke et al., 2004; Zien et al., 2000), splice sites (Degroeve, Saeys, De Baets, Rouze, & Van de Peer, 2005; Rätsch et al., 2005), or binding sites of transcription factors (O'Flanagan, Paillard, Lavery, & Sengupta, 2005; Sharan & Myers, 2005). The recent interest in short RNA such as antisense oligonucleotides or small interfering RNAs for the sequence-specific knockdown of messenger RNAs has also resulted in several works involving the classification of such sequences, which typically have a fixed length by nature (Camps-Valls, Chalk, Serrano-López, Martin-Guerrero, & Sonnhammer, 2004; Teramoto, Aoki, Kimura, & Kanaoka, 2005). Another important application field for these methods is immunoinformatics, including the prediction of peptides that can elicit an immune response (Bhasin & Raghava, 2004; Dönnes & Elofsson, 2002), or the classification of immunoglobulins collected from ill patients (Yu, Zavaljevski, Stevens, Yackovich, & Reifman, 2005; Zavaljevski, Stevens, & Reifman, 2002). In most of these applications, SVM lead to comparable if not better prediction accuracy than competing state-of-the-art methods such as neural networks.

Kernels for Variable-Length Sequences

Many problems in computational biology involve sequences of different lengths. For example, the automatic functional or structural annotation of genes found in sequenced genomes requires the processing of amino-acid sequences with no fixed length. Learning from variable-length sequences is a more challenging problem than learning from fixed-length sequences because there is no natural way to transform a variable-length string into a vector. For kernel methods, this issue boils down to the problem of defining kernels for variable-length strings, a topic that has deserved a lot of attention in the last few years and has given rise to a variety of ingenious solutions. As summarized in Table 1, three main approaches have been followed in the process of kernel design from strings: (a) computing an inner product in an explicitly defined feature space, (b) deriving a kernel from probabilistic models on strings, and (c) adapting widely used measures of similarity. These general strategies, surveyed in more detail in this section, are relevant beyond the specific problem of designing kernels for biological sequences. For example, we point out below strong similarities between some of these kernels and kernels developed for speech-recognition applications (surveyed in more detail in the chapters by Picone, Ganapathiraju, and Hamaker; and Wan).

The most common approach to make a kernel for strings, as for many other types of data, is to design explicitly a set of numerical features that can be extracted from strings, and then to form a kernel as a dot product between the resulting feature vectors. As an example, Leslie, Eskin, and Noble (2002) represent a sequence by the vector of counts of occurrences of all possible k-mers in the sequence for a given integer k, effectively

Table 1. A typology of kernels for variable-length biological sequences

Strategy	Example
Define a (possibly high-dimensional) feature space of interest	- Physico-chemical kernels (Wang et al., 2004 ; Zhang et al., 2003a) - Spectrum, mismatch kernels (Leslie et al., 2002, 2004) - Pairwise, motif kernels (Logan et al., 2001; Liao & Noble, 2003; Ben-Hur & Brutlag, 2003)
Derive a kernel from a generative model	- Fisher kernel (Jaakkola et al., 2000) - Pair HMM kernel (Watkins, 2000 ; Haussler, 1999) - Mutual information kernels (Cuturi & Vert, 2005) - Marginalized kernels (Tsuda et al., 2002; Kin et al., 2002; Kashima et al., 2004 ; Vert et al., 2006)
Derive a kernel from a measure of similarity	- Local alignment kernels (Saigo et al., 2004 ; Vert et al., 2004)

resulting in a vector of dimension a^k, where a is the size of the alphabet. As an example, the sequence AACGTCACGAA over the alphabet (A,C,G,T) is represented by the 16-dimensional vector (2,2,0,0,1,0,2,0,1,0,0,1,0,1,0,0) for $k = 2$; here, the features are the counts of occurrences of each 2-mer AA, AC, ..., TG, TT lexicographically ordered. The resulting spectrum kernel between this sequence and the sequence ACGAAA, defined as the linear product between the two 16-dimensional representation vectors, is equal to 9. It should be noted that although the number of possible k-mers easily reaches the order of several thousands as soon as k is equal to 3 or 4, the classification of sequences by SVM in this high-dimensional space results in fairly good results. A major advantage of the spectrum kernel is its fast computation; indeed, the set of k-mers appearing in a given sequence can be indexed in linear time in a tree structure, and the inner product between two vectors is linear with respect to the nonzero coordinates, that is, at most linear in the total lengths of the sequences. Several variants to the basic spectrum kernel have also been proposed, including, for example, kernels based on counts of k-mers appearing with up to a few mismatches in the sequences (Leslie, Eskin, Cohen, Weston, & Noble, 2004).

Another natural approach to represent variable-length strings by fixed-length numerical vectors is to replace each letter by one or several numerical features, such as physicochemical properties of amino acids, and then to extract features from the resulting variable-length numerical time series using classical signal processing techniques such as Fourier transforms (Wang, Yang, Liu, Xu, & Chou, 2004) or autocorrelation analysis (S.-W. Zhang, Pan, Zhang, Zhang, & Wang, 2003). For example, if denote P numerical features associated with the successive letters of a sequence of length P, then the autocorrelation function r_j for a given $j > 0$ is defined by:

$$r_j = \frac{1}{n-j}\sum_{i=1}^{n-j} f_i f_{i+j}.$$

One can then keep a fixed number of these coefficients, for example , and create an n-dimensional vector to represent each sequence.

Finally, another popular approach to design features and therefore kernels for biological sequences is to "project" them onto a fixed dictionary of sequences or motifs using classical similarity measures, and to use the resulting vector of similarities as a feature vector. For example, Logan, Moreno, Suzek, Weng, and Kasif (2001) represent each sequence by a 10,000-dimensional vector indicating the presence of 10,000 motifs of the BLOCKS database; similarly, Ben-Hur and Brutlag (2003) use a vector that indicates the presence or absence of about 500,000 motifs in the eMOTIF database, requiring the use of a tree structure to compute efficiently the kernel without explicitly storing the 500,000 features. Liao and Noble (2003) represent each sequence by a vector of sequence similarities with a fixed set of sequences.

A second general strategy that has been followed for kernel design is to derive kernel functions from probabilistic models. This strategy has been particularly motivated by the fact that, before the interest in string kernels grew, a number of ingenious probabilistic models had been defined to represent biological sequences or families of sequences, including, for example, Markov and hidden Markov models for protein sequences, and stochastic context-free grammars for RNA sequences (Durbin, Eddy, Krogh, & Mitchison, 1998). Several authors have therefore explored the possibility to use such models to make kernels, starting with the seminal work of T. Jaakkola, Diekhans, and Haussler (2000) that introduced the Fisher kernel. The Fisher kernel is a general method to extract a fixed number of features from any data **x** for which a parametric probabilistic model P_θ is defined. Here, θ represents a continuous n-dimensional vector of parameters for the probabilistic model, such as transition and emission probabilities for a hidden Markov model, and each P_θ is a probability distribution. Once a particular parameter θ_0 is chosen to fit a given set of objects, for example, by maximum likelihood, then an n-dimensional feature vector for each individual object **x** can be extracted by taking the gradient in the parameter space of the log likelihood of the point:

$$\phi(\mathbf{x}) = \nabla_\theta \log P_\theta(\mathbf{x}).$$

The intuitive interpretation of this feature vector, usually referred to as the Fisher score in statistics, is that it represents how changes in the n parameters affect the likelihood of the point **x**. In other words, one feature is extracted for each parameter of the model; the particularities of the data point are seen from the eyes of the parameters of the probabilistic model. The Fisher kernel is then obtained as the dot product of these n-dimensional vectors, eventually multiplied by the inverse of the Fisher information matrix to render it independent of the parametrization of the model.

A second line of thought to make a kernel out of a parametric probabilistic model is to use the concept of mutual information kernels (Seeger, 2002), that is, kernels of the form:

$$\kappa(\mathbf{x}, \mathbf{z}) = \int P_\theta(\mathbf{x}) P_\theta(\mathbf{z}) d\mu(\theta),$$

where $d\mu$ is a prior distribution on the parameter space. Here, the features correspond to the likelihood of the objects under all distributions of the probabilistic model; objects are considered similar when they have large likelihood under similar distributions. An important difference between the kernels seen so far is that here, no explicit extraction of finite-dimensional vectors can be performed. Hence, for practical applications, one must chose probabilistic models that allow the computation of the integral above. This was carried out by Cuturi and Vert (2005) who present a family of variable-length Markov models for strings and an algorithm to perform the exact integral over parameters and models in the same time, resulting in a string kernel with linear complexity in time and memory with respect to the total length of the sequences.

Alternatively, many probabilistic models for biological sequences, such as hidden Markov models, involve a hidden variable that is marginalized over to obtain the probability of a sequence, which be written as:

$$P(\mathbf{x}) = \sum_h P(\mathbf{x}, \mathbf{h}).$$

For such distributions, Tsuda, Kin, and Asai (2002) introduced the notion of a marginalized kernel, obtained by marginalizing a kernel for the complete variable over the hidden variable. More precisely, assuming that a kernel κ_0 for objects of the form (**x**,**h**) is defined, the marginalized kernel for observed objects *X* is given by:

$$\kappa(\mathbf{x}_1, \mathbf{x}_2) = \sum_{\mathbf{h}_1, \mathbf{h}_2} \kappa_0((\mathbf{x}_1, \mathbf{h}_1), (\mathbf{x}_2, \mathbf{h}_2)) P(\mathbf{h}_1 \mid \mathbf{x}_1) P(\mathbf{h}_2 \mid \mathbf{x}_2).$$

In order to motivate this definition with a simple example, let us consider a hidden Markov model with two possible hidden states to model sequences with two possible regimes, such as introns and exons in eukaryotic genes. In that case, the hidden variable corresponding to a sequence **x** of length *P* is a binary sequence **h** of length *P* describing the states along the sequence. For two sequences $\mathbf{x}_1$ and $\mathbf{x}_2$, if the correct hidden states $\mathbf{h}_1$ and $\mathbf{h}_2$ were known, such as the correct decomposition into introns and exons, then it would make sense to define a kernel taking into account the specific decomposition of the sequences into two regimes; for example, the kernel for complete data could be a spectrum kernel restricted to the exons, that is, to positions with a particular state. Because the actual hidden states are not known in practice, the marginalization over the hidden state of this kernel using an adequate probabilistic model can be interpreted as an attempt to apply the kernel for complete data by guessing the hidden variables. Similar to the mutual information kernel, marginalized kernels can often not be expressed as inner products between feature vectors, and they require computational tricks to be computed. Several beautiful examples of such kernels for various probabilistic models have been worked out, including hidden Markov models for sequences (Tsuda et al.; Vert, Thurman, & Noble, 2006), stochastic context-free grammars for RNA sequences (Kin, Tsuda, & Asai, 2002), and random walk models on graphs for molecular structures (Kashima, Tsuda, & Inokuchi, 2004).

Following a different line of thought, Haussler (1999) introduced the concept of convolution kernels for objects that can be decomposed into subparts, such as sequences or trees. For example, the concatenation of two strings $\mathbf{x}_1$ and $\mathbf{x}_2$ results in another string $\mathbf{x} = \mathbf{x}_1 \mathbf{x}_2$. If two initial string kernels $\kappa_1$and κ_2 are chosen, then a new string kernel is obtained by the convolution of the initial kernels following the equation:

$$\kappa(\mathbf{x}, \mathbf{z}) = \sum_{\mathbf{x}_1 \mathbf{x}_2 = \mathbf{x}} \sum_{\mathbf{z}_1 \mathbf{z}_2 = \mathbf{z}} \kappa_1(\mathbf{x}_1, \mathbf{z}_1) \kappa_2(\mathbf{x}_2, \mathbf{z}_2).$$

Here, the sum is over all possible decompositions of **x** and **z** into two concatenated subsequences. The rationale behind this approach is that it allows the combination of different kernels adapted to different parts of the sequences, such as introns and exons or gaps and aligned residues in alignment, without knowing the exact segmentation of the sequences. Besides proving that the convolution of two kernels is a valid kernel, Haussler (1999) gives several examples of convolution kernels relevant for biological sequences; for example, he shows that the joint probability *P*(**x**,**z**) of two sequences under a pair hidden Markov model (HMM) is a valid kernel under mild assumptions, a property that was also proved independently by Watkins (2000).

Finally, a third general strategy to design kernels for strings is to start from classical measures of similarity between strings and turn them into kernels. In the case of biological sequences, for example, a scheme known as Smith-Waterman local alignment score (Smith & Waterman, 1981) is widely used to compare sequences. Saigo, Vert, Ueda, and Akutsu (2004) show that a slight modification of this measure of similarity is a valid kernel, called the local alignment kernel. In fact, the proof of the positive definiteness of the new kernels builds on the study of convolution kernels presented above (Vert, Saigo, & Akutsu, 2004). Interestingly, the Smith-Waterman alignment score between biological sequences optimizes an alignment between sequences by dynamic programming, just like dynamic time warping (DTW) algorithms compare variable-length sequences in speech recognition. DTW kernels that violate the positive semidefiniteness assumption (surveyed in the chapter by Wan) have been proposed for speaker-recognition applications; adapting the local alignment kernel to this context, that is, replacing a Viterbi optimization with a forward summation in the dynamic programming algorithm, would result in a valid DTW kernel for speech applications. The local alignment kernel gives excellent results on the problem of detecting remote homologs of proteins, suggesting that combining domain knowledge (in the form of a relevant measure of similarity turned into a kernel) with kernel methods is a promising research direction.

These kernels for variable-length sequences have been widely applied, often in combination with SVM, to various classification tasks in computational biology. Examples include the prediction of protein structural or functional classes from their primary sequence (Cai, Wang, Sun, & Chen, 2003; Ding & Dubchak, 2001; T. Jaakkola et al., 2000; Karchin, Karplus, & Haussler, 2002; Vert et al., 2004), the prediction of the subcellular localization of proteins (Hua & Sun, 2001b; Matsuda, Vert, Saigo, Ueda, Toh, & Akutsu, 2005; Park & Kanehisa, 2003), the classification of transfer RNA (Kin et al., 2002) and noncoding RNA (Karklin, Meraz, & Holbrook, 2005), the prediction of pseudoexons and alternatively spliced exons (Dror, Sorek, & Shamir, 2005; X. H.-F. Zhang, Heller, Hefter, Leslie, & Chasin, 2003b), the separation of mixed plant-pathogen expressed sequence tag (EST) collections (Friedel, Jahn, Sommer, Rudd, Mewes, & Tetko, 2005), the classification of mammalian viral genomes (Rose, Turkett, Oroian, Laegreid, & Keele, 2005), and the prediction of ribosomal proteins (Lin, Kuang, Joseph, & Kolatkar, 2002).

This short review of kernels developed for the purpose of biological sequence classification, besides highlighting the dynamism of research in kernel methods resulting from practical needs in computational biology, naturally raises the practical question of which kernel to use for a given application. Although no clear answer has emerged yet, some lessons can be learned from early studies. First, there is certainly no kernel universally better than others, and the choice of kernel should depend on the targeted application. Intuitively, a kernel for a classification task is likely to work well if it is based on features relevant to the task; for example, a kernel based on sequence alignments, such as the local alignment kernel, gives excellent results on remote homology detection problems, while a kernel based on the global content of sequences in short subsequences, such as the spectrum kernel, works well for the prediction of subcellular localization. Although some methods for the systematic selection and combination of kernels are starting to emerge (see the next section), an empirical evaluation of different kernels on a given problem seems to be the most common way to chose a kernel. Another important point to notice, besides the classification accuracy obtained with a kernel, is its computational cost. Indeed, practical applications often involve data sets of thousands or tens of thousands of sequences, and the computational cost of a method can become a critical factor in this context, in particular in an online setting. The kernels presented above differ a lot in their computational cost, ranging from fast linear-time kernels like the spectrum kernel to slower kernels like the quadratic-time local alignment kernel. The final choice of kernel for a given application often results from a trade-off between classification performance and computational burden.

OTHER APPLICATIONS AND FUTURE TRENDS

Besides the important applications mentioned in the previous sections, several other attempts to import ideas of kernel methods in computational biology have emerged recently. In this section, we highlight three promising directions that are likely to develop quickly in the near future: the engineering of new kernels, the development of methods to handle multiple kernels, and the use of kernel methods for graphs in systems biology.

More Kernels

The power of kernel methods to process virtually any sort of data as soon as a valid kernel is defined has recently been exploited for a variety of data, besides high-dimensional data and sequences. For example, Vert (2002) derives a kernel for phylogenetic profiles, that is, a signature indicating the presence or absence of each gene in all sequenced genomes. Several recent works have investigated kernels for protein 3-D structures, a topic that is likely to expand quickly with the foreseeable availability of predicted or solved structures for whole genomes (Borgwardt, Ong, Schönauer, Vishwanathan, Smola, & Kreigel, 2005; Dobson & Doig, 2003). For smaller molecules, several kernels based on planar or 3-D structures have emerged, with many potential applications in computational chemistry (Kashima et al., 2004; Mahé, Ueda, Akutsu, Perret, & Vert, 2005; Swamidass, Chen, Bruand, Phung, Ralaivola, & Baldi, 2005). This trend to develop more and more kernels, often designed for specific data and applications, is likely to continue in the future because it has proved to be a good approach to obtain efficient algorithms for real-world applications. A nice by-product of these efforts, which is still barely exploited, is the fact that any kernel can be used by any kernel method, paving the way to a multitude of applications such as clustering (Qin, Lewis, & Noble, 2003; see also the chapter by Pochet, Ojeda, De Smet, De Bie, Suykens, & De Moor) or data visualization (Komura, Nakamura, Tsutsumi, Aburatani, & Ihara, 2005).

Integration of Heterogeneous Data

Operations on kernels provide simple and powerful tools to integrate heterogeneous data or multiple kernels; this is particularly relevant in computational biology, where biological objects can typically be described by heterogeneous representations, and the availability of a large number of possible kernels for even a single representation raises the question of choice or combination of kernels. Suppose, for instance, that one wants to perform a functional classification of genes based on their sequences, expression over a series of experiments, evolutionary conservation, and position in an interaction network. A natural approach with kernel methods is to start by defining one or several kernels for each sort of data, that is, string kernels for the gene sequences, vector kernels to process the expression profiles, and so forth. The apparent heterogeneity of data types then vanishes as one simply obtains a family of kernel functions . In order to learn from all data simultaneously, the simplest approach is to define an integrated kernel as the sum of the initial kernels:

$$\kappa = \sum_{i=1}^{p} \kappa_i .$$

The rationale behind this sum is that if each kernel is a simple dot product, then the sum of dot products is equal to the dot product of the concatenated vectors. In other words, taking a sum of kernels amounts to putting all features of each individual kernel together; if different features in different kernels are relevant for a given problem, then one expects the kernel method trained on the integrated kernel to pick those relevant features. This idea was pioneered by the authors of Pavlidis, Weston, Cai, and Noble (2002), in which gene expression profiles and gene phylogenetic profiles are integrated to predict the functional classes of genes, effectively integrating evolutionary and transcriptional information.

An interesting generalization of this approach is to form a convex combination of kernels of the form:

$$\kappa = \sum_{i=1}^{p} w_i \kappa_i ,$$

where the w_i are nonnegative weights. Lanckriet, De Bie, Cristianini, Jordan and Noble (2004) propose a general framework based on semidefinite programming to optimize the weights and learn a discrimination function for a given classification task simultaneously. Promising empirical results on gene functional classification show that by integrating several kernels, better results can be obtained than with each individual kernel.

Finally, other kernel methods can be used to compare and search for correlations between heterogeneous data. For example, Vert and Kanehisa (2003) propose to use a kernelized version of canonical correlation analysis (CCA) to compare gene expression data on the one hand with the position of genes in the metabolic network on the other hand. Each type of data is first converted into a kernel for genes, the information about gene positions in the metabolic network being encoded with the so-called diffusion kernel (Kondor & Vert, 2004). These two kernels define embeddings of the set of genes into two Euclidean spaces, in which correlated directions are detected by CCA. It is then shown that the directions detected in the feature space of the diffusion kernel can be interpreted as clusters in the metabolic network, resulting in a method to monitor the expression patterns of metabolic pathways.

Kernel Methods in Systems Biology

Another promising field of research where kernel methods can certainly contribute is systems biology, which roughly speaking focuses on the analysis of biological systems of interacting molecules, in particular, biological networks.

A first avenue of research is the reconstruction of biological networks from high-throughput data. For example, the prediction of interacting proteins to reconstruct the interaction network can be posed as a binary classification problem—given a pair of proteins, do they interact or not?—and can therefore be tackled with SVM as soon as a kernel between pairs of proteins is defined. However, the primary data available are about individual proteins. In order to use SVM in this context, it is therefore natural to try to derive kernels for pairs of proteins from kernels for single proteins. This has been carried out, for example, by Bock and Gough (2001) who characterize each protein by a vector and concatenate two such individual vectors to represent a protein pair. Observing that there is usually no order in a protein pair, Martin, Roe, and Faulon (2005) and Ben-Hur and Noble (2005) propose to define a kernel between pairs (*A*,*B*) and (*C*,*D*) by the equation:

$$\kappa((A,B),(C,D)) = \kappa_0(A,C)\kappa_0(B,D) + \kappa_0(A,D)\kappa_0(B,C),$$

where κ_0 denotes a kernel for an individual protein and κ is the resulting kernel for pairs of proteins. The rationale behind this definition is that in order to match the pair (*A*,*B*) with the pair (*C*,*D*), one can either try to match *A* with *C* and *B* with *D*, or to match *A* with *D* and *B* with *C*. Reported accuracies on the problem of protein-interaction prediction are very high, confirming the potential of kernel methods in this fast-moving field.

A parallel approach to network inference from genomic data has been investigated by Yamanishi, Vert, and Kanehisa (2004), who show that learning the edges of a network can be carried out by first mapping the vertices, for example, the genes, onto a Euclidean space, and then connecting the pairs of points that are close to each other in this embedding. The problem then becomes that of learning an optimal embedding of the vertices, a problem known as distance metric learning that recently caught the attention of the machine learning community and for which several kernel methods exist (Vert & Yamanishi, 2005).

Finally, several other emerging applications in systems biology, such as inference on networks (Tsuda & Noble, 2004) or the classification of networks (Middendorf et al., 2004), are likely to be subject to increasing attention in the future due to the growing interest and amount of data related to biological networks.

CONCLUSION

This brief survey, although far from being complete, highlights the impressive advances in the applications of kernel methods in computational biology in the last 5 years. More than just importing well-established algorithms to a new application domain, biology has triggered the development of new algorithms and methods, ranging from the engineering of various kernels to the development of new methods for learning from multiple kernels and for feature selection. The widespread diffusion of easy-to-use SVM software and the ongoing integration of various kernels and kernel methods in major computing environments for bioinformatics are likely to foster again the use of kernel methods in computational biology as long as they will provide state-of-the-art methods for practical problems. Many questions remain open regarding, for example, the automatic choice and integration of kernels, the possibility to incorporate prior knowledge in kernel methods, and the extension of kernel methods to more general kernels that are positive definite, suggesting that theoretical developments are also likely to progress quickly in the near future.

REFERENCES

Aebersold, R., & Mann, M. (2003). Mass spectrometry-based proteomics. *Nature, 422*(6928), 198-207.

Aliferis, C., Hardin, D., & Massion, P. (2002). Machine learning models for lung cancer classification using array comparative genomic hybridization. In *Proceedings of the 2002 American Medical Informatics Association (AMIA) Annual Symposium* (pp. 7-11).

Ambroise, C., & McLachlan, G. (2002). Selection bias in gene extraction on the basis of microarray gene-expression data. In *Proceedings of the National Academy of Science USA, 99*(10), 6562-6566.

Bao, L., & Sun, Z. (2002). Identifying genes related to drug anticancer mechanisms using support vector machine. *FEBS Letters, 521*, 109-114.

Ben-Dor, A., Bruhn, L., Friedman, N., Nachman, I., Schummer, M., & Yakhini, Z. (2000). Tissue classification with gene expression profiles. *Journal of Computational Biology, 7*(3-4), 559-583.

Ben-Hur, A., & Brutlag, D. (2003). Remote homology detection: A motif based approach. *Bioinformatics, 19*(Suppl. 1), i26-i33.

Ben-Hur, A., & Noble, W. S. (2005). Kernel methods for predicting protein-protein interactions. *Bioinformatics, 21*(Suppl. 1), i38-i46.

Bhasin, M., & Raghava, G. P. S. (2004). Prediction of CTL epitopes using QM, SVM, and ANN techniques. *Vaccine, 22*(23-24), 3195-3204.

Bock, J. R., & Gough, D. A. (2001). Predicting protein-protein interactions from primary structure. *Bioinformatics, 17*(5), 455-460.

Bock, J. R., & Gough, D. A. (2002). A new method to estimate ligand-receptor energetics. *Mol Cell Proteomics, 1*(11), 904-910.

Borgwardt, K. M., Ong, C. S., Schönauer, S., Vishwanathan, S. V. N., Smola, A. J., & Kriegel, H.-P. (2005). Protein function prediction via graph kernels. *Bioinformatics, 21*(Suppl. 1), i47-i56.

Cai, C., Wang, W., Sun, L., & Chen, Y. (2003). Protein function classification via support vector machine approach. *Mathematical Biosciences, 185*(2), 111-122.

Camps-Valls, G., Chalk, A., Serrano-Lopez, A., Martin-Guerrero, J., & Sonnhammer, E. (2004). Profiled support vector machines for antisense oligonucleotide efficacy prediction. *BMC Bioinformatics, 5*(135), 135.

Chen, Y., Lin, Y., Lin, C., & Hwang, J. (2004). Prediction of the bonding states of cysteines using the support vector machines based on multiple feature vectors and cysteine state sequences. *Proteins, 55*(4), 1036-1042.

Cuturi, M., & Vert, J.-P. (2005). The context-tree kernel for strings. *Neural Networks, 18*(4), 1111-1123.

Degroeve, S., Saeys, Y., De Baets, B., Rouze, P., & Van de Peer, Y. (2005). SpliceMachine: Predicting splice sites from high-dimensional local context representations. *Bioinformatics, 21*, 1332-1338.

DeRisi, J. L., Iyer, V. R., & Brown, P. O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science, 278*(5338), 680-686.

Ding, C., & Dubchak, I. (2001). Multi-class protein fold recognition using support vector machines and neural networks. *Bioinformatics, 17*, 349-358.

Dobson, P., & Doig, A. (2003). Distinguishing enzyme structures from non-enzymes without alignments. *Journal of Molecular Biology, 330*(4), 771-783.

Dönnes, P., & Elofsson, A. (2002). Prediction of MHC class I binding peptides, using SVMHC. *BMC Bioinformatics, 3*(1), 25.

Dror, G., Sorek, R., & Shamir, R. (2005). Accurate identification of alternatively spliced exons using support vector machine. *Bioinformatics, 21*(7), 897-901.

Durbin, R., Eddy, S., Krogh, A., & Mitchison, G. (1998). *Biological sequence analysis: Probabilistic models of proteins and nucleic acids.* Cambridge University Press.

Friedel, C. C., Jahn, K. H. V., Sommer, S., Rudd, S., Mewes, H. W., & Tetko, I. V. (2005). Support vector machines for separation of mixed plant-pathogen EST collections based on codon usage. *Bioinformatics, 21*, 1383-1388.

Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D. W., Schummer, M., & Haussler, D. (2000). Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics, 16*(10), 906-914.

Guermeur, Y., Lifschitz, A., & Vert, R. (2004). A kernel for protein secondary structure prediction. In B. Schölkopf, K. Tsuda, & J. Vert (Eds.), *Kernel methods in computational biology* (pp. 193-206). MIT Press.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning, 46*(1/3), 389-422.

Haussler, D. (1999). *Convolution kernels on discrete structures* (Tech. Rep. No. UCSC-CRL-99-10). Santa Cruz: University of California.

Hua, S., & Sun, Z. (2001a). A novel method of protein secondary structure prediction with high segment overlap measure: Support vector machine approach. *Journal of Molecular Biology, 308*(2), 397-407.

Hua, S., & Sun, Z. (2001b). Support vector machine approach for protein subcellular localization prediction. *Bioinformatics, 17*(8), 721-728.

Huang, T. M., & Kecman, V. (2005). Gene extraction for cancer diagnosis by support vector machines: An improvement. *Artificial Intelligence in Medicine (Special Issue on Computational Intelligence Techniques in Bioinformatics), 35*, 185-194.

International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature, 409*(6822), 860-921.

Jaakkola, T., Diekhans, M., & Haussler, D. (2000). A discriminative framework for detecting remote protein homologies. *Journal of Computational Biology, 7*(1,2), 95-114.

Jaakkola, T. S., Diekhans, M., & Haussler, D. (1999). Using the Fisher kernel method to detect remote protein homologies. In *Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology*, 149-158.

Karchin, R., Karplus, K., & Haussler, D. (2002). Classifying G-protein coupled receptors with support vector machines. *Bioinformatics, 18*, 147-159.

Karklin, Y., Meraz, R. F., & Holbrook, S. R. (2005). Classification of non-coding RNA using graph representations of secondary structure. In *Pacific Symposium on Biocomputing* (pp. 4-15).

Kashima, H., Tsuda, K., & Inokuchi, A. (2004). Kernels for graphs. In B. Schölkopf, K. Tsuda, & J. Vert (Eds.), *Kernel methods in computational biology* (pp. 155-170). MIT Press.

Kim, J. H., Lee, J., Oh, B., Kimm, K., & Koh, I. (2004). Prediction of phosphorylation sites using SVMs. *Bioinformatics, 20*(17), 3179-3184.

Kin, T., Tsuda, K., & Asai, K. (2002). Marginalized kernels for RNA sequence data analysis. In R. Lathtop, K. Nakai, S. Miyano, T. Takagi, & M. Kanehisa (Eds.), *Genome informatics 2002* (pp. 112-122). Universal Academic Press.

Komura, D., Nakamura, H., Tsutsumi, S., Aburatani, H., & Ihara, S. (2005). Multidimensional support vector machines for visualization of gene expression data. *Bioinformatics, 21*(4), 439-444.

Kondor, R., & Vert, J.-P. (2004). Diffusion kernels. In B. Schölkopf, K. Tsuda, & J. Vert, (Eds.), *Kernel methods in computational biology* (pp. 171-192). MIT Press.

Krishnapuram, B., Carin, L., & Hartemink, A. (2004). Joint classifier and feature optimization for comprehensive cancer diagnosis using gene expression data. *Journal of Computational Biology, 11*(2-3), 227-242.

Lanckriet, G. R. G., De Bie, T., Cristianini, N., Jordan, M. I., & Noble, W. S. (2004). A statistical framework for genomic data fusion. *Bioinformatics*, *20*(16), 2626-2635.

Leslie, C., Eskin, E., & Noble, W. (2002). The spectrum kernel: A string kernel for SVM protein classification. In *Proceedings of the Pacific Symposium on Biocomputing 2002* (pp. 564-575).

Leslie, C. S., Eskin, E., Cohen, A., Weston, J., & Noble, W. S. (2004). Mismatch string kernels for discriminative protein classification. *Bioinformatics*, *20*(4), 467-476.

Li, T., Zhang, C., & Ogihara, M. (2004). A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression. *Bioinformatics, 20*(15), 2429-2437.

Liao, L., & Noble, W. (2003). Combining pairwise sequence similarity and support vector machines for detecting remote protein evolutionary and structural relationships. *Journal of Computational Biology, 10*(6), 857-868.

Lin, K., Kuang, Y., Joseph, J. S., & Kolatkar, P. R. (2002). Conserved codon composition of ribosomal protein coding genes in Escherichia coli, mycobacterium tuberculosis and saccharomyces cerevisiae: Lessons from supervised machine learning in functional genomics. *Nucleic Acids Research, 30*(11), 2599-2607.

Listgarten, J., Damaraju, S., Poulin, B., Cook, L., Dufour, J., Driga, A., et al. (2004). Predictive models for breast cancer susceptibility from multiple single nucleotide polymorphisms. *Clinical Cancer Research*, *10*(8), 2725-2737.

Logan, B., Moreno, P., Suzek, B., Weng, Z., & Kasif, S. (2001). *A study of remote homology detection* (Tech. Rep. No. CRL 2001/05). Compaq Cambridge Research laboratory.

Mahé, P., Ueda, N., Akutsu, T., Perret, J.-L., & Vert, J.-P. (2005). Graph kernels for molecular structure-activity relationship analysis with support vector machines. *Journal of Chemical Information and Modeling, 45*(4), 939-951.

Martin, S., Roe, D., & Faulon, J.-L. (2005). Predicting protein-protein interactions using signature products. *Bioinformatics, 21*(2), 218-226.

Matsuda, A., Vert, J.-P., Saigo, H., Ueda, N., Toh, H., & Akutsu, T. (2005). A novel representation of protein sequences for prediction of subcellular location using support vector machines. *Protein Science, 14*(11), 2804-2813.

Meinicke, P., Tech, M., Morgenstern, B., & Merkl, R. (2004). Oligo kernels for datamining on biological sequences: A case study on prokaryotic translation initiation sites. *BMC Bioinformatics, 5*, 169.

Meireles, S., Carvalho, A., Hirata, R., Montagnini, A., Martins, W., Runza, F., et al. (2003). Differentially expressed genes in gastric tumors identified by cDNA array. *Cancer Letters, 190*(2), 199-211.

Middendorf, M., Ziv, E., Adams, C., Hom, J., Koytcheff, R., Levovitz, C., et al. (2004). Discriminative topological features reveal biological network mechanisms. *BMC Bioinformatics, 5*, 181.

Model, F., Adorjan, P., Olek, A., & Piepenbrock, C. (2001). Feature selection for DNA methylation based cancer classification. *Bioinformatics, 17*(Supp. 1), S157-S164.

Moler, E. J., Chow, M. L., & Mian, I. S. (2000). Analysis of molecular profile data using generative and discriminative methods. *Physiological Genomics, 4*(2), 109-126.

Mukherjee, S., Tamayo, P., Mesirov, J. P., Slonim, D., Verri, A., & Poggio, T. (1998). *Support vector machine classification of microarray data* (Tech. Rep. No. 182).

O'Donnell, R. K., Kupferman, M., Wei, S. J., Singhal, S., Weber, R., O'Malley, B., et al. (2005). Gene expression signature predicts lymphatic metastasis in squamous cell carcinoma of the oral cavity. *Oncogene, 24*(7), 1244-1251.

O'Flanagan, R. A., Paillard, G., Lavery, R., & Sengupta, A. M. (2005). Non-additivity in protein-DNA binding. *Bioinformatics, 21*(10), 2254-2263.

Park, K.-J., & Kanehisa, M. (2003). Prediction of protein subcellular locations by support vector machines using compositions of amino acids and amino acid pairs. *Bioinformatics, 19*(13), 1656-1663.

Passerini, A., & Frasconi, P. (2004). Learning to discriminate between ligand-bound and disulfide-bound cysteines. *Protein Engineering Design and Selection, 17*(4), 367-373.

Pavlidis, P., Weston, J., Cai, J., & Noble, W. (2002). Learning gene functional classifications from multiple data types. *Journal of Computational Biology, 9*(2), 401-411.

Pochet, N., De Smet, F., Suykens, J. A. K., & De Moor, B. L. R. (2004). Systematic benchmarking of microarray data classification: Assessing the role of non-linearity and dimensionality reduction. *Bioinformatics, 20*(17), 3185-3195.

Qin, J., Lewis, D. P., & Noble, W. S. (2003). Kernel hierarchical gene clustering from microarray expression data. *Bioinformatics, 19*(16), 2097-2104.

Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, C., Angelo, M., et al. (2001). Multiclass cancer diagnosis using tumor gene expression signatures. In *Proceedings of the National Academy of Science USA, 98*(26), 15149-15154.

Rätsch, G., Sonnenburg, S., & Schölkopf, B. (2005). RASE: Recognition of alternatively spliced exons in c. elegans. *Bioinformatics, 21*(Suppl. 1), i369-i377.

Res, I., Mihalek, I., & Lichtarge, O. (2005). An evolution based classifier for prediction of protein interfaces without using protein structures. *Bioinformatics, 21*(10), 2496-2501.

Rose, J. R., Turkett, W. H. J., Oroian, I. C., Laegreid, W. W., & Keele, J. (2005). Correlation of amino acid preference and mammalian viral genome type. *Bioinformatics, 21*(8), 1349-1357.

Saigo, H., Vert, J.-P., Ueda, N., & Akutsu, T. (2004). Protein homology detection using string alignment kernels. *Bioinformatics, 20*(11), 1682-1689.

Schena, M., Shalon, D., Davis, R., & Brown, P. (1995). Quantitative monitoring of gene expression patterns with a complimentary DNA microarray. *Science, 270,* 467-470.

Schölkopf, B., Tsuda, K., & Vert, J.-P. (2004). *Kernel methods in computational biology.* MIT Press.

Seeger, M. (2002). Covariance kernels from Bayesian generative models. *Advances in Neural Information Processing Systems, 14,* 905-912.

Segal, N. H., Pavlidis, P., Antonescu, C. R., Maki, R. G., Noble, W. S., DeSantis, D., et al. (2003). Classification and subtype prediction of adult soft tissue sarcoma by functional genomics. *American Journal of Pathology, 163*(2), 691-700.

Segal, N. H., Pavlidis, P., Noble, W. S., Antonescu, C. R., Viale, A., Wesley, U. V., et al. (2003). Classification of clear-cell sarcoma as a subtype of melanoma by genomic profiling. *Journal of Clinical Oncology, 21*(9), 1775-1781.

Sharan, R., & Myers, E. W. (2005). A motif-based framework for recognizing sequence families. *Bioinformatics, 21*(Suppl. 1), i387-i393.

Shawe-Taylor, J., & Cristianini, N. (2004). *Kernel methods for pattern analysis.* Cambridge University Press.

Shipp, M. A., Ross, K. N., Tamayo, P., Weng, A. P., Kutok, J. L., Aguiar, R. C. T., et al. (2002). Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nature Medicine, 8*(1), 68-74.

Smith, T., & Waterman, M. (1981). Identification of common molecular subsequences. *Journal of Molecular Biology, 147,* 195-197.

Steiner, G., Suter, L., Boess, F., Gasser, R., de Vera, M. C., Albertini, S., et al. (2004). Discriminating different classes of toxicants by transcript profiling. *Environmental Health Perspectives, 112*(12), 1236-1248.

Su, Y., Murali, T., Pavlovic, V., Schaffer, M., & Kasif, S. (2003). RankGene: Identification of diagnostic genes based on expression data. *Bioinformatics, 19*(12), 1578-1579.

Swamidass, S. J., Chen, J., Bruand, J., Phung, P., Ralaivola, L., & Baldi, P. (2005). Kernels for small molecules and the prediction of mutagenicity, toxicity and anti-cancer activity. *Bioinformatics, 21*(Suppl. 1), i359-i368.

Teramoto, R., Aoki, M., Kimura, T., & Kanaoka, M. (2005). Prediction of siRNA functionality using generalized string kernel and support vector machine. *FEBS Letters, 579*(13), 2878-2882.

Tsuda, K., Kin, T., & Asai, K. (2002). Marginalized kernels for biological sequences. *Bioinformatics, 18,* S268-S275.

Tsuda, K., & Noble, W. (2004). Learning kernels from biological networks by maximizing entropy. *Bioinformatics, 20,* i326-i333.

Valentini, G. (2002). Gene expression data analysis of human lymphoma using support vector machines and output coding ensembles. *Artificial Intelligence in Medicine, 26*(3), 281-304.

Venter, J. C. e. a. (2001). The sequence of the human genome. *Science*, *291*(5507), 1304-1351.

Vert, J.-P. (2002). A tree kernel to analyze phylogenetic profiles. *Bioinformatics*, *18*, S276-S284.

Vert, J.-P., & Kanehisa, M. (2003). Extracting active pathways from gene expression data. *Bioinformatics, 19*, 238ii-234ii.

Vert, J.-P., Saigo, H., & Akutsu, T. (2004). Local alignment kernels for biological sequences. In B. Schölkopf, K. Tsuda, & J. Vert, (Eds.), *Kernel methods in computational biology* (pp. 131-154). Cambridge, MA: MIT Press.

Vert, J.-P., Thurman, R., & Noble, W. S. (2006). Kernels for gene regulatory regions. In Y. Weiss, B. Schölkopf, & J. Platt (Eds.), *Advances in Neural Information Processing Systems 18* (pp. 1401-1408). Cambridge, MA: MIT Press.

Vert, J.-P., & Yamanishi, Y. (2005). Supervised graph inference. *Advances in Neural Information Processing Systems*, *17*, 1433-1440.

Wagner, M., Naik, D., & Pothen, A. (2003). Protocols for disease classification from mass spectrometry data. *Proteomics*, *3*(9), 1692-1698.

Wang, M., Yang, J., Liu, G.-P., Xu, Z.-J., & Chou, K.-C. (2004). Weighted-support vector machines for predicting membrane protein types based on pseudo-amino acid composition. *Protein Engineering Design and Selection*, *17*(6), 509-516.

Watkins, C. (2000). Dynamic alignment kernels. In A. J. Smola, P. L. Bartlett, B. Schölkopf, & D. Schuurmans (Eds.), *Advances in large margin classifiers* (pp. 39-50). Cambridge, MA: MIT Press.

Williams, R., Hing, S., Greer, B., Whiteford, C., Wei, J., Natrajan, R., et al. (2004). Prognostic classification of relapsing favorable histology Wilms tumor using cDNA microarray expression profiling and support vector machines. *Genes Chromosomes Cancer*, *41*(1), 65-79.

Wu, B., Abbott, T., Fishman, D., McMurray, W., Mor, G., Stone, K., et al. (2003). Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data. *Bioinformatics*, *19*(13), 1636-1643.

Yamanishi, Y., Vert, J.-P., & Kanehisa, M. (2004). Protein network inference from multiple genomic data: A supervised approach. *Bioinformatics*, *20*, i363-i370.

Yan, C., Dobbs, D., & Honavar, V. (2004). A two-stage classifier for identification of protein-protein interface residues. *Bioinformatics*, *20*(Suppl. 1), i371-i378.

Yoon, Y., Song, J., Hong, S., & Kim, J. (2003). Analysis of multiple single nucleotide polymorphisms of candidate genes related to coronary heart disease susceptibility by using support vector machines. *Clinical Chemistry and Laboratory Medicine*, *41*(4), 529-534.

Yu, C., Zavaljevski, N., Stevens, F. J., Yackovich, K., & Reifman, J. (2005). Classifying noisy protein sequence data: A case study of immunoglobulin light chains. *Bioinformatics*, *21*(Suppl. 1), i495-i501.

Yuan, Z., Burrage, K., & Mattick, J. (2002). Prediction of protein solvent accessibility using support vector machines. *Proteins*, *48*(3), 566-570.

Zavaljevski, N., Stevens, F., & Reifman, J. (2002). Support vector machines with selective kernel scaling for protein classification and identification of key amino acid positions. *Bioinformatics*, *18*(5), 689-696.

Zhang, S.-W., Pan, Q., Zhang, H.-C., Zhang, Y.-L., & Wang, H.-Y. (2003). Classification of protein quaternary structure with support vector machine. *Bioinformatics*, *19*(18), 2390-2396.

Zhang, X. H.-F., Heller, K. A., Hefter, I., Leslie, C. S., & Chasin, L. A. (2003). Sequence information for the splicing of human pre-mRNA identified by support vector machine classification. *Genome Research*, *13*(12), 2637-2650.

Zien, A., Rätsch, G., Mika, S., Schölkopf, B., Lengauer, T., & Müller, K.-R. (2000). Engineering support vector machine kernels that recognize translation initiation sites. *Bioinformatics, 16*(9), 799-807.

This work was previously published in Kernal Methods in Bioengineering, Signal and Image Processing, edited by G. Camps-Valls, J. Rojo-Alvarez, and M. Martinez-Ramon, pp. 42-63, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 1.25
Information Extraction in Biomedical Literature

Min Song
Drexel University, USA

Il-Yeol Song
Drexel University, USA

Xiaohua Hu
Drexel University, USA

Hyoil Han
Drexel University, USA

INTRODUCTION

Information extraction (IE) technology has been defined and developed through the US DARPA Message Understanding Conferences (MUCs). IE refers to the identification of instances of particular events and relationships from unstructured natural language text documents into a structured representation or relational table in databases. It has proved successful at extracting information from various domains, such as the Latin American terrorism, to identify patterns related to terrorist activities (MUC-4). Another domain, in the light of exploiting the wealth of natural language documents, is to extract the knowledge or information from these unstructured plain-text files into a structured or relational form. This form is suitable for sophisticated query processing, for integration with relational databases, and for data mining. Thus, IE is a crucial step for fully making text files more easily accessible.

BACKGROUND

The advent of large volumes of text databases and search engines have made them readily available to domain experts and have significantly accelerated research on bioinformatics. With the size of a digital library commonly exceeding millions of documents, rapidly increasing, and covering a wide range of topics, efficient and automatic extraction of meaningful data and relations has become a challenging issue. To tackle this issue, rigorous studies have been carried out recently to apply IE to biomedical data. Such research efforts began to be called biomedical literature mining or text mining in bioinformatics (de Bruijn & Martin, 2002; Hirschman et al., 2002; Shatkay & Feldman, 2003). In this article, we review recent advances in applying IE techniques to biomedical literature.

MAIN THRUST

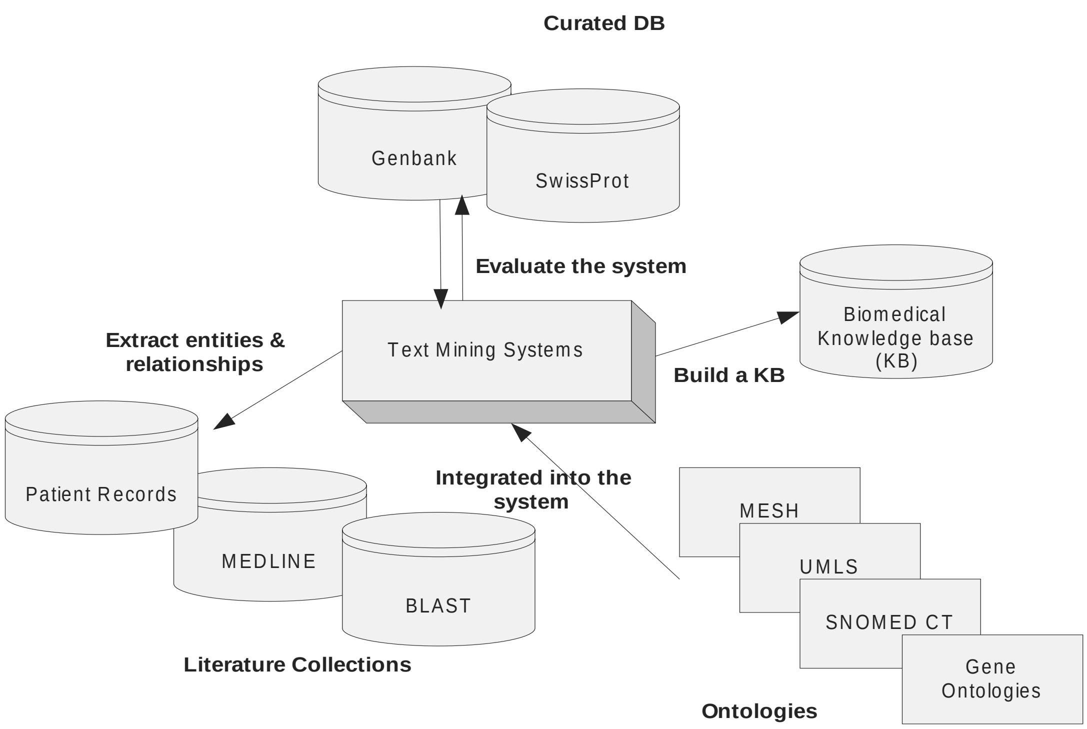
This article attempts to synthesize the works that have been done in the field. Taxonomy helps us understand the accomplishments and challenges in this emerging field. In this article, we use the following set of criteria to classify the biomedical literature mining related studies:

1. What are the target objects that are to be extracted?
2. What techniques are used to extract the target objects from the biomedical literature?
3. How are the techniques or systems evaluated?
4. From what data sources are the target objects extracted?

Target Objects

In terms of what is to be extracted by the systems, most studies can be broken into the following two major areas: (1) named entity extraction such as proteins or genes; and (2) relation extraction, such as relationships between proteins. Most of these studies adopt information extraction techniques using curated lexicon or natural language processing for identifying relevant tokens such as words or phrases in text (Shatkay & Feldman, 2003).

Figure 1. An overview of a biomedical literature mining system



In the area of named entity extraction, Proux et al. (2000) use single word names only with selected test set from 1,200 sentences coming from Flybase. Collier, et al. (2000) adopt Hidden Markov Models (HMMs) for 10 test classes with small training and test sets. Krauthammer et al. (2000) use BLAST database with letters encoded as 4-tuples of DNA. Demetriou and Gaizuaskas (2002) pipeline the mining processes, including hand-crafted components and machine learning components. For the study, they use large lexicon and morphology components. Narayanaswamy et al. (2003) use a part of speech (POS) tagger for tagging the parsed MEDLINE abstracts. Although Narayanaswamy and his colleagues (2003) implement an automatic protein name detection system, the number of words used is 302, and, thus, it is difficult to see the quality of their system, since the size of the test data is too small. Yamamoto, et al. (2003) use morphological analysis techniques for preprocessing protein name tagging and apply support vector machine (SVM) for extracting protein names. They found that increasing training data from 390 abstracts to 1,600 abstracts improved F-value performance from 70% to 75%. Lee et al. (2003) combined an SVM and dictionary lookup for named entity recognition. Their approach is based on two phases: the first phase is identification of each entity with an SVM classifier, and the second phase is post-processing to correct the errors by the SVM with a simple dictionary lookup. Bunescu, et al. (2004) studied protein name identification and protein-protein interaction. Among several approaches used in their study, the main two ways are one using POS tagging and the other using the generalized dictionary-based tagging. Their dictionary-based tagging presents higher F-value. Table 1 summarizes the works in the areas of named entity extraction in biomedical literature.

The second target object type of biomedical literature extraction is relation extraction. Leek (1997) applies HMM techniques to identify gene names and chromosomes through heuristics. Blaschke et al. (1999) extract protein-protein interactions based on co-occurrence of the form "… p1…I1… p2" within a sentence, where p1, p2 are proteins, and I1 is an interaction term. Protein names and interaction terms (e.g., activate, bind, inhibit) are provided as a dictionary. Proux (2000) extracts an interact relation for the gene

Table 1. A summary of works in biomedical entity extraction

Author	Named Entities	Database	No of Words	Learning Methods	F value
Collier et al. (2000)	Proteins and DNA	MEDLINE	30000	HMM	73
Krauthammer et al. (2000)	Gene and Protein	Review articles	5000	Character sequence mapping	75
Demetriou and Gaizauskas (2002)	Protein, Species, and 10 more	MEDLINE	30,000	PASTA template filing	83
Narayanaswamy (2003)	Protein	MEDLINE	302	Hand –crafted rules and co-occurrence	75.86
Yamamoto et al. (2003)	Protein	GENIA	1600 abstracts	BaseNP recognition	75
Lee et al. (2003)	Protein DNA RNA	GENIA	10,000	SVM	77
Bunescu (2004)	Protein	MEDLINE	5,206	RAPIER, BWI, TBL, k-NN , SVMs, MaxEnt	57.86

entity from Flybase database. Pustejovsky (2002) extracts an inhibit relation for the gene entity from MEDLINE. Jenssen, et al. (2001) extract a gene-gene relations based on co-occurrence of the form "… g1…g2…" within a MEDLINE abstracts, where g1 and g2 are gene names. Gene names are provided as a dictionary, harvested from HUGO, LocusLink, and other sources. Although their study uses 13,712 named human genes and millions of MEDLINE abstracts, no extensive quantitative results are reported and analyzed. Friedman, et al. (2001) extract a pathway relation for various biological entities from a variety of articles. In their work, the precision of the experiments is high (from 79-96%). However, the recalls are relatively low (from 21-72%). Bunescu et al. (2004) conducted protein/protein interaction identification with several learning methods, such as pattern matching rule induction (RAPIER), boosted wrapper induction (BWI), and extraction using longest common subsequences (ELCS). ELCS automatically learns rules for extracting protein interactions using a bottom-up approach. They conducted experiments in two ways: one with manually crafted protein names and the other with the extracted protein names by their name identification method. In both experiments, Bunescu, et al. compared their results with human-written rules and showed that machine learning methods provide higher precisions than human-written rules. Table 2 summarizes the works in the areas of relation extraction in biomedical literature.

Techniques Used

The most commonly used extraction technique is co-occurrence based. The basic idea of this technique is that entities are extracted based on frequency of co-occurrence of biomedical named entities such as proteins or genes within sentences. This technique was introduced by Blaschke, et al. (1999). Their goal was to extract information from scientific text about protein interactions among a predetermined set of related programs. Since Blaschke and his colleagues' study, numerous other co-occurrence-based systems have been proposed in the literature. All are associated with information extraction of biomedical entities from the unstructured text corpus. The common denominator of the co-occurrence-based systems is that they are based on co-occurrences of names or identifiers of entities, typically along with activation/dependency terms. These systems

Table 2. A summary of relation extraction for biomedical data

Authors	Relation	Entity	DB	Learning Methods	Precision	Recall
Leek (1997)	Location	Gene	OMIM	HMM	80%	36%
Blaschke (1999)	Interact	Protein	MEDLINE	Co-occurrence	n/a	n/a
Proux (2000)	Interact	Gene	Flybase	Co-occurrence	81%	44%
Puste-jovsky (2001)	Inhibit	Gene	MEDLINE	Co-occurrence	90%	57%
Jenssen (2001)	Location	Gene	MEDLINE	Co-occurrence	n/a	n/a
Friedman (2001)	Pathway	Many	Articles	Co-occurrence & thesauri	96%	63%
Bunescu (2004)	Interact	Protein	MEDLINE	RAPIER, BWI, ELCS	n/a	n/a

are differentiated one from another by integrating different machine learning techniques such as syntactical analysis or POS tagging, as well as ontologies and controlled vocabularies (Hahn et al., 2002; Pustejovsky et al., 2002; Yakushiji et al., 2001). Although these techniques are straightforward and easy to develop, from the performance standpoint, recall and precision are much lower than any other machine-learning techniques (Ray & Craven, 2001).

In parallel with co-occurrence-based systems, the researchers began to investigate other machine learning or NLP techniques. One of the earliest studies was done by Leek (1997), who utilized Hidden Markov Models (HMMs) to extract sentences discussing gene location of chromosomes. HMMs are applied to represent sentence structures for natural language processing, where states of an HMM correspond to candidate POS tags, and probabilistic transitions among states represent possible parses of the sentence, according to the matches of the terms occurring in it to the POSs. In the context of biomedical literature mining, HMM is also used to model families of biological sequences as a set of different utterances of the same word generated by an HMM technique (Baldi et al., 1994).

Ray and Craven (2001) have proposed a more sophisticated HMM-based technique to distinguish fact-bearing sentences from uninteresting sentences. The target biological entities and relations that they intend to extract are protein subcellular localizations and gene-disorder associations. With a predefined lexicon of locations and proteins and several hundreds of training sentences derived from Yeast database, they trained and tested the classifiers over a manually labeled corpus of about 3,000 MEDLINE abstracts. There have been several studies applying natural language tagging and parsing techniques to biomedical literature mining. Friedman, et al. (2001) propose methods parsing sentences and using thesauri to extract facts about genes and proteins from biomedical documents. They extract interactions among genes and proteins as part of regulatory pathways.

Evaluation

One of the pivotal issues yet to be explored further in biomedical literature mining is how to evaluate the techniques or systems. The focus of the evaluation conducted in the literature is on extraction accuracy. The accuracy measures used in IE are precision and recall ratio. For a set of N items, where N is either terms, sentences, or documents, and the system needs to label each of the terms as positive or negative, according to some criterion (positive, if a term belongs to a predefined document category or a term class). As discussed earlier, the extraction accuracy is measured by precision and recall ratio. Although these evaluation techniques are straightforward and are well accepted, recall ratios often are criticized in the field of information retrieval, when the total number of true positive terms is not clearly defined.

In IE, an evaluation forum similar to TREC in information retrieval (IR) is the Message Understanding Conference (MUC). Participants in MUC tested the ability of their systems to identify entities in text to resolve co-reference, extract and populate attributes of entities, and perform various other extraction tasks from written text. As identified by Shatkay and Feldman (2003), the important challenge in biomedical literature mining is the creation of gold-standards and critical evaluation methods for systems developed in this very active field. The framework of evaluating biomedical literature mining systems was recently proposed by Hirschman, et al. (2002). According to Hirschman, et al. (2002), the following elements are needed for a successful evaluation: (1) challenging problem; (2) task definition; (3) training data; (4) test data; (5) evaluation methodology and implementation; (6) evaluator; (7) participants; and (8) funding. In addition to these elements for evaluation, the existing biomedical literature mining systems encounter the issues of portability and scalability, and these issues need to be taken into consideration of the framework for evaluation.

Data Sources

In terms of data sources from which target biomedical objects are extracted, most of the biomedical data mining systems focus on mining MEDLINE abstracts of National Library of Medicine. The principal reason for relying on MEDLINE is related to complexity. Abstracts occasionally are easier to mine, since many papers contain less precise and less well supported sections in the text that are difficult to distinguish from more informative sections by machines (Andrade & Bork, 2000). The current version of MEDLINE contains nearly 12 million abstracts stored on approximately 43GB of disk space. A prominent example of methods that target entire papers is still restricted to a small number of journals (Friedman et al., 2000; Krauthammer et al., 2002). The task of unraveling information about function from MEDLINE abstracts can be approached from two different viewpoints. One approach is based on computational techniques for understanding texts written in natural language with lexical, syntactical, and semantic analysis. In addition to indexing terms in documents, natural language processing (NLP) methods extract and index higher-level semantic structures composed of terms and relationships between terms. However, this approach is confronted with the variability, fuzziness, and complexity of human language (Andrade & Bork, 2000). The Genies system (Friedman et al., 2000; Krauthammer et al., 2002), for automatically gathering and processing of knowledge about molecular pathways, and the Information Finding from Biological Papers (IFBP) transcription factor database are natural language processing based systems.

An alternative approach that may be more relevant in practice is based on the treatment of text with statistical methods. In this approach, the possible relevance of words in a text is deduced from the comparison of the frequency of different words in this text with the frequency of the same words in reference sets of text. Some of the major methods using the statistical approach are AbXtract and the automatic pathway discovery tool of Ng and Wong (1999). There are advantages to each of these approaches (i.e., grammar or pattern matching). Generally, the less syntax that is used, the more domain-specific the system is. This allows the construction of a robust system relatively quickly, but many subtleties may be lost in the interpretation of sentences. Recently, GENIA corpus has been used for extracting biomedical-named entities (Collier et al., 2000; Yamamoto et al., 2003). The reason for the recent surge of using GENIA corpus is because GENIA provides annotated corpus that can be used for all areas of NLP and IE applied to the biomedical domain that employs supervised learning. With the explosion of results in molecular biology, there is an increased need for IE to extract knowledge to build databases and to search intelligently for information in online journal collections.

FUTURE TRENDS

With the taxonomy proposed here, we now identify the research trends of applying IE to mine biomedical literature.

1. A variety of biomedical objects and relations are to be extracted.
2. Rigorous studies are conducted to apply advanced IE techniques, such as Random Common Field and Max Entropy based HMM to biomedical data.
3. Collaborative efforts to standardize the evaluation methods and the procedures for biomedical literature mining.
4. Continue to broaden the coverage of curated databases and extend the size of the biomedical databases.

CONCLUSION

The sheer size of biomedical literature triggers an intensive pursuit for effective information extraction tools. To cope with such demand, the biomedical literature mining emerges as an interdisciplinary field that information extraction and machine learning are applied to the biomedical text corpus.

In this article, we approached the biomedical literature mining from an IE perspective. We attempted to synthesize the research efforts made in this emerging field. In doing so, we showed how current information extraction can be used successfully to extract and organize information from the literature. We surveyed the prominent methods used for information extraction and demonstrated their applications in the context of biomedical literature mining

The following four aspects were used in classifying the current works done in the field: (1) what to extract; (2) what techniques are used; (3) how to evaluate; and (4) what data sources are used. The taxonomy proposed in this article should help identify the recent trends and issues pertinent to the biomedical literature mining.

REFERENCES

Andrade, M.A., & Bork, P. (2000). Automated extraction of information in molecular biology. *FEBS Letters*, *476*,12-7.

Blaschke, C., Andrade, M.A., Ouzounis, C., & Valencia, A. (1999). Automatic extraction of biological information from scientific text: Protein-protein interactions, *Proceedings of the First International Conference on Intelligent Systems for Molecular Biology.*

Bunescu, R. et al. (2004). Comparative experiments on learning information extractors for proteins and their interactions [to be published]. *Journal Artificial Intelligence in Medicine on Summarization and Information Extraction from Medical Documents.*

Collier, N., Nobata,C., & Tsujii, J. (2000). Extracting the names of genes and gene products with a hidden Markov model. *Proceedings of the 18th International Conference on Computational Linguistics (COLING2000).*

De Bruijn, B., & Martin, J. (2002). Getting to the (c)ore of knowledge: Mining biomedical literature. *International Journal of Medical Informatics*, *67*, 7-18.

Demetriou, G., & Gaizauskas, R. (2002). Utilizing text mining results: The pasta Web system. *Proceedings of the Workshop on Natural Language Processing in the Biomedical Domain*.

Friedman, C., Kra, P., Yu, H., Krauthammer, M., & Rzhetsky, A. (2001). GENIES: A natural-language processing system for the extraction of molecular pathways from journal articles. *Bioinformatics*, *17*, S74-82.

Hahn, U., Romacker, M., & Schulz, S. (2002). Creating knowledge repositories from biomedical reports: The MEDSYNDIKATE text mining system. *Proceedings of the Pacific Symposium on Biocomputing.*

Hirschman, L., Park, J.C., Tsujii, J., Wong, L., & Wu, C.H. (2002). Accomplishments and challenges in literature data mining for biology. *Bioinformatics*, *18*(12), 1553-1561.

Jenssen, T.K., Laegreid, A., Komorowski, J., & Hovig, E. (2001). A literature network of human genes for high-throughput analysis of gene expression. *Nature Genetics*, *28*(1), 21-8.

Krauthammer, M., Rzhetsky, A., Morozov P., & Friedman, C. (2000). Using BLAST for identifying gene and protein names in journal articles. *Gene*, *259*(1-2), 245-252.

Lee, K., Hwang, Y., & Rim, H. (2003). Two-phase biomedical NE recognition based on SVMs. *Proceedings of the ACL 2003 Workshop on Natural Language Processing in Biomedicine*.

Leek, T.R. (1997). *Information extraction using hidden Markov models* [master's theses]. San Diego, CA: Department of Computer Science, University of California.

Narayanaswamy, M., Ravikumar, K.E., & Vijay-Shanker, K. (2003). A biological named entity recognizer. *Proceedings of the Pacific Symposium on Biocomputing.*

Ng, S.K., & Wong, M. (1999). Toward routine automatic pathway discovery from on-line scientific text abstracts. *Proceedings of the Genome Informatics Series: Workshop on Genome Informatics.*

Proux, D., Rechenmann, F., & Julliard, L. (2000). A pragmatic information extraction strategy for gathering data on genetic interactions. *Proceedings of the International Conference on Intelligent System for Molecular Biology.*

Pustejovsky, J., Castano, J., Zhang, J., Kotecki, M., & Cochran, B. (2002). Robust relational parsing over biomedical literature: extracting inhibit relations. *Pacific Symposium on Biocomputing* (pp. 362-73).

Ray, S., & Craven, M. (2001). Representing sentence structure in hidden Markov models for information extraction. *Proceedings of the 17th International Joint Conference on Artificial Intelligence*, Seattle, Washington.

Shatkay, H., & Feldman, R. (2003). Mining the biomedical literature in the genomic era: An overview. *Journal of Computational Biology, 10*(6), 821-855.

Yakushiji, A., Tateisi, Y., Miyao,Y., & Tsujii, J. (2001). Event extraction from biomedical papers using a full parser. *Proceedings of the Pacific Symposium on Biocomputing.*

Yamamoto, K., Kudo, T., Konagaya, A., & Matsumoto, Y. (2003). Protein name tagging for biomedical annotation in text. *Proceedings of the ACL 2003 Workshop on Natural Language Processing in Biomedicine*.

KEY TERMS

F-Value: Combines recall and precision in a single efficiency measure (it is the harmonic mean of precision and recall): F = 2 * (recall * precision) / (recall + precision).

Hidden Markov Model (HMM): A statistical model where the system being modeled is assumed to be a Markov process with unknown parameters, and the challenge is to determine the hidden parameters from the observable parameters, based on this assumption.

Natural Language Processing (NLP): A subfield of artificial intelligence and linguistics. It studies the problems inherent in the processing and manipulation of natural language.

Part of Speech (POS): A classification of words according to how they are used in a sentence and the types of ideas they convey. Traditionally, the parts of speech are the noun, pronoun, verb, adjective, adverb, preposition, conjunction, and interjection.

Precision: The ratio of the number of correctly filled slots to the total number of slots the system filled.

Recall: Denotes the ratio of the number of slots the system found correctly to the number of slots in the answer key.

Support Vector Machine (SVM): A learning machine that can perform binary classification (pattern recognition) as well as multi-category classification and real valued function approximation (regression estimation) tasks.

This work was previously published in Encyclopedia of Data Warehousing and Mining, edited by J. Wang, pp. 615-620, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 1.26
Realizing Knowledge Assets in the Medical Sciences with Data Mining:
An Overview

Adam Fadlalla
Cleveland State University, USA

Nilmini Wickramasinghe
Cleveland State University, USA

ABSTRACT

This chapter provides insight into various areas within the medical field that strive to take advantage of different data mining techniques in order to realize the full potential of their knowledge assets. Specifically, this is done by discussing many of the limitations associated with conventional methods of diagnosis and showing how data mining can be used to improve these methods. Comparative analyses of different techniques associated with various areas within the medical field are outlined in order to identify the right technique for particular medical specialties. Furthermore, suggestions are provided to appropriately utilize the various data mining techniques thereby leading to effective and efficient knowledge management and knowledge utilization. In this chapter we highlight the potential of data mining in improving the exploratory as well as the predictive capabilities of conventional diagnostic methods in medical science.

INTRODUCTION

Knowledge management is an emerging business approach aimed at solving current business challenges to increase efficiency and effectiveness of core business processes while simultaneously fostering continuous creativity and innovation. Specifically, knowledge management through the use of various tools, processes and techniques combines germane organizational data, information and knowledge to create business value and enable an organization to capitalize on its intangible (e.g., knowledge) and intellectual assets so that it can effectively achieve its primary business goals as well as maximize its core business competencies (Swan et al., 1999; Davenport & Prusak, 1998). The need for knowledge management is based on a paradigm shift in the business environment where knowledge is central to organizational performance (Drucker, 1993).

Knowledge management offers organizations many tools, techniques and strategies to apply to their existing business processes. In essence then, knowledge management not only involves the production of information but also the capture of data at the source, the transmission and analysis of this data as well as the communication of information based on or derived from the data to those who can act on it (Swan et al., 1999). Fundamental to knowledge management is effectively integrating people, processes and technologies.

A pivotal technique in knowledge management is data mining which is used to discover new knowledge from existing data and information and thus grow the extant knowledge asset of the organization. This is particularly relevant to health care because not only is health care a knowledge-based industry, but it is also currently experiencing exponential growth in the collection of data and information primarily due to new legislative initiatives such as Managed Care and HIPAA (Health Information Portability and Accountability Act) in the US. This then makes it imperative for medical science to incorporate the benefits of this technique. We address this imperative by first discussing basic concepts of data mining and how they relate to the medical sciences. Next we elaborate upon key data mining techniques as well as their advantages and disadvantages and how they contribute to the building of important knowledge assets within health care.

BACKGROUND TO DATA MINING

In the literature, data mining is generally described at two levels: a broad perspective and a narrow perspective. While the broader perspective equates data mining to the process of Knowledge Discovery in Databases (KDD), the narrow perspective sees data mining as a step within this KDD process. In either case data mining can be defined as, "*The nontrivial extraction of implicit, previously unknown, and potentially useful information from data*" (Frawley et al., 1992). Data mining uses machine learning, as well as statistical and visualization techniques to discover and present knowledge in a form that is easily comprehensible to humans. Data mining involves sifting through huge amounts of data and extracting the relevant pieces of data for the particular analysis of a problem. More than just conventional data analysis (such as basic statistical methods), the technique makes heavy use of artificial intelligence. Often the emphasis is not as much on the extracting of data but more on the generating of a hypothesis, as in the case of exploratory data mining. Data mining also uses sophisticated statistical analysis and modeling techniques, which allow users to find useful information such as trends and patterns hidden in their business data. Data mining is one of the latest technologies to assist users deal with the abundance of data that they have collected over time. For example, this technique will help optimize business decisions, increase the value of each customer, enhance communication, and improve customer satisfaction. The retail industry has been using data mining technology to understand customer buying patterns, product warranty management, detection of fraud, and identification of good credit risks. Data Mining has become more popular over time due to the following reasons:

1. The main reason for the popularity of various data mining techniques is due to the large amount of data already collected and newly appearing data that requires processing beyond traditional approaches. The amount of data collected by various businesses, scientific, medical and governmental organizations around the world is enormous. It is impossible for human analysts to cope with the ever-growing and overwhelming amounts of data.
2. When a person analyses the data, he/she is liable to make errors due to the inadequacy of the human brain (i.e., the bounded ratio-

nality problem) to solve complex multifactor dependencies in the data and sometimes a lack of objectiveness in such an analysis. A human always tries to derive results based upon previous experiments and experiences gained from investigating other systems, unlike data mining which simply reflects what the data is conveying without preconceived hypotheses.

3. One more advantage of data mining is that, particularly in the case of large amounts of data, this process involves a much lower cost than hiring a team of experts. Although this technique does not discard the human involvement, it significantly simplifies the job and allows an analyst who is not proficient in statistics or programming to manage the process of extracting knowledge from data (Mega Computer Intelligence).

DATA MINING IN MEDICAL SCIENCES

The medical sciences offer a unique opportunity to apply the many techniques of data mining. This is because health care generates mountains of administrative data about patients, hospitals, utilization, claims, etc. In addition, clinical trials, electronic patient records and computer supported disease management increasingly produce large amounts of clinical data. This data, both the administrative and clinical, is a strategic resource for health care institutions since it represents a raw form of their knowledge

Data mining discovers the patterns and correlations hidden within this raw knowledge, i.e., the data repository. Furthermore, it enables health care professionals to use these patterns to aid in decision making and the establishment of revised and improved treatment protocols, and thereby enhance organizational performance.

Previous studies (Maria-Luiza et al., 2001) in various areas of the medical sciences have revealed that conventional methods of detecting symptoms or other health-related problems have been very costly and error prone. Due to the complexities and inconsistencies in these detection methods, the diagnoses which are based on the information gained from these methods can lead to outcomes that are sometimes dangerous and even could lead to a person's death. For example, during the prognosis of breast cancer, the main detection method available is mammography. Due to the high volume and variation in the stage of potential malignancy of tumors from mammograms that need to be read by physicians, the accuracy rate tends to decrease, and methods that focus on automatic reading of digital mammograms become highly desirable. It has been proven that double reading (by two different experts) of mammograms increases the accuracy but also naturally increases the costs. Thus, making it even more imperative to incorporate computer-aided diagnosis systems to assist medical professionals in achieving cost efficiency and diagnostic effectiveness and thereby enabling more appropriate and timely treatment.

In more litigious environments, the increasing risk to health care organizations and providers due to error in detection and interpretations has become extremely costly. Therefore, it is becoming a necessity to adopt new methods to facilitate not only more accurate detection and then treatment but also better preventative measures. Health care organizations have already accumulated large raw knowledge assets in the form of administrative and clinical data. What is now important for them to do is to maximize the potential of this strategic asset, hence the need for embracing data mining.

DATA MINING TECHNIQUES AND THEIR ROLE IN HEALTHCARE

Data mining techniques are not only used in the detection of diseases but they also are beneficial in helping to compare the different procedures required for a prognosis. For example, a physician who has newly started in practice can learn from the association of different procedures to certain diagnoses, which is the result of exploratory data mining and thus can take advantage of these findings to more effectively treat their patients, rather than depending more on the prolonged "trial and error" diagnostic path which is both more time consuming and a lower quality of care approach. The following data mining techniques are recognized for being of great benefit to many areas in business, engineering, as well as other industries. Health care should not be an exception in the application of these techniques:

- Association Rules
- Clustering
- Neural Networks
- Decision Trees

While we acknowledge that there are numerous data mining techniques, we focus on these techniques since they are some of the major techniques that are most suitable in our opinion to the medical sciences. The first two techniques are used for exploratory data mining, the latter two techniques are used for predictive data mining. We will first outline the major steps involved in data mining in order to achieve the final goal of knowledge creation before we describe each of the above data mining techniques.

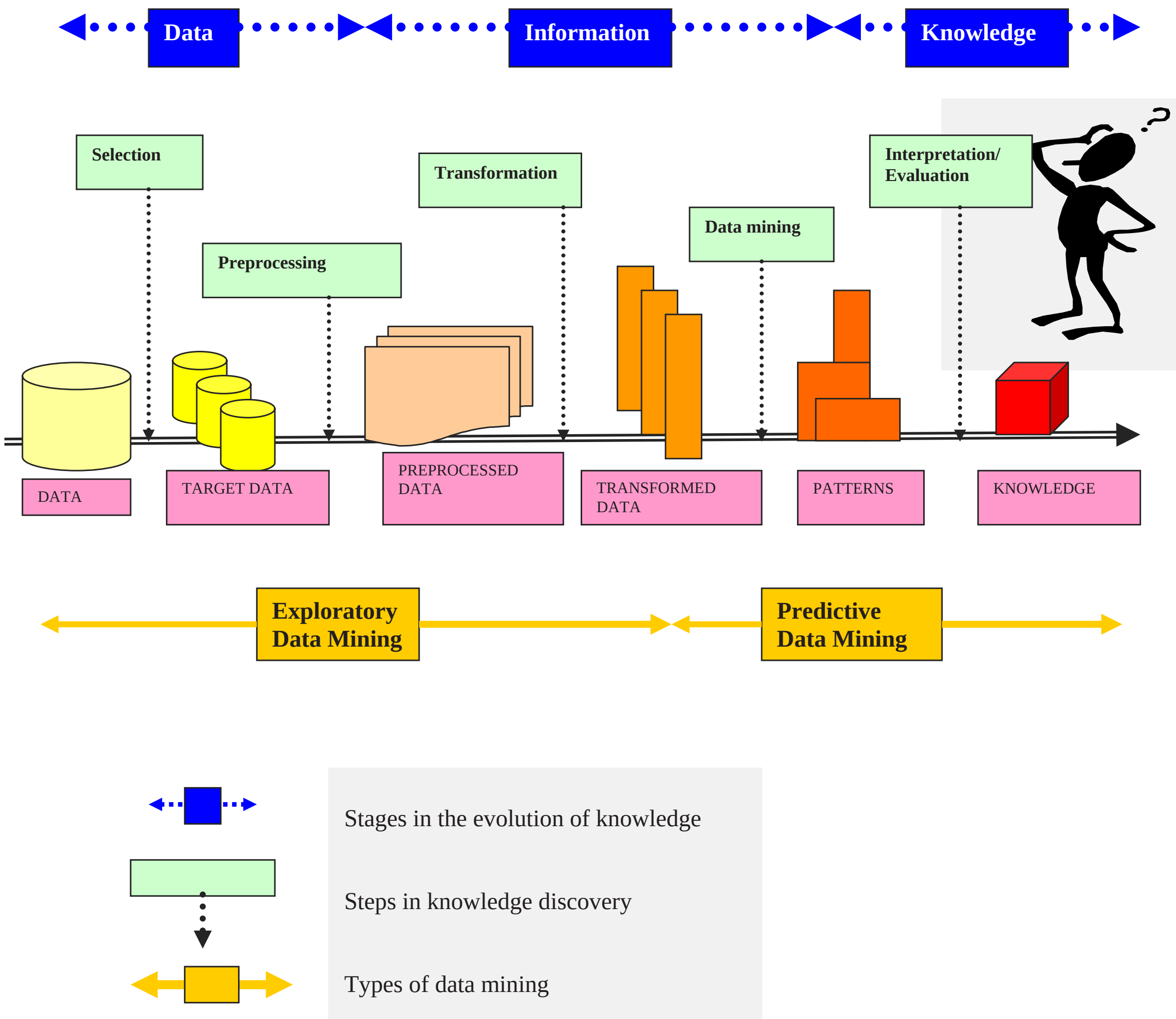
Knowledge Discovery Process

Figure 1 shows the knowledge discovery process, the evolution of knowledge from data through information to knowledge (Fayyad et al., 1996) and the types of data mining (exploratory and predictive) and their interrelationships. It is essential to emphasize here the importance of the interaction with the medical professionals and administrators who should always play a crucial and indispensable role in a knowledge discovery process, as depicted in *Figure 1* in the interpretation step. This is particularly true when we take into consideration features that are specific to the medical databases. For example, more and more medical procedures employ imaging as a preferred diagnosing tool. Thus, there is a need to develop methods for efficient mining in databases of images, which is inherently more difficult than mining in numerical databases. Other significant features include but are not limited to security and confidentiality concerns and the fact that the physician's interpretation of images, signals, or other clinical data, is written in unstructured English, which is also very difficult to mine (McGee, 1997). Some important data issues that data mining is most useful in helping organizations wrestle with include: huge volumes of data, dynamic data, incomplete data, imprecise data, noisy data, missing attribute values, redundant data, and inconsistent data.

Figure 1 also shows how data goes through the following process steps before being used for any decision-making:

- **Selection:** selecting the data according to some criteria, e.g., all those people who are suffering from or at risk of cardiac complications.
- **Preprocessing:** this is the data cleansing stage where certain unwanted information which may not be relevant or useful to the analysis is removed.
- **Transformation:** the data is not merely transferred but also changed using various mathematical manipulations (such as logarithmic transformations).
- **Data mining:** this stage is concerned with the extraction of patterns from the data. It includes choosing a data-mining algorithm, which is appropriate to discover a particular pattern in the data.
- **Interpretation and evaluation:** this is where human interaction and intervention is essential, specifically the patterns identified

Figure 1. Overview of the knowledge discovery process



by the system are interpreted into knowledge by humans and thereby redundant or irrelevant patterns are removed while patterns deemed useful are translated into potential treatment decisions.

Association Rule Mining

Association rules are used to discover relationships between attribute sets for a given input pattern. Such relationships do not necessarily imply causation, they are only associations. For example, an association rule that can be derived from medical data could be that 80% of the cases that display a given symptom are diagnosed with a similar condition and hence improves diagnostic capabilities. These patterns (associations) are not easily discovered using other data mining techniques. The support of an association rule is the percentage of cases which include the antecedent of the rule, while the confidence of the association rule is the percentage of cases where both the antecedent and the consequence of the rule are displayed. Only rules whose support and confidence exceed predetermined thresholds are considered useful. The classic algorithm used to generate these rules is the Apriori *algorithm* (Laura, 1990).

Advantages of Association Rule

- Association rules are readily understandable.
- Association rules are best suited for categorical data analysis
- It is widely used in hospitals to maintain patient's records.
- The outcomes are easy to interpret and explain and thus easy to use in the aiding of decision making.

Disadvantages of Association Rule Mining

- Generate too many rules and sometimes these are even trivial rules.
- The association rules are not expressions of cause and effect, rather they are descriptive relationships in particular databases, so there is no formal testing to increase the predictive power of these rules.
- Insight, analysis and explanation by health care professionals are usually required to identify the new and useful rules and thereby achieve the full benefits from such association rules.

Clustering

In clustering we are trying to develop groupings that are internally homogenous, mutually exclusive and collectively exhaustive. For example, in the study and treatment of chromosomal and DNA-related problems the clustering technique is important. This technique is an exploratory data mining technique. The outcome from the clustering process can then be used as input into a decision tree or neural network (Berkhin, 2002).

The most frequently used clustering method is k-means. This is a geometrical method, which uses the distance from the average location of all the members of a particular cluster to place a specific data point. The whole data field is divided into numbers and then these numbers are normalized. The value of each field is interpreted as the distance from the origin along corresponding axes. The initial clusters are randomly defined and computationally refined during the clustering process. The working of the clustering technique is dependent on two main criteria: (1) the members of a cluster should be most similar to each other, and (2) members of any two different clusters should be most dissimilar.

In most cases clusters are usually mutually exclusive but in some instances they may be overlapping, probabilistic or have hierarchical structures. In k-means a data point is assigned to the cluster which has the nearest centroid (i.e., the nearest mean). Clustering requires the data in numeric form since it works by assigning the cluster points accordingly. This process of assigning points to clusters continues until points stop changing positions (i.e., cluster hopping).

Advantages of Clustering

- The main strength of clustering is that it is an undirected knowledge discovery technique.
- The clustering can be used as a preparatory technique for other data mining techniques such as decision trees or neural networks.
- The outcome of clustering can be visually represented and hence easily understood.
- Creating clusters reduces the complexity of the problem by subdividing the problem space into more manageable partitions.
- The more separable the data points the more effective clustering is.

Disadvantages of Using Clustering

- Clustering represents a snap shot of the data at a certain point in time and thus may not be as useful in highly dynamic situations.
- Sometimes the clusters generated may not even have a practical meaning.
- Sometimes it is possible not to spot the cluster since you do not know what you are looking for.
- Clustering can be computationally expensive.

Neural Networks

The technique of neural networks is modeled after the human brain and normally consists of many input nodes, one or more hidden (middle) layer nodes and one or more output nodes. The input and output nodes relate to each other through the hidden layer. The input layer represents the raw information that is fed into the network. The hidden layer represents a computational layer that transforms the inputs coming from the input layer into inputs to the output layer. The behavior of the output layer depends on the activity of the hidden layer where the weights between the hidden and output layers are used as a reconciliation mechanism to help minimize the difference between the actual and desired outputs.

The outcome of a neural network is improved through the minimization of an error function, i.e., namely the difference between a desired output and an actual output value. The most widely used algorithm used to minimize this error function is known as backpropagation. Each input pattern is evaluated individually and if its value exceeds a predetermined threshold, then a pre-specified rule fires (i.e., is activated) whereby its outcome is fed forward to the next layer. The firing rule is an important concept in neural networks and accounts for their high flexibility since it determines how one calculates whether a subsequent neuron (node) should fire for any given input pattern.

The most important application of neural networks is pattern recognition. The network is trained to associate specific output patterns with input patterns. The power of neural networks comes into play in its predictive abilities, i.e., associating an input pattern that has not previously been classified with a specific output pattern. In such cases, the network will most likely give the output that corresponds to a pre-classified input pattern that is least different from the new input pattern.

Neural networks are mainly used in the medical sciences in recognizing disease types from various scans such as MRI or CT scans. The neural networks learn by example and therefore the more examples we feed into the neural network the more accurate its predictive capabilities become. Neural networks can process a large number of medical records, each of which includes information on symptoms, diagnoses, and treatments for a particular case. The use of neural network as a potential tool in medical science is exemplified by its use in the study of mammograms. In breast cancer detection the primary task is detection of a tumorous cell in the early stages. The best probability for a successful cure of this disease is in its early detection. Therefore, the power of neural networks lies in that they could be used to detect minute changes in tissue patterns (a key indicator of the existence of malignant cells) that are often difficult to detect with the human eye.

Advantages of Neural Networks

- Neural networks are good classification and prediction techniques when the results of the model are more important than the understanding of how the model works.
- Neural networks are very robust in that they can be used to model any type of relationship implied by the input patterns.
- Neural networks can easily be implemented to take advantage of the power of parallel computers with each processor simultaneously doing its own calculations.

- Neural networks are also very robust in situations where the data is noisy.

Disadvantages of Neural Networks

- The key problem with neural networks is the difficulty to explain its outcome. Unlike decision trees, neural networks use complex nonlinear modeling that does not produce rules and hence it is hard to justify one's decision.
- Significant preprocessing and preparation of the data is required.
- Neural networks will tend to over-fit the data unless implemented carefully. This is due to the fact that the neural networks have a large number of parameters which can fit any data set arbitrarily well.
- Neural networks require extensive training time unless the problem is small.

Decision Trees

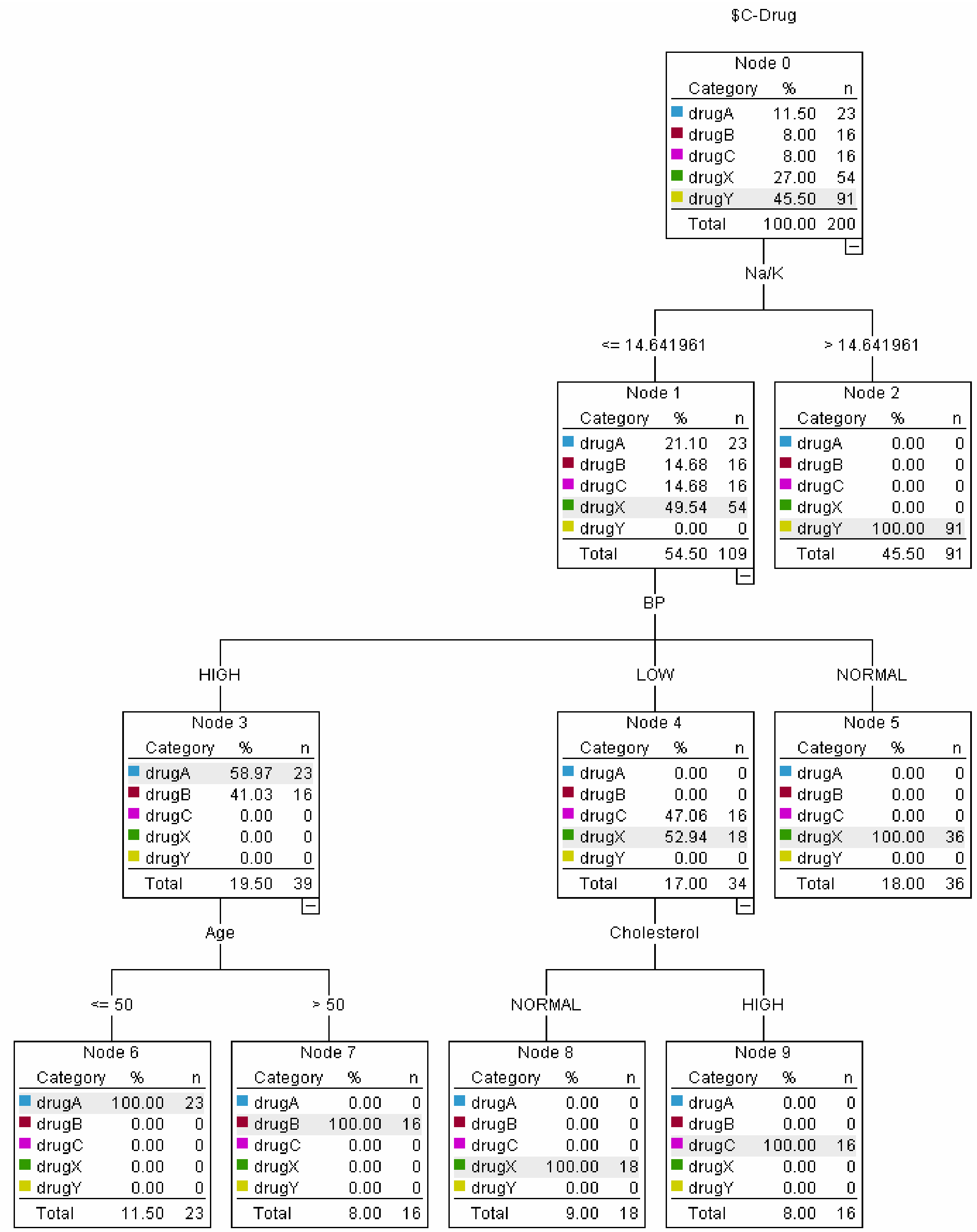
In critical decision situations, mistakes could be costly and have far reaching impacts. Thus data mining techniques are adopted in an attempt to minimize such mistakes. Decision trees split the available information in a treelike form and then arrive at a final decision by continuously refining the decision choices. The decision is usually made based on the choice between binary outcomes. For example, consider the binary decision of choosing between two methods—surgery and radiation—in the case of cancer treatments.

Decision making permeates health care but is of particular significance in the treatment of life-threatening diseases such as cancer. The decision tree then becomes a particularly powerful tool in such circumstances. Particularly in the case of cancer, early detection is critical since the disease grows rapidly and secondaries are more likely to develop in the meantime. A principal decision-making aspect is to decide quickly upon the specific treatment technique and then administer it and proceed with the delivery of care. For example, in *Figure 2* we can see a simple decision tree that tries to model the underlying decision problem of which drug to administer under which circumstances/conditions. At the root (the top node), the data is split into two partitions with respect to this decision problem, where one partition reflects cases where the Na/K ratio is less than or equal to 14.6 and it is not clear which drug should be administered, while the other partition represents the cases where the Na/K ratio is greater than 14.6 and it is clear that Drug Y is the drug of choice. Partition 1 therefore needs to be further subdivided into three sub-partitions; namely high (partition 1.1), low (partition 1.2) or normal (partition 1.3) blood pressure cases. In the case of partition 1.1 we can see that the choice is narrowed to Drug A or Drug B, so further sub-partitioning is required (namely, partition 1.1.1 and partition 1.1.2) and is performed on age in order to get clear decisions. It then becomes clear that age is a deciding factor between administering Drug A or Drug B—something that could not be seen from partition 1.1 or even less obvious from partition 1.

Advantages Of Decision Tree

- The graphical representation of a decision tree makes it a convenient and user-friendly modeling technique since it becomes very easy to visually follow the appropriate decision path and thereby facilitate accurate decision making.
- The decision tree algorithms are not only very fast and efficient to implement but the results are also unambiguous and thus easy to interpret. This feature of easy interpretability becomes of even greater significance in the medical sciences because often times doctors must justify their treatment decisions, such as in litigation instances.
- Decision trees can handle categorical (non-numeric) decision variables which are

Figure 2. Data mining resulting in the decision tree – each path from the tree root down represents a rule (i.e. a type of pattern)
Knowledge stage: Knowledge
Type of data mining: Predictive



common in the medical sciences such as in *Figure 2* where the decision variable is the drug to be administered.

- Decision trees can handle modeling situations where there is missing data, a situation that better mirrors practice.
- Decision trees prioritize the variables, using those with the most predictive power early in the partitioning process, hence using the most informative data first.

Disadvantages of Decision Trees

- When a decision variable is continuous, a categorization scheme needs to be developed first before applying the decision tree. Any weaknesses in this categorization scheme will be reflected in the outcome of the decision tree technique.
- There are dependencies between generations of splits (i.e., partition 1 impacts partitions 1.1, 1.2, etc.).
- The way in which decision trees handle numeric variables sometimes leads to loss of information due to loss of detail.

CONTRIBUTION TO KNOWLEDGE ASSETS

Irrespective of the specific data mining technique adopted, the significant common outcome of the application of all of these techniques is the generation of new knowledge. This newly created knowledge grows the extant knowledge base of the organization and thus not only adds value to its intangible assets but also increases its overall organizational value as new management techniques, such as the balanced scorecard, have demonstrated (Kaplan & Norton, 1996). In today's knowledge-based economy sustainable strategic advantages are gained more from an organization's knowledge assets than from its more traditional types of assets. Therefore, processes, tools and techniques that serve to grow the knowledge assets of an organization and thereby increase their value are strategic necessities to effectively compete in today's economy.

Healthcare is noted for using leading edge medical technologies and embracing new scientific discoveries to enable better cures for diseases and better means to enable early detection of most life threatening diseases. However, the healthcare industry globally, and in the US specifically, has been extremely slow to adopt key business processes (such as knowledge management) and techniques (such as data mining) (Wickramasinghe et al., 2003; Wickramasinghe & Mills, 2001). "Despite its information-intensive nature, the healthcare industry invests only 2% of gross revenues in information technology, compared with 10% for other information-intensive industries" (Bates et al., 2003). Furthermore, "[e]ven though US medical care is the world's most costly, its outcomes are mediocre compared with other industrial nations" (Bates et al., 2003). Therefore, making more of an investment in key business processes and techniques is a strategic imperative for the US healthcare industry if it is to achieve a premier standing with respect to high value, high quality and high accessibility of its healthcare delivery system.

In the final report compiled by the Committee on the Quality of Healthcare in America (Crossing the Quality Chasm, 2001), it was noted that improving patient care is integrally linked to providing high quality healthcare. Furthermore, in order to achieve a high quality of healthcare the committee identified six key aims — namely that healthcare should be: (1) safe: avoiding injuries to patients from the care that is intended to help them, (2) effective: providing services based on scientific knowledge to all who could benefit and refraining from providing services to those who will not benefit (i.e., avoiding under-use and overuse), (3) patient-centered: providing care that is respectful of and responsive to individual patient preferences, needs, and values and ensuring that patient values guide all clinical decisions, (4) timely: reducing waiting and sometimes harmful delays for both those receiving care and those who give care, (5) efficient: avoiding waste and (6) equitable: providing care that does not vary in quality based on personal characteristics.

Most of the poor quality connected with healthcare is related to a highly fragmented delivery system that lacks even rudimentary clinical information capabilities resulting in poorly designed care processes characterized by unnecessary duplication of services and long waiting times and delays (ibid). The development and application of sophisticated information systems is essential to address these quality issues and improve efficiency, yet healthcare delivery has been relatively untouched by the revolution of information technology, new business management processes such as knowledge management or new techniques such as data mining that are transforming so many areas of business today (Wickramasinghe et al., 2003; Wickramasinghe & Mills, 2001; Bates et al., 2003; Crossing the Quality Chasm, 2001; Wickramasinghe, 2000; Stegwee & Spil, 2001; Wickramasinghe & Silvers, 2002).

CONCLUSIONS

This chapter attempted to provide a survey of the four major data mining techniques and their application to the medical science field in order to realize the full potential of the knowledge assets in healthcare. We also presented an enhanced framework of the knowledge discovery process to highlight the interrelationships between data, information and knowledge as well as between knowledge creation and the key steps in data mining. There is no single data mining technique that will be best under all circumstances in healthcare as well as other industries. However, a comparative analysis of the various techniques as they are used in medical science suggests the following:

1. Neural networks are general and flexible thus can model situations with either numeric or non-numeric data. Further, they can handle noisy data effectively. However, their major limitation is that it is difficult to understand the reasoning behind their outcomes.
2. Decision trees on the other hand are very intuitive to understand but are not as flexible, nor as tolerant with noisy data.
3. Clustering provides a powerful exploratory data mining technique, however it can also be very computationally expensive. Further, it could sometimes generate clusters that are difficult to justify in practice. Clustering can be used as a first step for either decision trees or neural networks.
4. Association rules are very general in nature and their outcomes are very easy to understand, since these outcomes are made up of nested if-then rules. However, they require human insights to identify which rules are significant and useful and which are trivial.

Finally, we discussed the importance of data mining to the growing of the organizational knowledge assets and argued why this is of such significance in today's knowledge-based economy for healthcare. Clearly, as much as there is a need for such techniques in creating knowledge-based organizations in healthcare, there is still much to be done before knowledge management enabled through the adoption of these data mining techniques is diffused en mass throughout healthcare organizations and thereby enabling these organizations to realize the full benefits of their knowledge assets. This chapter then has served as an overview of how to realize the full potential of knowledge assets in the medical sciences using various key data mining techniques. There is naturally scope for future research to take a more detailed view of these techniques within the medical sciences.

REFERENCES

Antonie, M. L., Zaiane, O. R., & Coman, A. (2001, August). *Application of Data Mining Techniques for Medical Image Classification.* DataBase Laborarатory, Department of Computing Science, University of Alberta, Canada.

Bates, D., Ebell, M., Gotlieb, E., Zapp, J., & Mullins, H. (2003, Jan/Feb). A proposal for Electronic Records in U.S. *Primary Care JAMIA*, 10(1), 1-9.

Berkhin, P. (2002). *Survey of Clustering Data Mining Techniques.* Accrue Software, Inc.

Crossing the Quality Chasm - a New Health System for the 21st Century. (2001). Committee on Quality of Health Care in America Institute of Medicine. Washington, D.C.: National Academy Press.

Davenport, T. & Prusak, L. (1998). *Working Knowledge*. Boston, MA: Harvard Business School Press.

Drucker, P. (1993). *Post-Capitalist Society.* New York: Harper Collins.

Fayyad, Piatetsky-Shapiro, & Smyth (1996). From data mining to knowledge discovery: An overview. In Fayyad, Piatetsky-Shapiro, Smyth, & Uthurusamy, *Advances in Knowledge Discovery and Data Mining.* Menlo Park, CA: AAAI Press/The MIT Press.

Frawley, W., Piatetsky-Shapiro, G., & Matheus, C. (1992, Fall). Knowledge discovery in databases: An overview. *AI Magazine*, 213-228.

Kaplan, R. & Norton, D. (1996). *The Balanced Scorecard: Translating Strategy into Action*. Boston, MA: Harvard Business School Press.

Kolbasuk McGee, M. (1997, September). High-tech healing. *Information Week*.

Laura, Lanzarini, Ing. A. De Giusti, (1990). *Pattern Recognition in Medical Images using Neural Networks.*

Mega Computer Intelligence. Data Mining 101. Retrieved from the World Wide Web: http://www.megaputer.com/dm/dm101.php3.

Stegwee, R. & Spil, T. (2001). *Strategies for Healthcare Information Systems.* Hershey, PA: Idea Group Inc.

Swan, J., Scarbrough, H. & Preston, J. (1999). Knowledge management – The next fad to forget people? *Proceedings of the 7th European Conference in Information* Systems.

Wickramasinghe, N. (2000). IS/IT as a Tool to Achieve Goal Alignment: A theoretical Framework. *International J. Healthcare Technology Management*, 2(1/2/3/4), 163-180.

Wickramasinghe, N. & Mills, G. (2001). MARS: The electronic medical record system the core of the kaiser galaxy. *International Journal Healthcare Technology Management*, 3(5/6), 406-423.

Wickramasinghe, N. & Silvers, J. B. (2003). IS/IT the Prescription To Enable Medical Group Practices Attain Their Goals. Healthcare Management Science, 6, 75-86.

Wickramasinghe, N., Fadlalla, A., Geisler, E. & Schaffer, J. (2003). Knowledge Management and Data Mining Strategic Imperatives for Health Care. *Proceedings of the 3rd Hospital of the Future conference,* Warwick.

This work was previously published in Creating Knowledge-Based Healthcare Organizations, edited by N. Wickramasinghe, J.N.D. Gupta & S. Sharma, pp. 164-178 , copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 1.27
Data Mining Medical Digital Libraries

Colleen Cunningham
Drexel University, USA

Xiaohua Hu
Drexel University, USA

INTRODUCTION

Given the exponential growth rate of medical data and the accompanying biomedical literature, more than 10,000 documents per week (Leroy et al., 2003), it has become increasingly necessary to apply data mining techniques to medical digital libraries in order to assess a more complete view of genes, their biological functions and diseases. Data mining techniques, as applied to digital libraries, are also known as text mining.

BACKGROUND

Text mining is the process of analyzing unstructured text in order to discover information and knowledge that are typically difficult to retrieve. In general, text mining involves three broad areas: Information Retrieval (IR), Natural Language Processing (NLP) and Information Extraction (IE). Each of these areas are defined as follows:

- **Natural Language Processing:** a discipline that deals with various aspects of automatically processing written and spoken language.
- **Information Retrieval:** a discipline that deals with finding documents that meet a set of specific requirements.
- **Information Extraction:** a sub-field of NLP that addresses finding specific entities and facts in unstructured text.

MAIN THRUST

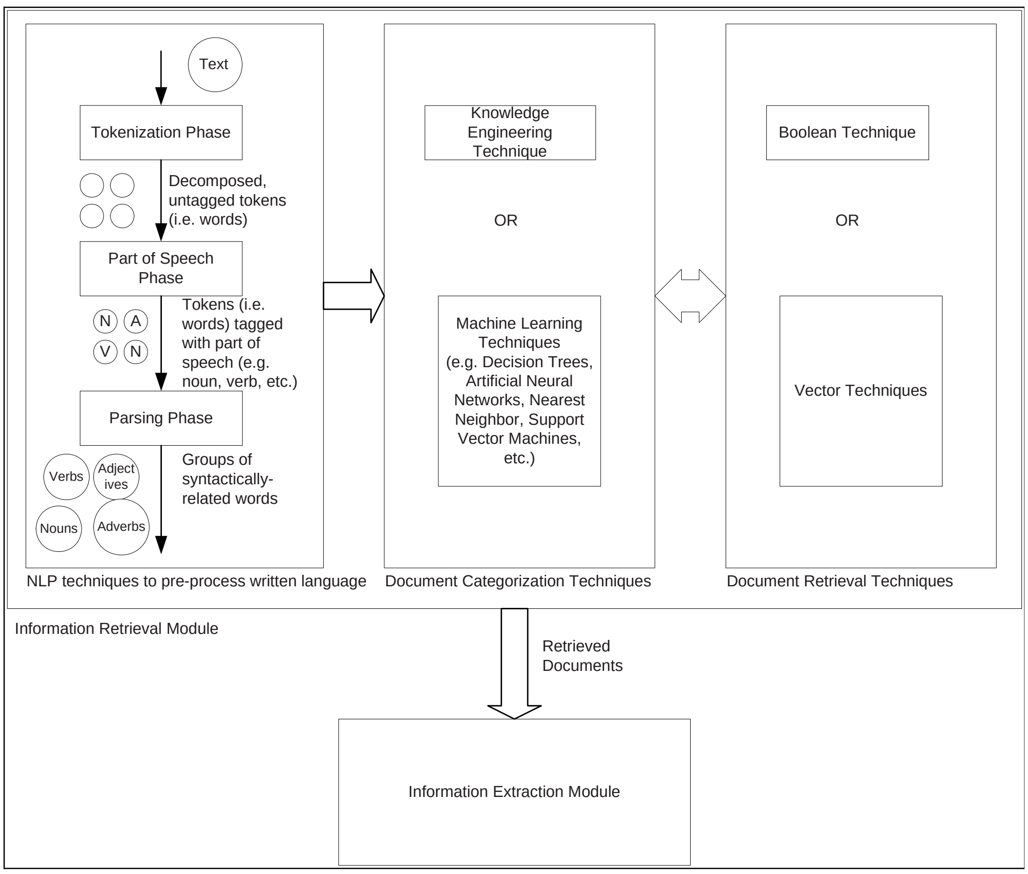
The current state of text mining in digital libraries is provided in order to facilitate continued research, which subsequently can be used to develop large-scale text mining systems. Specifically, an overview of the process, recent research efforts and practical uses of mining digital libraries, future trends and conclusions are presented.

Text Mining Process

Text mining can be viewed as a modular process that involves two modules: an information retrieval module and an information extraction module. presents the relationship between the modules and the relationships between the phases within the information retrieval module. The former module involves using NLP techniques to pre-process the written language and using techniques for document categorization in order to find relevant documents. The latter module involves finding specific and relevant facts within text. NLP consists of three distinct phases: (1) tokenization, (2) parts of speech (PoS) tagging and (3) parsing. In the tokenization step, the text is decomposed into its subparts, which are subsequently tagged during the second phase with the part of speech that each token represents (e.g., noun, verb, adjective, etc.). It should be noted that generating the rules for PoS tagging is a very manual and labor-intensive task. Typically, the parsing phase utilizes shallow parsing in order to group syntactically related words together because full parsing is both less efficient (i.e., very slow) and less accurate (Shatkay & Feldman, 2003). Once the documents have been pre-processed, then they can be categorized.

There are two approaches to document categorization: Knowledge Engineering (KE) and Machine Learning (ML). Knowledge Engineering requires the user to manually define rules, which can consequently be used to categorize documents into specific pre-defined categories. Clearly, one of the drawbacks of KE is the time that it would take a person (or group of people) to manually construct and maintain the rules. ML, on the other hand, uses a set of training documents to learn the rules for classifying documents.

Figure 1. Overview of text mining process



Specific ML techniques that have successfully been used to categorize text documents include, but are not limited to, Decision Trees, Artificial Neural Networks, Nearest Neighbor and Support Vector Machines (SVM) (Stapley et al., 2002). Once the documents have been categorized, then documents that satisfy specific search criteria can be retrieved.

There are several techniques for retrieving documents that satisfy specific search criteria. The Boolean approach returns documents that contain the terms (or phrases) contained in the search criteria; whereas, the vector approach returns documents based upon the term frequency-inverse document frequency (TF x IDF) for the term vectors that represent the documents. Variations of clustering and clustering ensemble algorithms (Iliopoulous et al., 2001; Hu, 2004), classification algorithms (Marcotte et al., 2001) and co-occurrence vectors (Stephens et al., 2001) have been successfully used to retrieve related documents. An important point to mention is that the terms that are used to represent the search criteria as well as the terms used to represent the documents are critical to successfully and accurately returning related documents. However, terms often have multiple meanings (i.e., polysemy) and multiple terms can have the same meaning (i.e., synonyms). This represents one of the current issues in text mining, which will be discussed in the next section.

The last part of the text mining process is information extraction, of which the most popular technique is co-occurrence (Blaschke & Valencia, 2002; Jenssen et al., 2001). There are two disadvantages to this approach, each of which creates opportunities for further research. First, this approach depends upon assumptions regarding sentence structure, entity names, and etcetera that do not always hold true (Pearson, 2001). Furthermore, this approach relies heavily on completeness of the list of gene names and synonyms and summarizes the modular process of text mining.

Research to Address Issues in Mining Digital Libraries

The issues in mining digital libraries, specifically medical digital libraries, include scalability, ambiguous English and biomedical terms, non-standard terms and structure and inconsistencies between medical repositories (Shatkay & Feldman, 2003). Most of the current text mining research focuses on automating information extraction (Shatkay & Feldman, 2003). The scalability of the text mining approaches is of concern because of the rapid rate of growth of the literature. As such, while most of the existing methods have been applied to relatively small sample sets, there has been an increase in the number of studies that have been focused on scaling techniques

Table 1. General text mining process

General Step	General Purpose	General Issues	General Solutions
IR	Identify and retrieve relevant docs	aliases, synonyms & homonyms	
IE	Find specific and relevant facts within text [e.g., find specific gene entities (entity extraction) & relationships between specific genes (relationship extraction)]	aliases, synonyms & homonyms	Controlled vocabularies (ontologies) [e.g., gene terms and synonyms: LocusLink (Pruitt & Maglott, 2001), SwissProt (Boeckmann et al., 2003), HUGO (HUGO, 2003), National Library of Medicine's MeSH (NLM, 2003)]

to apply to large collections (Pustejovsky et al., 2002; Jenssen et al., 2002). One exception to this is the study by Jenssen et al. (2001) in which the authors used a predefined list of genes to retrieve all related abstracts from PubMed that contained the genes on the predefined list.

Since mining digital libraries relies heavily on the ability to accurately identify terms, the issues of ambiguous terms, special jargon and the lack of naming conventions are not trivial. This is particularly true in the case of digital libraries where the issue is further compounded by non-standard terms. In fact, a lot of effort has been dedicated to building ontologies to be used in conjunction with text mining techniques (Boeckmann et al., 2003; HUGO, 2003; Liu et al., 2001; NLM, 2003; Oliver et al., 2002; Pruitt & Maglott, 2001; Pustejovsky et al., 2002). Manually building and maintaining ontologies, however, is a time consuming effort. In light of that, there have been several efforts to find ways of automatically extracting terms to incorporate into and build ontologies (Nenadic et al., 2002; Ono et al., 2001). The ontologies are subsequently used to match terms. For instance, Nenadic et al. (2002) developed the Tagged Information Management System (TIMS), which is an XML-based Knowledge Acquisition system that uses ontology for information extraction over large collections.

Uses of Text Mining in Medical Digital Libraries

There are many uses for mining medical digital libraries that range from generating hypotheses (Srinivasan, 2004) to discovering protein associations (Fu et al., 2003). For instance, Srinivasan (2004) developed MeSH-based text mining methods that generate hypotheses by identifying potentially interesting terms related to specific input. Further examples include, but are not limited to: uncovering uses for thalidomide (Weeber et al., 2003), discovering functional connections between genes (Chaussabel & Sher, 2002) and identifying viruses that could be used as biological weapons (Swanson et al., 2001). He summarizes some of the recent uses of text mining in medical digital libraries.

FUTURE TRENDS

The large volume of genomic data resulting and the accompanying literature from the Human Ge-

Table 2. Some uses of text mining medical digital libraries

Application	Technique	Supporting Literature
Building gene networks	Co-occurrence	Jenssen et al., 2001; Stapley & Benoit, 2000; Adamic et al., 2002
Discovering protein association	Unsupervised cluster learning and vector classification	Fu et al., 2003
Discovering gene interactions	Co-occurrence	Stephens et al., 2001; Chaussabel & Sher, 2002
Discovering uses for thalidomide	Mapping phrases to UMLS concepts	Weeber et al., 2001
Extracting and combining relations	Rule-based parser and co-occurrence	Leroy et al., 2003
Generating hypotheses	Incorporating ontologies (e.g., mapping terms to MeSH)	Srinivasan, 2004; Weeber et al., 2003
Identifying biological virus weapons		Swanson et al., 2001

nomic project is expected to continue to grow. As such, there will be a continued need for research to develop scalable and effective data mining techniques that can be used to analyze the growing wealth of biomedical data. Additionally, given the importance of gene names in the context of mining biomedical literature and the fact that there are a number of medical sources that use different naming conventions and structures, research to further develop ontology will play an important part in mining medical digital libraries. Finally, it is worth mentioning that there has been some effort to link the unstructured text documents within medical digital libraries with their related structured data in data repositories.

CONCLUSION

Given the practical applications of mining digital libraries and the continued growth of available data, mining digital libraries will continue to be an important area that will help researchers and practitioners gain invaluable and undiscovered insights into genes, their relationships, biological functions, diseases and possible therapeutic treatments.

REFERENCES

Blaschke, C., & Valencia, A. (2002). The frame-based module of the SUISEKI information extraction system. IEEE Intelligent Systems, Special Issue on Intelligent Systems in Biology, 17(2), 14-20.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., & Schneider, M. (2003). The SWISS-PROT Protein Knowledgebase and its Supplement TrEMBL in 2003. Nucleic Acids Research, 31(1), 365-370.

Chaussabel, D., & Sher, A. (2002). Mining microarray expression data by literature profiling. Genome Biology, 3(10), research0055.1-0055.16.

De Bruijn, B., & Martin, J. (2002). Getting to the core of knowledge: Mining biomedical literature. International Journal of Medical Informatics, 67, 7-18.

Fu, Y., Mostafa, J., & Seki, K. (2003). Protein association discovery in biomedical literature. In Proceedings of the Third ACM/IEEE-CS Joint Conference on Digital Libraries (pp.113-115).

Hu, X. (2004). Integration of cluster ensemble and text summarization for gene expression. In Proceedings of the 2004 IEEE Symposium of Bioinformatics and Bioengineering.

HUGO. (2003). HUGO (The Human Genome Organization) Gene Nomenclature Committee. Retrieved from http://www.gene.ucl.ac.uk/nomenclature

Iliopolous, I., Enright, A.J., & Ouzounis, C.A. (2001). Textquest: Document clustering of Medline abstracts for concept discovery in molecular biology. In Proceedings of the Pacific Symposium on Biocomputing (PSB) (pp. 384-395).

Jenssen, T.K., Laegrid, A., Komorowski, J., & Hovig, E. (2001). A literature network of human genes for high-throughput analysis of gene expression. Nature Genetics, 28(1), 21-28.

Leroy, G., Chen, H., Martinez, J.D., Eggers, S., Flasey, R.R., Kislin, K.L., Huang, Z., Li, J., Xu, J., McDonald, D.M., & Ng, G. (2003). Genescene: Biomedical text and data mining. In Proceedings of the Third ACM/IEEE-CS joint conference on Digital Libraries (pp. 116-118).

Liu, H. Lussier, Y.A., & Friedman, C. (2001). Diambiguating ambiguous biomedical terms in biomedical narrative text: an unsupervised method-abstract. Journal of Biomedical Informatics, 34(4), 249-261.

Marcotte, E.M, Xenarios, I., & Eisenberg, D. (2001). Mining literature for protein-protein interactions. Bioinformatics, 17(4), 359-363.

NLM. (2003). MeSH: Medical Subject Headings. Retrieved from http://www.nlm.nih.gov/mesh/

Oliver, D.E., Rubin, D.L., Stuart, J.M., Hewett, M., Klein, T.E., & Altman, R.B. (2002). Ontology development for a pharmacogenetics knowledge base. Proceedings of Pacific Symposium on Biocomputing (PSB)-2002, 7, 65-76.

Pearson, H. (2001). Biology's name game. Nature, 411(6838), 631-632.

Pruitt, K.D., & Maglott, D.R. (2001). RefSeq and LocusLink:NCBI gene-centered resources. Nucleic Acids Research, 29(1), 137-140.

Pustejovsky, J., Castano, J., Zhang, J., Kotecki, M., & Cochran, B. (2002). Robust relational parsing over biomedical literature: Extracting inhibit relations. Proceedings of Pacific Symposium on Biocomputing (PSB)-2002, 7, 362-373.

Shatkay, H., & Feldman, R. (2003). Mining the biomedical literature in the genomic era: An overview. Journal of Computational Biology, 10(6), 821-855.

Srinivasan, P. (2004). Text Mining: Generating hypotheses from MEDLINE. Journal of the American Society for Information Science and Technology, 55(5), 396-413.

Stapley, B.J., Kelley, L.A., & Sternberg, M.J. (2002). Predicting the sub-cellular location of proteins from text using support vector machines. Proceedings of the Pacific Symposium on Biocomputing (PSB), 7, 374-385.

Stephens, M., Palakal, M., Mukhopadhyay, S., Raje, R., & Mostafa, J. (2001). Detecting gene relations from Medline abstracts. In Proceedings of the Pacific Symposium on Biocomputing (PSB) (pp. 483-496).

Swanson, D.R., Smalheiser, N.R., & Bookstein, A. (2001). Information discovery from complementary literatures: Categorizing viruses as potential weapons. Journal of the American Society for Information Science, 52(10), 797-812.

Weeber, M., Klein, H., Berg, L., & Vos, R. (2001). Using concepts in literature-based discovery: Simulating Swanson's Raynaud-Fish Oil and Migraine-Magnesium discoveries. Journal of the American Society for Information Science, 52(7), 548-557.

Weeber, M., Vos, R., Klein, H., de Jong-Van den Berg, L.T.W., Aronson, A., & Molema, G. (2003). Generating hypotheses by discovering implicit associations in the literature: A case report for new potential therapeutic uses for Thalidomide. Journal of the American Medical Informatics Association, 10(3), 252-259.

KEY TERMS

Bibliomining: Data mining applied to digital libraries to discover patterns in large collections.

Bioinformatics: Data mining applied to medical digital libraries.

Clustering: An algorithm that takes a dataset and groups the objects such that objects within the same cluster have a high similarity to each other, but are dissimilar to objects in other clusters.

Information Extraction: A sub-field of NLP that addresses finding specific entities and facts in unstructured text.

Information Retrieval: A discipline that deals with finding documents that meet a set of specific requirements.

Machine Learning: Artificial intelligence methods that use a dataset to allow the computer to learn models that fit the data.

Natural Language Processing: A discipline that deals with various aspects of automatically processing written and spoken language.

Supervised Learning: A machine learning technique that requires a set of training data, which consists of known inputs and a priori desired outputs (e.g., classification labels) that can subsequently be used for either prediction or classification tasks.

Unsupervised Learning: A machine learning technique, which is used to create a model based upon a dataset; however, unlike supervised learning, the desired output is not known a priori.

This work was previously published in Encyclopedia of Data Warehousing and Mining, edited by J. Wang, pp. 278-282, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Section II
Development and Design Methodologies

This section provides in-depth coverage of conceptual architectures, frameworks and methodologies related to the design and implementation of medical information systems and technologies. Throughout these contributions, research fundamentals in the discipline are presented and discussed. From broad examinations to specific discussions on particular frameworks and infrastructures, the research found within this section spans the discipline while also offering detailed, specific discussions. Basic designs, as well as abstract developments, are explained within these chapters, and frameworks for designing successful e-health applications, mobile healthcare systems, and clinical decision support systems are discussed.

Chapter 2.1
Building Better E-Health Through a Personal Health Informatics Pedagogy

E. Vance Wilson
University of Wisconsin–Milwaukee, USA

ABSTRACT

E-health use is increasing worldwide, but no current e-health paradigm fulfills the complete range of user needs for online health services. This dilemma clouds a number of issues surrounding e-health, as promoters of e-commerce, personal health records, and consumer health informatics paradigms attempt to create encompassing e-health within the constraints of each unique perspective. In the long term, the most important of these issues may be the need to develop an e-health pedagogy that offers conceptual grounding and course curricula in order to effectively represent all facets of e-health. To address that issue, this article introduces a personal health informatics (PHI) paradigm that incorporates the best features of preceding paradigms by integrating informatics, personal, and healthcare perspectives. Drawing from PHI, a pedagogical framework is proposed to guide instruction in the design and development of encompassing e-health.

INTRODUCTION

Use of e-health continues to expand worldwide. Harris Interactive reports that the number of Americans who have searched for health information online has increased to 117 million, and 85% of the these individuals searched within the month prior to being surveyed (Krane, 2005). Outside the U.S. and Europe, e-health use has grown more slowly (Holliday & Tam, 2004). However, further expansion seems likely as the World Health Organization and similar groups undertake efforts to increase availability of e-health in developing nations (Kwankam, 2004; WHO, 2005).

While some aspects of successful e-health are well-established, such as the need to provide encyclopedic health content, other aspects are less obvious. For example:

- Which services should be deployed online, and how should users interface with these services?
- If communication is offered, what is the best way to coordinate this to balance the needs of the public with those of healthcare

representatives (e.g., physicians and clinic staff)?
- How should personal health records (PHR) be incorporated into e-health, who owns the data in these records, and what (if any) data should PHR share with records of the healthcare provider, insurer(s), and payer(s), such as an employer or a government agency?

These are not idle questions to the health informatics and IT professionals who must design and deploy e-health applications. However, the area of e-health currently is underserved (or not served at all) in health informatics and IT curricula. This is a situation that should be remedied as soon as possible, given the large number of healthcare providers who currently are investing in e-health as an important part of organizational strategy (Lazarus, 2001; Martin, Yen & Tan, 2002).

In developing effective curricula for designers and developers of e-health, I propose that it will be helpful to view e-health from a user-centered perspective that can incorporate best practices of current e-health paradigms without being limited by their constraints. This article presents the foundational concepts that underlie this perspective and proposes a new pedagogy.

PARADIGMS OF E-HEALTH

E-health is a broad domain with numerous published definitions that primarily address the convergence of healthcare and Internet technology (Oh, Rizo, Enkin, & Jadad, 2005). A frequently cited definition by Eysenbach (2001) highlights e-health's interdisciplinary underpinnings.

E-health is an emerging field in the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the Internet and related technologies. In a broader sense, the term characterizes not only a technical development, but also a state-of-mind, a way of thinking, an attitude, and a commitment for networked, global thinking, to improve health care locally, regionally, and worldwide by using information and communication technology. (Eysenbach, 2001, p. e20)

Historically, each of the paradigms discussed in the following sections has played an important role in developing e-health to its current state. Yet, I will argue that none of these paradigms fulfills the complete range of user needs for online health services, and for this reason, none presents a suitable basis for pedagogy in design and development of encompassing e-health.

E-Commerce Paradigm

Prominent early developers of e-health services operated within an e-commerce paradigm in which vendors expected to profit from users paying directly for products and services acquired through the site or from advertisers paying for exposure to users. Typically, vendors were not affiliated with healthcare providers, so they could not provide services that link individuals with their own physician, clinic, or pharmacy. Although numerous vendors developed e-health within the e-commerce paradigm, few survived the ensuing shakeout due to failure to provide value to customers or to adequately control costs, lack of effective revenue models, or simple inability to ensure sustainable competitive advantage (Itagaki, Berlin, & Schatz, 2002; Rovenpor, 2003). Among the prominent e-health ventures representing the e-commerce paradigm, such as DrKoop.com, MediConsult.com, and PlanetRx.com, only a handful remains. The best-known of these is WebMD, which provides an exceptionally wide range of health services but continues to struggle toward profitability.

Personal Health Record Paradigm

A second approach to e-health that initially was based on profit motives is the personal health record (PHR) paradigm. The PHR is defined as follows:

An electronic, universally available, lifelong resource of health information needed by individuals to make health decisions. Individuals own and manage the information in the PHR, which comes from healthcare providers and the individual. The PHR is maintained in a secure and private environment, with the individual determining rights of access. The PHR is separate from and does not replace the legal record of any provider. (AHIMA, 2005, p. 64A)

The idea of computerizing personal health information is not new. A number of vendors introduced PHR products during the 1990s, both as stand-alone software and as online services. To date, demand for commercial PHR products has been low—only 2% of individuals who maintain health records use purchased PHR software (Taylor, 2004)—and none of these products has gained financial success (Holt, 2005). However, several factors may prompt increased use of PHRs in the near term. First, the U.S. public gained increased awareness of the value of portable health records as a result of the massive human displacement and damage to healthcare facilities sustained from hurricanes during 2005. This awareness is likely to affect public sentiment and governmental policies for some time to come (Kloss, 2005). Second, PHR software is continuing to improve. Current offerings have become easier to use and are more fully capable of integrating health resources beyond the individual's immediate data, such as linking to relevant medical journals and FDA drug information (Campbell, 2005). Finally, researchers are beginning to study how individuals go about maintaining health records in non-computerized settings, and this research is clarifying the opportunities and constraints that PHR developers face. For example, two recent studies in this area assess the importance of privacy and security in health records (Taylor, 2004) and shed light on the diverse strategies that individuals prefer to employ in storing health records (Moen & Brennan, 2005).

Consumer Health Informatics Paradigm

Although healthcare providers did not participate strongly in early stages of e-health development (Lazarus, 2001), these organizations have been instrumental in developing an increasing amount of e-health within a consumer health informatics (CHI) paradigm. CHI is "a branch of medical informatics that analyzes consumers' needs for information; studies and implements methods for making information accessible to consumers; and models and integrates consumers' preferences into medical information systems" (Eysenbach, 2000, p. 1713). Where e-commerce vendors hoped to achieve competitive advantage, healthcare provider organizations appear to be driven more by concerns of achieving competitive equity, both with competing providers and commercial e-health sites. Because CHI is produced by healthcare provider organizations, e-health produced within this paradigm has been oriented toward providing added customer service, increasing healthcare delivery quality, and containing costs rather than selling products, services, or advertising access. Although the CHI paradigm is designed to address needs of the individual user, it inherently takes an organizational perspective that imposes a passive view of users as consumers rather than suppliers of information. Thus, providing PHRs and integrating external patient data typically have not been high priorities for e-health created within the CHI paradigm.

A New Paradigm: Personal Health Informatics

Each of the three historical paradigms identifies an area of need in the population of e-health users. However, none addresses the total range of needs that e-health users identify. These include communicating with physicians and clinic staff, arranging services (e.g., scheduling appointments and renewing prescriptions), checking bills and making payments, viewing lab results, accessing records of procedures/tests/immunizations, receiving online alerts, monitoring chronic conditions, and interacting with online support communities (e.g., patient chat groups), in addition to accessing encyclopedic health information (Fox, 2005; Harris, 2000; Taylor & Leitman, 2002). Furthermore, individuals want the ability to control their personal data, to ensure that recordkeeping aspects of e-health are interoperable and portable, and to be assured of privacy and security (Holt, 2005; Taylor, 2004). In order to capture the complete set of knowledge and skills that will be necessary to encompass e-health services, I propose that it is essential to consider a new paradigm that can integrate best practices of the three paradigms discussed previously. Because of its user-centered focus, which views e-health from the perspective of individual users in the role of information provider as well as information consumer, I term this paradigm *personal health informatics* (PHI). PHI is defined as the knowledge, skills, practices, and research perspectives necessary to develop e-health that is effective, efficient, encompassing, and user-centered.

FOUNDATIONS OF A PHI PEDAGOGY

A conceptual model of PHI is presented in Figure 1. Four structural components define the content of PHI within three focal areas. Web e-service infrastructure is the hardware, software, and networking capabilities that support all e-health functions. Because e-health is primarily service-oriented, the informatics focus in PHI centers on infrastructure that is specialized for e-service presentation and delivery. Personal health management and user-centered development comprise the personal focus that PHI presents to individual users. User-centered development methods provide tools for eliciting user needs, designing solutions, and evaluating the utility of these solutions in meeting user needs. Personal health management addresses individual practices as well as psychological, social, and cultural aspects of the management, storage, and retrieval of personal health information. Personal health management and the health informatics domain combine in the healthcare focus of PHI. Content within the health informatics domain addresses the skills, knowledge, and surrounding use of IT within the subject health area(s).

In order for students to enter the health informatics and IT workforce with sufficient preparation to develop e-health that can fully support individual users' needs, training will be needed in each of the structural areas of PHI, which are presented in Figure 1. None of these areas is beyond the grasp of graduate or advanced undergraduate students; however, the combination of content areas represents a set of specializations that rarely are combined in current curricula. Thus, for the time being, PHI training necessarily will have to be pieced together from existing textbooks, research reports, and industry cases. I address each curricular area in the following sections.

E-Service Web Infrastructure Curriculum

E-health in the PHI paradigm is essentially a form of e-service, defined as services provided over electronic networks such as the Internet (Rust & Lemon, 2001) Thus, most texts that focus on development and administration of Web infrastructure for e-services will be applicable to PHI. Texts that address the general topic of e-commerce typically are less applicable, as large portions of these dwell on factors that are not central to e-health, such as general business models, marketing and consumer behavior issues, and methods for increasing online sales. Web infrastructure changes rapidly, so texts in this area must not be allowed to become stale. New texts should be chosen, based upon inclusion of emerging technologies such as mobile devices, speech recognition, and natural language parsers, as well as effective coverage of more mature technologies such as XHTML, database connectivity, and network communications.

User-Centered Development Curriculum

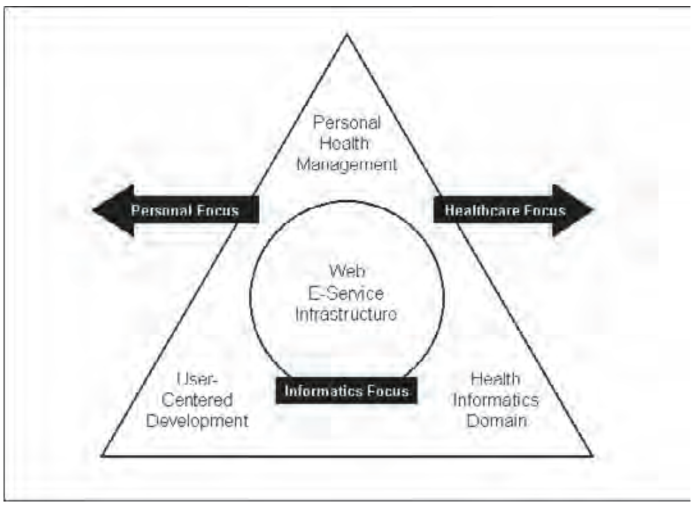
User-centered development refers to a software development process that incorporates the user viewpoint, assesses user needs, and validates that user needs are met. There are two complementary approaches to this process that will be valuable to students of PHI. The first of these is the design guide approach. Design guides emphasize generic or context-specific design principles, such as the following example for use of color on Web sites:

In the physical world, color has embedded meaning, usually culturally based. In the American culture, at a simplistic level, red means stop and green means go, red means hot and blue means cold, black means death and white means birth, and green represents money and greed. These embedded meanings of color differ from culture to culture; for instance, the color white is symbolic not of birth, but of a funeral in Japan or India. (Lazar, 2001, p. 158)

Design guides are valuable in establishing general guidelines in conventional use of design features and in broadening students' appreciation of ways that others' perspectives may differ from their own.

The second approach emphasizes the design process by establishing an algorithmic approach or how-to guidance for developing functionality and user interface features. For example, the

Figure 1. Conceptual model of key PHI structural components



SALVO method (Wilson & Connolly, 2000) guides students through five development stages. The first three of these are planning stages in which designers specify user needs, adopt a technology-specific design guide, and leverage appropriate matches between technology and user needs in order to enhance user abilities and to support user disabilities. Once planning stages are initially completed, developers perform visualization techniques, including mockups to elicit user feedback early in the development cycle. Visualization is followed by an observation stage, in which designers observe and evaluate attempts by users to perform representative tasks using the mockup or completed software. SALVO takes an iterative approach to development, and it is accepted throughout the development cycle that prior stages and sequences of stages can be repeated, as necessary. The design process approach provides PHI students with a methodology and specific tools for customizing e-health in order to meet the needs of a specific user audience, which may differ substantially, for example, between a macular degeneration support group and the general public.

Personal Health Management Curriculum

Personal health management is an emerging area of research, and new findings likely will emerge that add appreciably to current knowledge in this field. Although it may be too early to look for texts that focus on personal health management and good empirical studies, and cases are available that would be highly applicable to PHI. One exemplar is a study by Moen and Brennan (2005) that investigates strategies for health information management (HIM) that are applied by individuals in their own homes. Their research identifies four HIM strategies:

- **Just in Time.** Information is kept with the person at most times in anticipation of immediate need.
- **Just at Hand.** Information is stored in familiar, highly accessible locations, often serving as a reminder.
- **Just in Case.** Information is stored in order to be accessible in case of need, but out of view.
- **Just Because.** Information is retained in interim storage pending a decision to file or to discard it.

This study presents several important implications for developers of HIM software that would be highly relevant for PHI students to consider and discuss. These include effects of tradeoffs between information accessibility and visibility on the success of individuals' efforts to coordinate HIM, implications of the high degree of reliance that individuals place on paper-based records, and design decisions that would be necessary in order to align electronic records to support the differentiated storage strategies that individuals are found to prefer.

Health Informatics Domain Curriculum

The health informatics domain encompasses use of IT to support the delivery of healthcare services (including communication, coordination, logistics, and other business processes) as well as applicable practices and procedures that surround the use of IT in the healthcare setting. It is important to recognize that while the health informatics domain is a key component of PHI, it is not the central element of e-health from the user's perspective. In order to preserve a user-centered focus, it is essential that distinctive components of PHI pedagogy should not be replaced by health informatics courses that focus on IT from a perspective that is internal to the organization (i.e., managerial or technical). Where PHI is taught within an existing health informatics program, a range of courses already will be in place in order to effectively cover topics in the domain. Where PHI is taught in technology programs, topics in the health informatics domain may be covered effectively by traditional health informatics or by medical informatics texts.

Suggested Modes of Curriculum Organization

As discussed previously, PHI can be integrated into the undergraduate or graduate curricula of existing health informatics programs or used to augment offerings in technology programs, such as information systems, computer science, or information science. As a specialization within a health informatics program that already incorporates domain coverage, in-depth PHI training can be achieved by a curriculum that offers one course-equivalent each in user-centered development and personal health management and up to two courses in Web e-service infrastructure. If two courses are offered, the first should focus on basic development methods, including XHTML, XML, and simple database connectivity, and the second on administration issues and advanced development methods including Web services, mobile applications, and emerging technologies.

As an augmentation to a technology program that already incorporates Web e-service infrastructure coverage, substantial depth of PHI training can be achieved by offering a specialization with one course equivalent in each of user-centered development, personal health management, and health informatics. A final alternative is to offer a basic introduction to each of these subjects within a single course equivalent. Although not ideal for career preparation, this alternative could be an effective way to direct technology-area undergraduates toward careers and further educational opportunities in the field of health informatics and IT.

CONCLUSION

Although not always profitable for commercial vendors, e-health has been a success for individual users. Major healthcare provider organizations have adopted e-health as a means of attaining competitive equity in the marketplace, thus providing greater opportunities for individuals to use e-health and creating a demand for trained health informatics and IT professionals to work in e-health. PHI presents a conceptual perspective that incorporates the best features of preceding paradigms and provides a framework for improving instruction in e-health design, development, and administration.

REFERENCES

AHIMA (2005). The role of the personal health record in the EHR. *Journal of AHIMA, 76*(7), 64A–64D.

Campbell, R.J. (2005). Getting to the good information: PHRs and consumer health informatics. *Journal of AHIMA, 76*(10), 46–49.

Eysenbach, G. (2000). Consumer health informatics. *British Medical Journal, 320*(7251), 1713–1716.

Eysenbach, G. (2001). What is e-health? *Journal of Medical Internet Research, 3*(2), e20.

Fox, S. (2005). Pew Internet & American Life Project. Retrieved December 13, 2005, from http://www.pewinternet.org

Harris Interactive. (2000). *Healthcare satisfaction study* (Harris Interactive/ARiA Marketing Final Report). Retrieved December 13, 2005, from http://www.harrisinteractive.com/news/downloads/HarrisAriaHCSatRpt.PDF

Holliday, I., & Tam, W. (2004). E-health in the East Asian tigers. *International Journal of Medical Informatics, 73*, 759-769.

Holt, M. (2005). An archaeology of the commercial PHR movement. In *Proceedings of the QTC-CGU Symposium: The many faces of person-centric electronic health systems: A national symposium to examine use-cases for healthy, chronically ill and disabled user communities*, Claremont, California, USA

Itagaki, M.W., Berlin, R.B., & Schatz, B.R. (2002). The rise and fall of e-health: Lessons from the first generation of Internet healthcare. *Medscape General Medicine, 4*(2). Retrieved December 13, 2005, from http://www.medscape.com/viewarticle/431144_Print

Kloss, L. (2005). Greater urgency, sharper focus for PHRs. *Journal of AHIMA, 76*(10), 27.

Krane, D. (2005). Number of "cyberchondriacs"—U.S. adults who go online for health information—increases to estimated 117 million. *Healthcare News, 8*(5), 1–7. Retrieved December 13, 2005, from http://www.harrisinteractive.com/news/newsletters_healthcare.asp

Kwankam, S.Y. (2004). What e-health can offer. *Bulletin of the World Health Organization, 82*(10), 800–802.

Lazar, J. (2001). *User-centered Web development*. Sudbury, MA: Jones and Bartlett.

Lazarus, I.R. (2001, June 1). Separating myth from reality in e-health initiatives. *Managed Healthcare Executive*, 33–36.

Martin, S., Yen, D.C., & Tan, J.K. (2002). E-health: Impacts of Internet technologies on various healthcare and services sectors. *International Journal of Healthcare Technology and Management, 4*(1,2), 71–86.

Moen, A., & Brennan, P.F. (2005). Health@Home: The work of health information management in the household (HIMH): Implications for consumer health informatics (CHI) innovations. *Journal of the American Medical Informatics Association, 12*(6), 648–656.

Oh, H., Rizo, C., Enkin, M., & Jadad, A. (2005). What is e-health(3): A systematic review of published definitions. *Journal of Medical Internet Research, 7*(1), e1.

Rovenpor, J. (2003). Explaining the e-commerce shakeout. *E-Service Journal, 3*(1), 53–76.

Rust, R., & Lemon, K.N. (2001). E-service and the consumer. *International Journal of E-Commerce, 5*(3), 85–102.

Taylor, H. (2004). Two in five adults keep personal or family health records and almost everybody thinks this is a good idea: Electronic health records likely to grow rapidly. *Health Care News, 4*(10), 1–5. Retrieved December 13, 2005 from http://www.harrisinteractive.com/news/newsletters_healthcare.asp

Taylor, H., & Leitman, R. (2002). Patient/physician online communication: Many patients want it, would pay for it, and it would influence their choice of doctors and health plans. *Health Care News, 2*(8), 1–4. Retrieved December 13, 2005 from http://www.harrisinteractive.com/news/newsletters_healthcare.asp

WHO. (2005). *E-health: Report by the Secretariat*. World Health Organization, 58th World Health Assembly.

Wilson, E.V., & Connolly, J.R. (2000). SALVO: A basic method for user interface development. *Journal of Informatics Education & Research, 1*(2), 29–39.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 1, Issue 3, edited by J. Tan, pp. 69-76, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 2.2
Web Service Design Concepts and Structures for Support of Highly Interconnected E-Health Infrastructures:
A Bottom-Up Approach

Adamantios Koumpis
ALTEC S.A., Greece

ABSTRACT

In this chapter we present organizational aspects that appear when considering the case of interconnecting and integrating different compartments of a modern hospital. While the information and communication technologies provide advanced and powerful means for the creation of coherent information supply services, such as the Web service and ontology technologies, there is a lack of appropriate organizational metaphors that will enable the successful assimilation of these technologies, helping to aid in the improvement of critical cost parameters that concentrate a large part of the hospital's management resources, while also helping to improve the knowledge capital and the intangible and immaterial assets of any particular hospital, which are considered as the most essential and scarce resource. In this paper we presents a technology-based approach for solving interoperability problems at the service level, and we deliberately adopt a problem-solving approach that has successfully been adopted by the European IST Project ARTEMIS.

INTRODUCTION

What we more intensively experience is that the Web is moving from being a collection of pages toward a collection of services that interoperates through the Internet (Paolucci, Kawamura, Payne, & Sycara, 2002). According to the same source:

Web services provide a new model of the web in which sites exchange dynamic information on demand. This change is especially important for the e-business community, because it provides an opportunity to conduct business faster and more efficiently. Indeed, the opportunity to manage supply chains dynamically to achieve the greatest advantage on the market is expected to create great value added and increase productivity. On the other hand, automatic management of supply chain opens new challenges: first, web services should be able to locate automatically other services that provide a solution to their problems, second, services should be able to interoperate to compose automatically complex services. In this paper we concentrate on the first problem: the location of web services on the basis of the capabilities that they provide.

In this chapter our concern is the application of some Web service design concepts and structures in order to support highly interconnected e-health infrastructures; though health for sure constitutes a special application domain with several idiosyncrasies and singular characteristics, there is a great extent of paradigms and analogies that can be drawn from the area of e-business and supply chain management.

We believe that our contribution by means of the work conducted under the European IST project ARTEMIS (ARTEMIS, 2004), which was lead by professor Asuman Dogac of the Middle East Technical University, constitutes a success story that can be followed to form the basis of several VE ventures in the area of e-health as well as in other business domains.

Currently, the majority of the interoperability problems of medical information systems are two-fold (Bicer, Banu Laleci, Dogac, & Kabak 2005a; Bicer, Kilic, Dogac, & Banu Laleci, 2005b):

1. First there are multiple, incompatible, proprietary approaches to connecting disparate applications to clinical networks and information systems. As a result, for example it is not possible to integrate electronically the clinical patient records with critical emergency control information.
2. Secondly, when there are standards to achieve interoperability, there are more than one standard to represent the same information, which in turn creates an interoperability problem. For example, GEHR, CEN 13606 and openEHR are all standards for patient electronic health records.

The proposed model provides the most important entity of the healthcare industry, namely the hospital, with an ideal platform to achieve difficult organizational and technology integration problems. The proposed services as developed in our project ARTEMIS allow for seamless integration of disparate applications representing different and, at times, competing standards, thus allowing for a service to be invoked *on demand* pervasively by business processes, applications or people to fulfil a particular function. The latter forms the most important innovation of the presented work and a tangible contribution towards smarter hospitals that are capable to build dynamic information exchange and sharing infrastructures that might have the form of virtual enterprises. Though from a legal point of view there are many problems and difficulties, it is however important that such a goal will guide the investments in the health industry in Europe.

As will be further described, the innovation of our approach comes to the fact that our approach for design and management of services is implemented in a distributed service infrastructure according to a *preplanned usage of a multiple service actors' scheme*. The term distributed service infrastructure is used for description of an environment with the following characteristics:

1. It consists of a number of service flows that are executed using resources of several sites simultaneously.
2. That service flows communicate with each other by exchanging messages over a commonly agreed network of participants (in our case it is the network of the hospitals

or hospital units involved in the provision of a service).

Our efforts may be viewed from within the perspective of building the service flow execution kernel for mobile agent applications that may regarded as the high-end of the foreseen application service providers (ASP) market in terms of *aggregating* functionality requested by the particular differentiated users of the distributed service environment. In this respect, the approach we employ address the following two needs:

1. From an *operational* viewpoint, it focuses on the intersite aspects (timing and security) for remote interoperability of the participating hospital services. Intrasite, it will focus on the dynamic adaptation of the application to changes in the environment of a single hospital (unit).
2. From a *methodological* viewpoint, it focuses on the way to capture and validate dependability requirements and validates these requirements, on the way to derive from requirements the structure of the modeling approach, and on the use of modeling to drive the development and the assessment of the proposed solutions.

The building of the proposed services is based on:

- **Service elements:** Regarded from the service designer's perspective as these concern reusable elements that may be used for developing new services or enhancing/changing the functionality of existing (operational) ones
- **Service pages:** Entities upon which the hospital user (i.e., a doctor, a nurse or an administration employee) may regard services either for carrying out customization activities such as personalization of the access-to-service interface for example, for different user categories and different types of usage (data entry, retrieval of data, sophisticated query formulation and processing, etc.)

The approach taken helps in the creation of a significant competitive advantage and market knowledge and creates first-mover advantage in the addressed transition towards *service-oriented architectures.* This know-how is faster transferable from within the operational environment to those key divisions that will be acting as uptakers and adopters. Of course, a key objective has been how to get improved ideas to those who can effectively apply them. The main focus is to gain the technological capabilities and the necessary means (i.e., methods, practices and software components) so that any new services will be affordably priced for a segment of the hospital market that has been *largely unable to afford such services—namely small and medium sized hospitals with fewer than 200 employees such as the case of regional or prefecture-level hospital and / or health centers.*

The access-to-service environment under implementation enables the users of the service platform to:

- Establish an overall service flow direction, by means of providing linkage to a set of pre-programmed resources that are executed in the distributed (Internet) environment, such as ERP, patient recording applications, and so forth.
- Acquire resources for a particular service property which may be it a service flow, a service element, or a service "page".
- Provide "capabilities" to a service flow by means of integrates both structural and behavioral aspects from within a single perspective, which will be utilized to instantiate the actual service delivery at the end user's point.
- Execute the service flow by means of utilizing resources to accomplish the particularly

assigned service scenario.

The last may also be regarded also the "bottom-line" for the actual service delivery by a particular service flow to support the purpose of the latter's establishment (i.e., the reason for existence of that particular service flow in the overall hospital value chain).

The Artemis Project

Medicine is one of the few domains to have some domain knowledge in a computable form which can be exploited in defining the semantics of Web services. This semantic knowledge exists in "controlled vocabularies", or "terminologies". Some vocabularies are rich semantic nets, such as SNOMED-CT while others such as ICPC and ICD are little more than lexicons of terms. Although such vocabularies do not express all the possible semantics of basic concepts—they are generally limited to terms, definitions and some semantic relationships—they offer significant value in terms of expressing the semantic of Web services.

ARTEMIS is a project that runs under the 6th Framework Programme of the European Commission, and aims to develop a semantic Web services based interoperability framework for the health care domain. The interoperability problems of medical information systems are two fold:

- First there are multiple, incompatible, proprietary approaches to connecting disparate applications to clinical networks and information systems. As a result, for example it is not possible to integrate electronically the clinical patient records with billing information.
- Secondly, when there are standards to achieve interoperability, there are more than one standard to represent the same information, which in turn creates an interoperability problem. For example, GEHR, CEN 13606 and openEHR are all standards for patient electronic health records. The CEN 13606 model, for instance, describes a hierarchical representation for all data using the cluster and data_item classes. The GEHR model does the same thing, with its hierarchical_value and hierarchical_group classes.

Figure 1 gives an overview of the ARTEMIS architecture to enable the peer-to-peer networks as a platform for publishing, discovering and invoking semantically enriched Web services that wrap Medical Information System Applications based on the semantic Web initiative. To achieve this objective the following basic components will be provided by the ARTEMIS project:

- Semantically enriched and secure P2P Web service environment for medical information systems
- Semantic wrapper for Web service creation and composition that adapts the medical information system applications
- Service-based access layer for electronic health records
- User friendly interfaces for the healthcare organizations, such as for publishing, discovering and composing Web services

These components are presented in more detail in Figure 2.

Furthermore, in our approach we provide a P2P platform that with enhanced capabilities for publishing, discovering and invoking *semantically enriched medical Web services in a highly secure framework*.

This component will be a base for the medical informatics Web services to operate on. The objective of this component is to facilitate the discovery of medical Web services both from the individual peers in a network and also from the public service registries.

However neither existing P2P architectures nor the Web service registries provide facilities

Figure 1. An overview of the service-oriented architecture in the ARTEMIS project

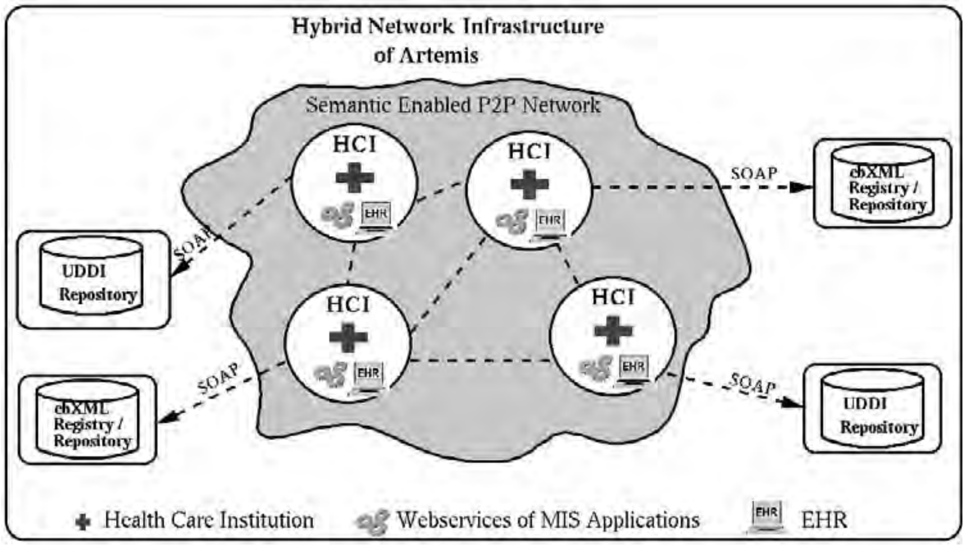
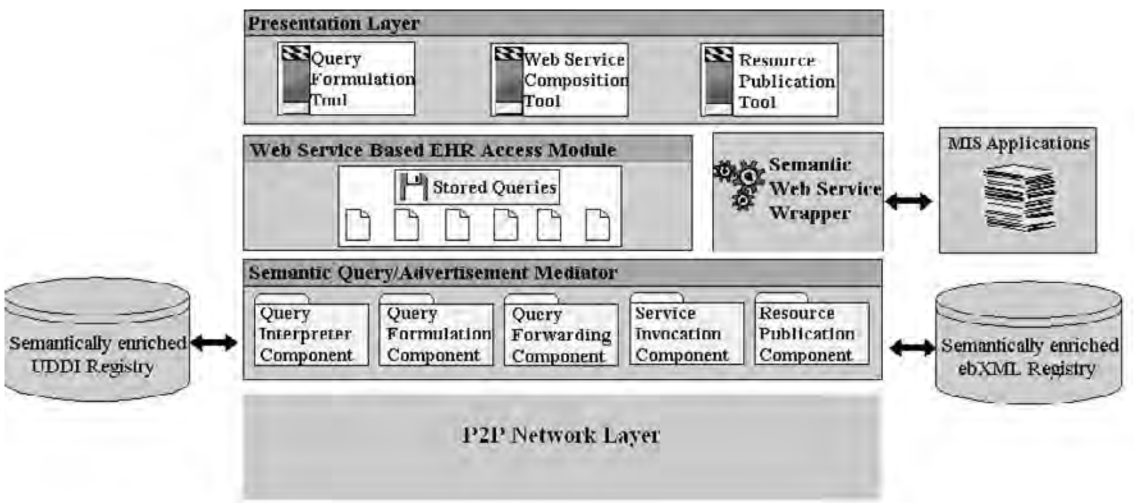


Figure 2. ARTEMIS components overview



for exploiting the semantics of the services, and P2P security models are still emerging mainly in the context of the WS-Security standardization roadmap (Dogac, Bicer, Banu Laleci, Kabak, Gurcan, & Eichelberg, 2006).

In order to achieve an integrated hybrid semantic based infrastructure, this component will be developed in three main steps. In the first step existing Web service registries will be enhanced to store domain specific ontologies, and services based on these ontologies; the second step will be incorporating service semantics in to P2P networks, particularly developing semantic based advertisement and discovery mechanisms for P2P networks. In the final step, semantically enriched service registries will be integrated to the P2P architecture.

POSITIONING OF THE SYSTEM AND THE APPROACH

Having in mind that some of the most essential problems that users, administrators, developers and vendors of information supply services in the health discipline, as well as in every application and service field, face today may be viewed under the common denominator of "interoperability" problems, the presented approach illustrates possible ways to address these problems. A design goal of the project is to provide a cohesive technological infrastructure independent of any specific implementation pathway and to contain features that are effective and easy to use in a broad range of representative networked service environments which may be subject to variable configurations. For this reason we recognize the following types and broad categories of users:

1. Platform and service vendors (may concern IT companies, content providers or—in case they exhibit competencies in any of them—as a specialization of the broker category)
2. Professional health service providers (as a specialization of the broker category)
3. Service developers (as a specialization of the broker category)
4. Service administrators (as a specialization of the broker category)
5. Service end users (i.e., hospitals—either public or private owned ones)
6. Information technology managers (as a specialization of the previous end user category)

These users participate in one or more of the following four stages in the development and usage of the health-based service infrastructures:

- **Establishment:** Implementing and deploying the presented service approach across the health information "supply chain."
- **Build:** Exercising the service elements to define a baseline service flow configuration (establishing the exchange paths between known service sources and targets as well as the various filtering mechanisms involved).
- **Operation:** Operating the service flow infrastructures.
- **Maintenance:** Exercising the introduced concepts to define changes in the distributed service configuration (e.g., to cover changes as "small" as the addition of new service elements in the overall service configuration and as "large" as merger with or replacement by another configuration such as in the case of replacing a service flow with a group of supplying service flows loosely linked and using a new distributed management scheme); this is a quite complex issue for which description may be regarded as outside the scope of the project. It concerns the "reverse" engineering of a service into a set of constituent services that should be chosen for support of an, for example, localization exercise (a global service gets localized

Table 1. Shows the forseen added values to the users

		Foreseen Added Value to the Users		
User category	**Stage**	**Problem or need**	**Tools and repositories**	**How the system promotes better service utilization**
Platform and Service vendors	Build	Must subscribe to standards for intervendor interconnect	– Web service infrastructure – Common Repository Facility – Tools for modeling, development, deployment and service management	– System provides a common "backplane" for pluggable subsystems. – It may be exploited as a globally usable notation for metaservice exchange protocols which enables flexible distribution of distributed services over a heterogeneous collection of information systems.
Professional Service Providers	Build	Must accumulate and reuse elements from service engagement	Third party and in-house tools that apply metaservices to concrete service-base catalogs and vice versa	Reusable, editable, and extensible metaservice should provide a first-level "asset base" that builds (new) value. This base of reusable elements starts a self-reinforcing feedback loop with continually increasing returns improved by engagement productivity for the users.
Professional Service Providers	Maintenance	Must modify Service process configuration: knowing what and where to modify; knowing dependency closure	Third party or in-house tools to manage reconfiguration editing of a service flow	System exposes the information required to modify a service flow model. Service context definition and self-describing features for the service flows are used to isolate dependency relationships.
Professional Service Providers, Dielemma Service Administrators	Maintenance	Must integrate existing tools and data which adhere to standards other than service flow model into a distributed service configuration environment.	Tools based on ability to incorporate metamodels of services and alternate service definition practices and standards.	System does or can subsume non-service representations. For example, may be elaborated in the future to contain any XML-based service model with a focus to domain-specific characteristics.
Service Administrators	Build	Must establish and manage expressions, relationships, and lineage over multiple service-based schemata.	Tools that use built-in facilities to define schema content, relationships, and lineage.	System design is based on need to manage such information at multiple levels. The basic Web services will have to be designed to allow navigation of metaservices correlated to schemata.
Service Administrators	Maintenance	Must add, subtract, re-partition, reallocate, or merge service resources in deployment configuration.	Service management tools.	System consists of models of metaservices that assist in making such changes and allow impact of these changes to be assessed.

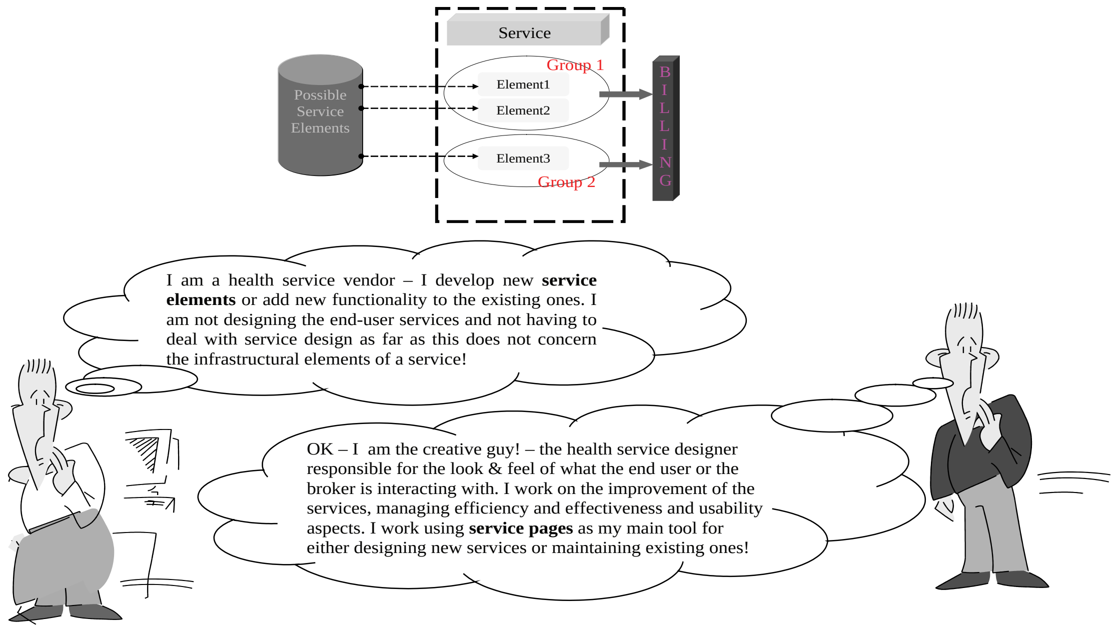
and a set of local service points are now assigned the responsibility for running the service). In the following we present some usage scenarios that illustrate activities in the *build* and *maintenance* steps that clearly demonstrate the value addedness of the approach.

In regard to positioning the added value of the project in terms of linkage with the business opportunity for the project and its real applicability and potential for adoption in the health sector within a supportive uptake environment that would favorably sustain business development in that specific area, we note that there is certain potential in coupling the project work with developments in the application service provider (ASP) market segment. In our context we adopt the usual definition for an application service provider as a 3rd-party service firm which deploys, manages, and remotely hosts a both the various application and service portfolios through centrally-located servers in a "rental" or lease agreement as well as the related business model for operating them. Our work in the project helps a revisit to the topic of ASPs for two main reasons:

- We now see that there is a strong future for ASP related businesses.
- A great deal has changed on the ASP competitive landscape.

Over the past months multiple variations on the basic ASP model have emerged which warrant review and analysis. In addition, the ASP business model is beginning to face serious challenges regarding its viability and broader market acceptance, linking those models to the potential of providing through ASPs the means for operating into a networked enterprise environment.

Figure 3. Distinction of roles in the operation of the system



In its "purest" sense, the ASP model invokes the delivery of software as a service. In exchange for accessing the application, the client renders rental-like payments; in this respect, an ASP facilitates a remote, centrally-managed "rent-an-application" service for the client. The emphasis is placed on the *use and not on the ownership* of the application. The client no longer owns the application or the responsibilities associated with initial and ongoing maintenance. The client, through an Internet browser, accesses remote, centralized computer servers hosting the application. Only the results from the application are managed locally by the client.

Currently the ability of an ASP to sell "mission-critical" solutions to early ASP adopters hinges to a large degree on whether the application being hosted was designed specifically for delivery over the Internet. In point of fact, many of the "mission-critical" packages sold by highly respected vendors were *not* developed for a hosted / services delivery methodology.

Where the added value of our approach comes to the stage relates to the fact that our approach for design and management of services is implemented in a distributed service infrastructure according to planned (actually: envisaged) usage of a multiple actors scheme. The term distributed service infrastructure is used for description of an environment with the following characteristics:

1. It consists of a number of service flows that are executed using resources of several sites simultaneously.
2. The service flows communicate with each other by exchanging messages over a commonly agreed network of participants (in our case it is the network of the health organizations i.e., hospitals and regional health centers involved).

In this respect our efforts in the project may be viewed from within the perspective of building the service flow execution kernel for mobile agent applications that may regarded as the high-end of the foreseen ASP market in terms of *aggregating* functionality requested by the particular differentiated users of the distributed service environment.

SETTING THE STAGE FOR THE VE BUSINESS DESCRIPTION

According to "The Emerging European Health Telematics Industry" market analysis report (2000), developed by Deloitte (2000) on assignment by the European Commission-Directorate General Information Society, "the healthcare sector represents 6% of the overall current European IT market" and the "consolidation on the supply side" constitutes on the most important "vital conditions" for the market growth.

The ARTEMIS platform can have a significant role in e-health, as it may form the basis for realizing most of the e-services that all health institutions will be providing to their beneficiaries.

The ARTEMIS platform aims at delivering a *semantic Web services based interoperability support* and an associated software engineering methodology specific to the issue of e-health data, systems and service interoperability. The approach of the ARTEMIS project to the addressed problem is holistic. It targets in satisfying the needs of all stakeholders in the provision of such e-services, namely health sector IT staff (administrators), health sector domain experts (medical experts and doctors), managers of health sector authorities (managers) and patients, businesses that do business with health authorities (e.g., in terms of hospital procurements, etc.) or any other health sector employees (end users). The approach of ARTEMIS also covers all steps and processes involved in the e-health services provision, as it will be demonstrated further on in the project, and shall cover other related parties, such as insurance companies.

In this chapter, the opportunities and drivers for the project are analyzed, the ARTEMIS platform is positioned amongst other relative products and technologies, the available services are outlined, and lastly, the Unique Selling Proposition of the ARTEMIS project is presented.

Business Opportunity: Business Drivers

According to the above mentioned survey the drivers for the provision of e-services as those we are developing in ARTEMIS (2004) include:

- As the market is made up of hardware, software and services, the role played by services in 1998 was an estimated 26 % while the projection for 2002 was for this to raise to 37 %, becoming the largest component in terms of revenue ahead of the other two, and with an incremental pattern
- "The main trend which can be observed in European hospitals, which presents a challenge in the coming years, consists of integrating the different areas of the HIS (medical, nursing, technical or administrative) into a common architecture.
- However, when looking more closely at hospital information systems, the picture is more fragmented. The analysis confirmed that currently there is no architecture which can integrate all modules of the HIS, but they are often run alongside each other in the forms of the clinical patient record, the laboratory information system, the patient administration system, etc." (p. 188).
- "Apart from internal integration of HIS, hospitals are also heading towards external integration on a regional, national and even an international level. Here, the emergence of care networks, involving many different healthcare actors, hopes to optimise patient management" (p. 192).

In the light of the above the main opportunity concerns an apparent need for common service platforms. Furthermore, it was identified that the main need that ARTEMIS aims to fulfill is to enable *otherwise disconnected systems* to communicate *over different standards*. This will result in services of one health institution being available to another, which will carry all users of the ARTEMIS platform to become members of the same domain.

Since ARTEMIS works over standards and enables to map and convert them to each other, it lowers the negotiation cost, when one system is integrated to or with another. As a result, some services which may require huge costs in terms of investment to hardware/staff may be presented to other institutions, resulting in much better *price discrimination* for suppliers, and new service opportunities for consumers of such services.

This service-oriented business model enables implementers of products to have a much larger market by selling / renting them as services over the ARTEMIS "network".

The basic business opportunity for the ARTEMIS project lies in that the health sector spends on integration forms—including cost of designing, management and handling—is massive. Bureaucracy is the inevitable consequence of the traditional method of handling the extreme volume of information and knowledge derived from them, while the low quality of service is another collateral outcome of this situation.

ARTEMIS aims at capitalizing on this opportunity by offering the means to the health authorities and institutions to reduce involved costs by the implementation of e-services (note: we refer to e-services for simplicity reasons; the correct terms for the ARTEMIS services is "semantically enriched Web services") and associated service management infrastructures. The appropriate design of e-services, according to ARTEMIS approach, will increase cost-efficiency and effectiveness. Moreover, the interaction between the health sector and businesses will be improved when the ARTEMIS tools are combined with best practices. Additionally, the platform is open to enhancements with facilities such as open source development paradigm and self-regulated communities of practice.

Identification of the ARTEMIS Platform

Taking into consideration the various software solutions available to health authorities, it is evident that the market is quite chaotic in the sense that there are no well-established products on the market for healthcare-specific enterprise application integration (EAI), but the rather inflationary estimates we mention previously (have) attract(ed) many vendors. This means that there are many projects, many announcements, but very few results so far. This is where ARTEMIS comes and fills the gap.

Only for Germany, where one of the partners of the ARTEMIS project came from, are two indicative projects to take into consideration, namely:

- The Bit4Health project, a big national initiative to establish a smartcard based PKI for the healthcare sector which would be used for electronic prescriptions, digital signatures (e.g., patient consent) and, at some point in the future) as an "access key" to a national electronic healthcare record, which is, however, totally unspecified at the moment. (http://www.dimdi.de/de/ehealth/karte/bit4health/index.htm)
- The "doctor 2 doctor" (D2D) and "VCS" communication standards, which aim at integration of private practices. While D2D is promoted by an organization representing medical doctors, VCS is promoted by healthcare IT industry. (http://www.kvno.de/mitglieder/d2d/; http://www.vdap.de/html/vcs/vcs.html)

Both projects have at most a few hundred installations so far, which is a very limited success given there are about 130,000 private practices in Germany.

ARTEMIS as a stand-alone product is characterized as a user-friendly service development tool, which can be complementary to the above software solutions. But more importantly, it can be offered as a solution, and furthermore it features a service platform which interacts with the various legacy systems and assists deployment parties in the generation and manipulation of online (transaction-based) services. Using ARTEMIS as a stand-alone product will simplify the development and maintenance of these e-services. On the other hand, using ARTEMIS with another installed IT system will not create any technical or process problems in the integration stage, due to the platform's open architecture.

The ARTEMIS platform constitutes an environment that enables health sector professionals in different levels (administrators, experts / doctors and managers) to develop and maintain services for both their own institution and for other authorities as end users. In this environment an employee with the necessary skills and domain knowledge will be able to use a predefined form template or create a new service in order to implement a new functionality or to edit an existing service.

The employee will be assisted by the ARTEMIS development environment (i.e., the ARTEMIS platform), which will automate parts of the development process. At the end of the development procedure the employee will be able to activate the service even if the connection to the back-end has not been implemented yet. The platform will retain all collected data and once the IT staff implements the connection to the back-end the data will be processed.

The types of the envisaged users of the system can be listed as follows:

- Users should know the backend legacy system.
- Users should know the input output structures of the applications from which the Web services will be created.

Additionally, we can open ARTEMIS certificate programs such as:

- ARTEMIS healthcare ontology engineer
- ARTEMIS healthcare service engineer
- ARTEMIS P2P engineer

Unique Selling Proposition of the ARTEMIS Services

A holistic approach in the management of services in the health sector reveals the Unique selling proposition of the ARTEMIS system. The system's approach to the introduction of semantically enriched Web services in the health sector very much takes into account, that the addressed area is a process and person related issue. It has to involve individual participants from different organizations and with different backgrounds, cultural elements and adopted practices, where continuous and situated elaboration and maintenance of the required knowledge through the involved parties is necessary (actually: a precondition if not a sine qua non). This means a shift from a technology oriented paradigm to an activity oriented paradigm, where involved human beings, administrative processes and technical resources represent the equally eligible units of analysis and where system design follows the concept of socio-technical systems as holistic and dynamic structures.

The ARTEMIS project suggests that by adopting the above *holistic approach* for the introduction of semantically enriched Web services in the health sector, most of the following problems, that health sector stakeholders face can be reduced or even eliminated:

- Complexity in creating e-services and most importantly difficulty in encapsulating domain expertise in them
- Difficulty in interoperability with existing IT systems within the organization and with external IT systems of other organizations
- Lack of user-friendliness for the end-user in the service creation phase
- Lack of mechanisms to encapsulate the knowledge as a whole also by transforming implicit knowledge into an explicit form. ARTEMIS shall provide tools for creating and mapping ontologies; of course, ARTEMIS aim is not proposing ontologies to the healthcare community, but providing all the necessary technology means so that if ontologies are available, then be also exploited to facilitate semantic interoperability.
- Lack of coherent process models for exploiting the use of "public" e-services within the health sector
- Organizational and cultural barriers, such as health sector employees' fear of new technology and new methods of work.

Target Market: A Dynamic Health Service Market

ARTEMIS supports the concept and the implementation of decentralization in European health care systems thus setting the stage for multiple different types of synergies and ventures to emerge and foster a culture of joining efforts and resources. Though such an exercise may face impediments from the legal environments on the one hand and from the tendency of private organizations to operate on short term profitability plans, it is yet an open challenge for all actors in Europe and internationally.

The concept of decentralization has become a cornerstone of health policymaking in an increasing number of Western European countries. Once confined only to the Nordic region and Switzerland, intergovernmental decentralization has now become a central principle of health policy in Spain and Italy, in a different context in the United Kingdom (to Scotland, Wales and Northern Ireland), and to a lesser degree in France, Portugal and more recently in Greece.

In certain respects, decentralization has become a synonym for a strengthening process of regionalization not just in the health sector, but in broader social as well as political arenas, serving as something of a cultural corrective in a Europe where more and more aspects of national sovereignty are being exercised at the supranational European Union level.

The best-known public administration effort to define decentralization is that of (Rondinelli, 1983), with a four-part, structurally oriented framework of:

- Devolution
- De-concentration
- Delegation
- Privatization

The economic argument in favor of decentralization typically has as its common denominator the belief that decentralization will create competition between different local actors, thus opening the door for local/regional-based small or medium-sized enterprises to innovate and provide cost-efficient services to the health care institutions for both sides (i.e., themselves and the particular health organizations). Not unlike economic assumptions about commodity markets generally, the economic argument here is that decentralization will enable different SMEs to provide different levels of *e*-health-related services, which will generate three economically-"prized" advantages:

- Greater efficiency in the service provision process (same is the case for the service conception, design and delivery processes)
- Greater choice for the health institutions of different alternative services
- Smaller complexity (bureaucracy) reduced times and improved responsiveness in the identification of a need and its fulfillment

Moreover, the complex intergovernmental relations created by the EU single market are seen as an extremely suitable opening to adopt such an economically-driven mechanism.

Of course, the basic and strongest from an economical standpoint argument regarding the disadvantages of decentralization focuses on the inefficiency and duplication of having multiple small service providers.

At this point it is interesting to carry out some type of sensitivity analysis regarding how many customers should an SME afford to have for ARTEMIS customers so that it can operate its business with different levels of profitability and amortization period. (Independently to this, there is also the concern (worry, rather) that such smaller operating units will be managerially weaker than larger ones.)

CONCLUSION

With the approach developed as part of the IST Project ARTEMIS, we provide the healthcare industry with an ideal platform to achieve difficult integration problems while also utilising the VE concept as the underlying basis for all different types of ventures and business partnerships. Our Web service model encapsulates already existing applications and access to documents in a standard way and incorporates service providers, service consumers and service registries.

It should be noted that currently most prominent Web service registries are universal description, discovery, integration (UDDI) and electronic business XML (ebXML). There are also very recent efforts to use peer-to-peer networks based on Web services. However both service registries and P2P architectures available do not provide semantically enriched search capabilities.

Furthermore, in ARTEMIS we provide extensions to these architectures to enable discovery of the Web services based on their semantic descriptions. As mentioned above, medicine is one of the few domains to have some domain knowledge in a computable form, which we exploit in defining the semantics of medical Web services.

The nature of business model research is extremely demanding as an exercise (Afuah, 2003; Alt & Zimmermann, 2001) and in many scientific research projects is not given adequate importance – however the reasoning behind a business model is not the understanding of a phenomenon, rather it is a problem-solution finding approach. In this respect, a particular technology may or may not facilitate the success of a business model; and similar to this, a particular market may or may not support the uptake of a particular solution as result of a more or less appropriate business model.

As a result of the approach taken, virtual enterprises (VEs) may emerge taking the form of joint ventures which will aim to provide operative and highly specialized value added services in all necessary e-health application fronts related to individuals, private companies and public health organizations, such as in the following application fields:

- Single point access to health services using technologies and methodologies developed within the project and which base on the ARTEMIS conceptual architecture and service modeling approach
- Heterogeneous hospital platform integration for support of online one-stop-shop services in various situations
- Design and lifecycle management of integrated hospital or other health care organization services from different back-offices, based on advanced interoperability frameworks and standards for the health sector

ACKNOWLEDGMENT

Work reported was partially funded by the European Commission under the contract IST-1-002103-STP, and under the e-Health priority. The authors would like to thank the European Commission for the generous support provided. More information about the ARTEMIS project the reader can find under in the official project Web site: http://www.srdc.metu.edu.tr/webpage/projects/artemis/.

REFERENCES

Afuah, A., & Tucci, C. (2003). *Internet business models and strategies*. Boston: McGraw Hill.

Alt, R., & Zimmermann, H. (2001). Introduction to special section – Business models. *Electronic Markets Journal, 11*(1).

ARTEMIS (2004). *IST project ARTEMIS: A Semantic Web service-based P2P infrastructure for the interoperability of medical information systems*. Coordinator Middle East Technical University, Software Research and Development Center, Turkey.

Bicer, V., Banu Laleci, G., Dogac, A., & Kabak, Y. (2005, October 19-21). Providing semantic interoperability in the healthcare domain through ontology mapping. In *Proceedings of the E-Challenges 2005 Conference*, Ljubljana, Slovenia.

Bicer, V., Kilic, O., Dogac, A., & Banu Laleci, G. (2005, October). Archetype-based semantic interoperability of Web service messages in the healthcare domain. *International Journal on Semantic Web & Information Systems, 1*(4).

Deloitte (2000). *The emerging European health telematics industry* (Market Analysis Report). European Commission-Directorate General Information Society.

Dogac, A., Bicer, V., Banu Laleci, G., Kabak, Y., Gurcan, Y., Eichelberg, M., & Riesmeier, J. (2006, February 12-16). Untangle the Web: Communicating through the Web services using semantics. In *Proceedings of the 2006 Annual HIMSS Conference*, San Diego, California.

Paolucci, M., Kawamura, T., Payne, T. R., & Sycara, K. (2002). Importing the semantic web in UDDI. In *Proceedings of the E-Services and the Semantic Web Workshop.*

Rondinelli, D. A. (1983). *Decentralization in developing countries*. (Staff Working Paper No. 581). Washington, DC: World Bank.

This work was previously published in Agent and Web Service Technologies in Virtual Enterprises, edited by N. Protogeros, pp. 327-340, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 2.3
The PsyGrid Experience:
Using Web Services in the Study of Schizophrenia

John Ainsworth
The University of Manchester, UK

Robert Harper
The University of Manchester, UK

ABSTRACT

The key aim of the PsyGrid project was the creation of an information system to ascertain and characterise a large, representative cohort of schizophrenics, beginning from their first episode of psychosis. The cohort was to be drawn from eight geographically dispersed regions of England, covering in total one-sixth of the entire population. In order to meet the current and future requirements we needed to build a secure distributed system, which not only could support remote data collection, but could also be integrated with other data sets, applications, and workflows for statistical analysis. We concluded that a service-oriented architecture was required and that the implementation technology should be Web services. In this article we present the design, deployment and operation of the PsyGrid data collection system as a case study in applying Web services to health informatics. The major problems we faced were related to the deployment of Web services into an existing network infrastructure, but overall found Web services to be the most suitable middleware technology.

INTRODUCTION

In the 1970's Archie Cochrane (Cochrane, 1972) and colleagues alerted the medical profession to the need to weed out subjectivity and anecdote from clinical practice. At the same time there was a move to improve the safety of medicines. Since then the evidence-based care movement has grown and is now accepted by most healthcare professionals to be best practice. However, there are serious problems with the evidence on which we base healthcare: it is expensive to produce; it takes a long time to produce; it takes a long time to influence clinical practice; it is based on clinicians' more often than patients' perceptions of important outcomes, which might not match; and it is crude—relating to the average participant and simple treatment definitions under ideal/trial conditions, often long ago—in other words, it gives a low-resolution picture of how your patient might respond to treatment. Most of the health informatics literature on electronic health records and "evidence into practice" is about weaving the existing evidence-base into healthcare decision-making. The role of clinical information systems in improving the evidence-base, however, has been neglected. This situation is changing: PsyGrid, which is funded under the UK Medical Research Council's E-Science programme, is focused on providing informatics employing e-Science principles for improving clinical trials and longitudinal studies in mental health research (Ainsworth et al., 2006).

PsyGrid employs service-oriented architecture computing techniques and technologies in the implementation of a system that will remove the barriers to epidemiological research. The process of epidemiological research has three phases—the establishment and characterisation of a large, representative cohort from a geographically distributed population; the integration of the cohort data with other data sources to provide additional characterisation; and the formulation of a hypothesis and generation of the corresponding predictions. For the establishment and characterisation of a cohort, many epidemiological studies use paper based data collection systems. A computer-based data collection system which enables geographically distributed data collection would alleviate much of the labour, tedium, and error that are inherent in paper-based data collection. Such a system is required to store personal, confidential medical data, and this data must be sent across a network from remote data entry clients to a data repository server. The immediate goals in developing the system were to ensure high quality data; ensure privacy and confidentiality of sensitive patient data; enable data collection to be performed at any location; ensure that the system was scaleable to cover larger populations and that it was highly available. In the long term the system needs to be enhanced and extended to provide a platform for the epidemiological study of schizophrenia and intervention research for predicting and preventing adverse outcomes.

We selected Web services as the underlying middleware because they offer firewall traversal, interoperability, modularity, and loose coupling which provides the ability to incrementally extend the capabilities of the system as it grows.

The health informatics findings from PsyGrid will be generalisable well beyond its disease focus. By treating every new patient as a participant in a longitudinal study, it will start to test a new model of care and research combined. We believe that this combination is essential to providing a timely and more flexible evidence base for future healthcare. This future could be called "high resolution healthcare"; it would encompass "personalised medicine", self-care decision-support, efficient and opportunistic clinical trials, complex (including genomic) epidemiology, and tactical development of local services based on local environmental factors and outcomes at the population level. High-resolution care and research requires information systems to link relevant data, methods and people in a clear and timely fashion.

In the following section we examine the clinical need for PsyGrid, and in then we analyse the requirements that flow from this. Afterwards, we provide a description of the NHS Information Technology infrastructure into which PsyGrid will be deployed. The rationale for using Web services is then presented, followed by an overview of the PsyGrid system architecture and the interactions between the Web services. We report our experience of implementing Web services, and provide a detailed description of the systems functionality and implementation. Thereafter, we capture our experiences of deploying and operating the system, and finally, we cover related work, future work, and a discussion of our findings.

CLINICAL OVERVIEW OF PSYGRID

Mental disorders are a large public health burden, whose treatment accounts for about 22% of the UK National Health Service (NHS) budget (Davies & Drummond, 1994). The major disorders are schizophrenia and bipolar disorder, each with a lifetime prevalence of about 0.5%. Both are major public health challenges, with costs of health and social care in the UK of over £2 billion annually, a similar order of magnitude to cancer or ischaemic heart disease. Schizophrenia usually starts in early adult life and leads to persistent disability in most cases. It arises out of a complex interaction of genetic and environmental risk factors.

The evidence base for the treatment of psychotic disorders is underdeveloped (NICE, 2003). Interventions can be divided into drug treatments, psychological treatments, and service-level interventions. Historically, randomised, controlled trials have been few in number and uneven in quality (Thornley & Adams, 1998). Rational service planning is constrained by the lack of knowledge currently about the comparative incidence rates and course of psychotic disorders in different regions of the UK. Models purporting to predict local service usage on the basis of demographic indicators are not well developed (Glover et al., 1998). For example, the extent to which an urban environment contributes to the risk of psychosis has only recently been understood and quantified (Pederson & Mortensen, 2001).

The focus of first episode psychosis (FEP) in PsyGrid arises out of a convergence of key questions in clinical science that require large, well characterised, representative cohorts to address them, along with recent developments in the NHS and NHS R&D to facilitate such research. Delayed detection and treatment is a widespread problem and predicts poor clinical outcome (Marshall et al.,2005).

Schizophrenia typically first occurs in one to two people in every 10,000 of a population each year. However, about two-thirds of episodes of schizophrenia start before 35; this proportion is higher in men, who more commonly suffer from the disorder anyway. PsyGrid subjects are drawn from eight different geographic regions of England that together cover one-sixth of the population. This coverage could be expected to yield up to 1,000 patients per year who would meet the eligibility criteria for the project.

REQUIREMENTS

In this section we discuss the key requirements of the system that shaped the system architecture.

- **Privacy and Confidentiality:** The PsyGrid data repository stores sensitive clinical data describing a patient's mental illness. Ensuring a patient's privacy and the confidentiality of their data was of prime importance. A fine-grained access control system was required which would only permit users with the necessary privileges to access and operate on the data in the repository. The ethical approval granted to PsyGrid further restricted the data we could store, mandating that the data stored be anonymous. However, removing certain identifying data items completely, such as post code and date of birth, restricts the hypotheses that can be tested and so these data items needed to be transformed such that they are still useful, but are not identifying.
- **Remote Data Entry:** The goal of the project is to study the occurrence and outcome of schizophrenia, beginning from a patient's first episode, over the course of the next 12 months. Subjects were to be drawn initially from eight geographically dispersed locales

covering in total one-sixth of the population of England, thus providing a highly representative sample. Consequently, a client-server architecture was required, where data collection clients provide the user interface for data entry, and the server hosts a centralised data repository, which persists the assessment data. In line with the privacy and confidentiality requirements, all communications between client and server had to be secure against interception attacks.

- **Paperless Data Collection:** PsyGrid was required to be a paperless data collection system, where the data is entered directly into the system at the time of collection. The implication of this requirement was that the system must be as natural and easy to use during a client interview as a pen and paper form. Since data capture is happening in real time it also implies that the system is *always* available for data entry. Due to the nature of the illness, if a schizophrenic is willing and able to be interviewed, then the opportunity should not be lost. The absence of a paper record from which the data could be recovered if necessary necessitates a back up system that can recover data up to the last committed transaction.
- **Off-line Data Entry:** The clinical researchers needed the capability to interview clients in any location, such as the hospital for the initial assessment or the client's home for any of the follow-up assessments. Consequently, a connection to the network cannot be guaranteed to be present and so the data entry system must be capable of working off-line on a portable computer. When a connection to the network becomes available, then the data entered whilst off-line will be uploaded. This requirement places a lot of complexity into the data entry client, and so a rich client application was preferable to Web-browser based solution.
- **High availability:** This requirement was implied by paperless data collection. The central data repository must be hosted on a high availability hardware platform, with data storage redundancy, to ensure continuous operation. The high availability architecture also must have the capability of performing a live upgrade to the system software and to modify a data set definition.
- **Scaleable:** The initial sizing of the system was for eight remote locations with 25 users using a single data set, which it was anticipated would contain approximately 1,500 subjects. However, we worked on the assumption that the initial deployed system would be require to support 32 remote locations with 100 users and four different data sets.
- **Data Quality:** Data quality (or the lack of it) is a major problem for all evidence based medical research. Electronic data entry systems can restrict the range of valid values and enforce them in a way that is not possible with pen and paper. Wherever possible we needed to restrict the ranges of data elements and provide an intuitive user interface for our end users.
- **Extension:** The data collection system is the first phase of the PsyGrid system. One vision for PsyGrid is to provide an epidemiology work-bench which will provide the capability to integrate the PsyGrid data set with other data sets, such as socio-economic data, which will enable the aetiology of schizophrenia to be studied in the population at large, which can be directly used in service planning for the treatment of the disease. The myGrid (Goble et al., 2003) toolkit provides this capability for the bioinformatics domain and could easily be adapted to epidemiology. This improved understanding of the lifestyle and environmental causes of schizophrenia, when combined with genetic information,

could be used to predict an individual's risk of developing schizophrenia, and so early intervention treatment could be effectively targeted.

- **Remote Management:** The PsyGrid system was required to support a geographically dispersed user base, with minimal technical support available on site. Therefore the goal was to build a system that could be managed remotely with no ongoing on site support required. All software upgrades needed to be performed remotely and ideally automatically. This requirement suggested to us that a browser-based Web application, which could be managed centrally, would be the easiest solution, but this solution would not support off-line data collection.
- **User Defined and User Managed:** PsyGrid was required to be simple to use and simple to extend. PsyGrid needed to support multiple independent data collection projects that could be designed and managed by their users. To simplify the creation of a new data set, the user interface for entering data should not need to be defined as well, as the data collection client application must be able to render this solely from the data set definition. The user interfaces for the management tools needed to be intuitive and easy to use.
- **Security Credential Management:** Based on our end user community we ruled out the possibility of using file system based Public Key Infrastructure (PKI). It is unrealistic to expect clinical users to deal with the complexity of managing certificates and keys themselves, and it would be a deterrent to the use of the system, which was not our aim. However, we did not want to preclude the use of a hardware token-based PKI approach in the future, and so the system needed to be designed with a PKI security infrastructure that would enable this to happen in the future. Consequently, the system needed to provide the ability to translate a user name and password credential into a PKI credential.
- **Federation of Systems:** Over time we envisage that multiple PsyGrid data repositories could be deployed. By supporting federation of these systems, much larger data sets could be used in analysis. In this federated scenario, there would be multiple instances of PsyGrid, where each one is operated by a different autonomous organisation and has its own user directory and security policy, which can be used collaboratively, such that a user from one PsyGrid in the federation can access another system without repeated authentication.

THE OPERATING ENVIRONMENT

The National Health Service (NHS) in England and Wales provides publicly available healthcare for the whole population. The NHS comprises of largely autonomous Trusts, which are responsible for procuring, deploying and operating their own IT infrastructure and applications. Consequently, no two Trusts are the same, and a large variety of infrastructure exists. To address the problems inherent in the NHS IT infrastructure, and to enable the deployment of NHS wide applications, such as the Electronic Care Record, the UK government embarked on the £6 billion "Connecting for Health" IT infrastructure project (formally the National Programme for IT). The aim of Connecting for Health is to modernise the NHS IT infrastructure and provide a set of global applications that will enable mobility of patients and their clinical data between Trusts. The first phase of the project was to deploy a new NHS-wide network infrastructure, which would provide high-speed links (>100Mb/s) between all Trusts via the backbone network. This infrastructure is commonly known as N3. The backbone would then be able to host applications global to the

NHS, which are known as spine services. These services include the National Care Record, the Secondary Usage Service and Picture Archiving and Communication Services, and will be rolled out in subsequent phases. The current state of the NHS IT infrastructure can best be described as transitional.

During the design of the PsyGrid data collection system we made the following assumptions about the NHS network infrastructure. Firstly, we assumed that the N3 high bandwidth infrastructure would be ubiquitous; secondly that each Trust would allow direct HTTP/HTTPS outbound connections to be made to any public NHS server in any trust on any port; and thirdly that inbound HTTP and HTTPS connections would be possible on ports 80, 443, and 8443 from any source address on the NHS network.

Motivation for Using Web Services

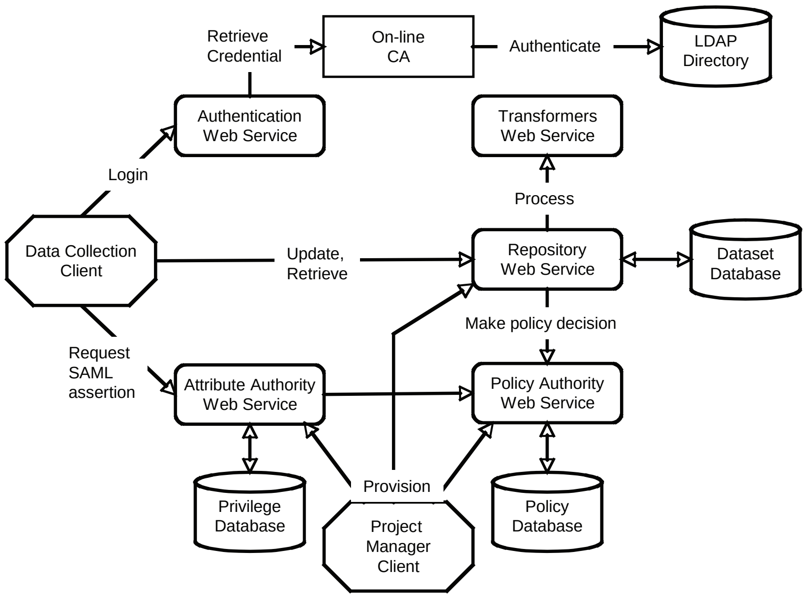
Service-oriented architectures have been successfully applied in the bioinformatics domain as typified by the myGrid (Goble et al., 2003) project. One of the motivations for the PsyGrid project was to provide the same ability to integrate and analyse multiple data sources using the same workflow approach based on Web services (Oinn et al., 2004). This was the primary factor in our decision to pursue a service-oriented architecture. By exposing our data repository this way, it enables the data to be analysed as part of a workflow that tests an epidemiological hypothesis. The need for a service-oriented architecture leads directly to Web services as the implementation technology. Perhaps the most compelling advantage Web services have over other technologies is the ability to pass through firewalls, which the analysis of the operating environment (presented previously) identified as particularly important. The availability of open-source implementations of Web service containers and protocol stacks also played a role, as PsyGrid was to be made freely available to the community, and it should be possible for other groups to deploy a PsyGrid system without purchasing any software. The existence of open Web service standards was also a factor in the decision, which would enable the PsyGrid system to be used from disparate implementations and operating systems, removing some of the barriers that are commonly found in heterogeneous environments such as the NHS. This is critical to fostering adoption of PsyGrid, as we do not have to mandate one particular platform or implementation. Finally, Web services bring with them the benefits of loose coupling. The system would be deployed and operational for data collection early in the project, long before much of the functionality for data integration and workflow-based statistical analysis could be completed. As additional functionality became available we needed the capability to add this into the operating system incrementally with no disruption to the deployed system.

However, we did have concerns about using Web services in the PsyGrid project. We were not aware of the use of Web services in the NHS and so we would not be able to benefit from the experience of others. Also, many of the Web services standards are new and potentially immature. This would be true of the open source implementation of these standards. Finally Web services have a reputation for being very slow in comparison to other technologies offering similar functionality. There was concern that a system based on Web services would not scale.

SYSTEM ARCHITECTURE

The data collection system has four main architectural components. The data repository stores the datasets, the data collection client provides the user interface for data entry, the security system provides authentication and authorization capability, and the project manager client provides the tools to setup and manage data collection projects. The data repository has been implemented as two

Figure 1. PsyGrid system architecture showing principal interactions between clients, Web services and databases.



Web services, and the security system is composed of three Web services. The data collection client and the project manager are clients of the Web services. The project manager client application is still under development and we do not describe it further. Figure 1 shows the architecture of the system and the principal interactions between the components.

Data Repository

The two Web services, which together form the data repository, are the Repository service and the Transformers service. The Repository service provides a Web service interface to the backend database, which stores the data set definitions and the collected patient data. The operations on the Web service can be grouped into two. The first set of operations enables the remote management of data set definitions such that they can be installed and updated. These operations are used by the Project Manager client application. The other set are used by the Data Collection client application and allow patient data to be added, updated, and retrieved. The implementations of these operations allow for client applications that must operate off-line. For instance, a client is not required to download an existing record when new documents are to be completed; the repository manages appending the new document instances to the existing record during the save process. The Repository service enforces access control on its exposed operations to ensure data privacy and confidentiality. The Transformers service provides a set of operations that enable data to be anonymised. For example, a SHA-1 operation is provided to return the hash value of the input. For any data value contained within a data set, a transformer can be specified. The input data type and the output data type are specified along with the URL of the Transformer Web service and the operation to be invoked. The Transformers Web service effectively provides a configurable data pre-processor, which is primarily used for de-identification of identifiable data between it being entered by a user and stored in the database.

Security System

The three services implemented for the security system are the Authentication service, the Attribute Authority and the Policy Authority. Together they provide a Role Based Access Control (RBAC) (Sandhu et al., 1996) system for PsyGrid. In RBAC, users are assigned privileges, and authorisation decisions are based upon the possession of the required privilege. There are three components in RBAC. The first is a privilege manager, which maps a user to their privileges. The second is a policy decision point (PDP) that is used to control access to a resource. The third component is a Policy Decision Function (PDF), which is used to make a decision on whether a user has sufficient privileges to access a resource.

The Authentication service provides a port type for the user to login to the system. It is essentially a Web service wrapper around an on-line certificate authority, which itself authenticates users against an LDAP directory. If authentication is successful, then the on-line certificate authority will issue an X.509 PKI credential, bound to the user's identity. This credential is then used to authenticate with the other Web services in the system which require mutual authentication over SSL.

In PsyGrid, the Attribute Authority (AA) provides the privilege management function. It stores the list of projects that are active on the system. For each project it records its name, a unique identifier, a list of the sub-groups of the project, and the roles that users can take on in this project. It also maintains a registry of users and their privileges. The user's privileges are maintained on a per project basis. For each project the user is a member of, the privileges granted to the user (role or group membership) are listed. The AA issues Security Assertion Mark-up Language (SAML) tokens that bind a user's identity to their privileges in a project. The AA digitally signs these statements, which guarantees their authenticity. Any entity that trusts the AA can accept its assertion about a user's privileges.

The Policy Authority (PA) maintains the security policy. It stores multiple policies, such that each data collection project can have a unique policy. A policy consists of statements, and each statement has an *action*, *target*, and a *rule*. The rule is a Boolean logic expression composed of operators (AND, OR, NOT) and privileges. A rule may be composed of many sub-expressions. The *action* is the operation the user wishes to perform, and the *target* is the resource on which they want to perform it. In the following example, in order to invoke the "getRecordSummary" method (the "action") on the data repository to retrieve the records owned by the North West hub (the "target"), then you must either be a Clinical Researcher belonging to the North West hub or be the Clinical Project Manager (the "rule"):

```
ACTION="ACTION_DR_GET_RECORD_SUMMARY", TARGET="NORTH_WEST_HUB",
RULE={{ROLE=CLINICAL_RESEARCHER AND GROUP=NORTH_WEST_HUB} OR
{ROLE=CLINICAL_PROJECT_MANAGER}}
```

The current policy for the First Episode Psychosis study has approximately 600 statements, covering some 30 actions and 10 targets.

The access enforcement function (AEF) provides the policy decision function. It is a client side API for the PA, which can be invoked from any Web service that protects a resource. The AEF requires the caller to supply the target and action, and either the users identity, or a signed SAML assertion that can be verified. The PA, AA and Repository all use the AEF to protect the operations that are exposed as Web services.

The on-line certificate authority is used to issue short-lived user credentials. However, an alternative source of authority, the off-line PsyGrid Infrastructure Certificate Authority (CA), is used to issue long-lived credentials. This gives us two levels of trust, determined by the certificate authority that issued the end entity's credentials. Those in possession of a credential from the PsyGrid Infrastructure CA, typically servers, are able to invoke services on behalf of other users. In this case, the identity presented during authentication and the identity of the subject in the SAML assertion need not be the same. This is known as delegation.

By using signed SAML assertions to identify a user's roles, and using a role-based access control system, then to federate multiple data collection systems only requires the policy decision function to accept SAML assertions signed by the other attribute authorities participating in the federation. In the current implementation, the PA is configured with a list of AA's that it trusts. This federated model is similar in design to the Shibboleth (Shibboleth, 2006) system but has been implemented for Web services, where as Shibboleth is used to protect Web pages.

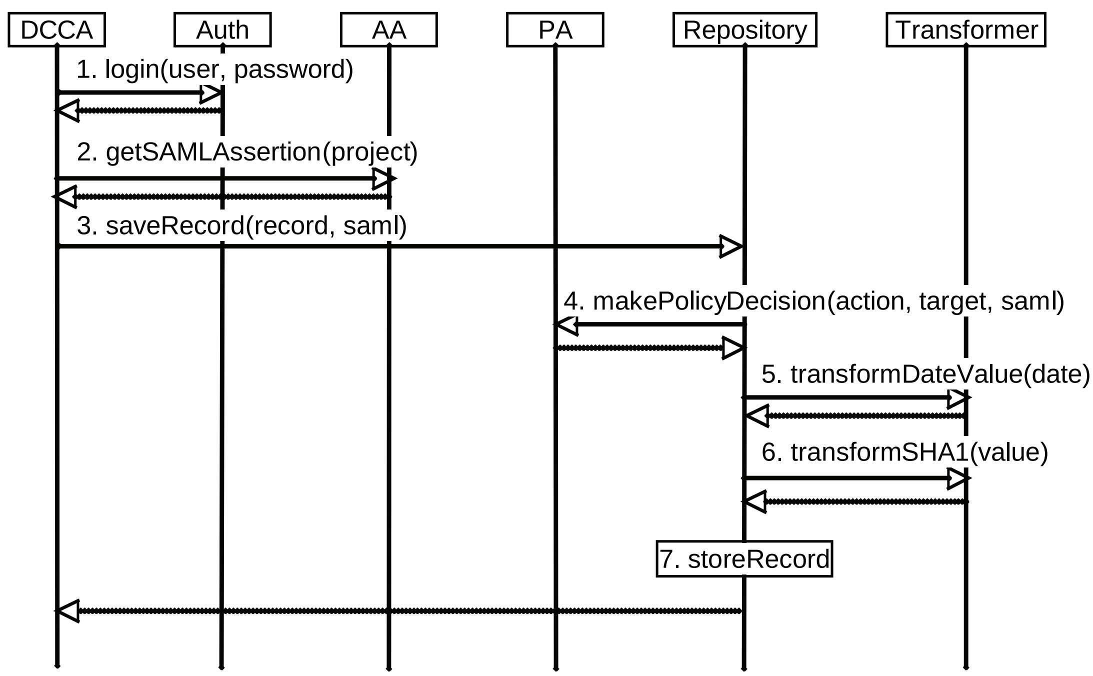
Data Collection Client Application (DCCA)

The PsyGrid data collection client application provides the functionality required for the entry, secure storage and secure transmission of data collected from longitudinal studies. To achieve this, it interacts with the data repository and the security subsystem. A Web-services interface is used by the data client to communicate with the other systems, including the data repository, attribute authority, policy authority, and certificate authority. All communication is encrypted through the use of SSL.

Inter Service Communication

To ensure confidentiality and privacy, all components in the PsyGrid system communicate over secure, encrypted communications links.

Figure 2. Interaction between services when the user logs in and saves a document



We selected transport layer security (TLS) in conjunction with HTTP as the transport for our SOAP-based Web services. We use TLS in mutual authentication mode to ensure both server and client can be sure of the identity of the other party. TLS was chosen over message level security (MLS), as specified in WS-Security, because it is widely available and implementations are mature.

The basic interaction between the client, the data repository and the security system can be illustrated by the example of a user logging into the system and saving a document, as shown in the interaction sequence in Figure 2. Prior to this the user has been working off-line and completed a patient assessment form, which they now wish to save.

1. The user launches the data collection client application and logs in with their user name and password. The DCCA passes the login information to the Authentication service, which attempts to authenticate with the credential repository. The credential repository in turn attempts to authenticate the user with the directory; if successful, it uses its on-line certificate authority to generate a temporary PKI credential for the user, which contains the user's Distinguished Name as the principal.
2. The DCCA next contacts the attribute authority, which issues a signed SAML assertion for the user, based on the Distinguished Name in the credential supplied during mutual authentication between DCCA and the AA, containing their privileges for the requested data collection project.
3. The DCCA then invokes the `saveRecord()`operation on the repository passing the Record which contains the document(s) to be saved and the signed SAML assertion.
4. The Web service uses the AEF to invoke the Policy Authority's `makePolicyDecision()` operation, including the *action* (saveRecord), *target* (determined from the owning group of the record) and SAML assertion. Based on the users privileges listed in the SAML assertion and the stored policy, the Policy Authority will either grant or deny access.
5. The document to be saved has a date entry marked as requiring date transformation. The date is to be converted from day-month-year to a month-year, and the appropriate transformer is invoked.
6. The document has another entry that is marked as requiring encryption. The SHA-1 encryption transformer is invoked.
7. The document is store in the Repository's database, and the unique identifier of the persisted object is returned as an indication of success.

Once the user has logged into the system then temporary credentials can be used until they expire. The DCCA will automatically refresh the end user's temporary credential before it becomes invalid by accessing the `login()` operation of the authentication service, using user name and password which were cached by the client when the user logged in. This is done without further interaction from the user. Thus the infrastructure is secured by PKI, and yet there is no burden on the user of managing any more than a user name and password. It would not require any changes to the backend system to integrate a hardware token approach; only the DCCA would be affected.

Interface Description

Each of the Web service exposes a number of operations, and the parameters required range in complexity from simple integers or strings, to complex documents that describe an entire data set. For the Repository, Authentication and Transformer Web services, the interface description in WSDL is generated from the Java implementation of the interface. For the PA and

Table 1. Examples of operations exposed by the PsyGrid Web services

Web Service	Operation	Input Parameters	Output	Description
Transformer	encrypt	Value:string	String	Returns a SHA-1 hash of the input string
Authentication	login	userId:String password:char[]	String	Authenticates the user and returns a PKI credential store in a Java KeyStore that is base-64 encoded and returned as a string.
Repository	getRecordsByGroups	project:String groups:String[] saml:String	Record[]	Returns an array of Java Record objects (as defined by the Repository data model), which exists in the specified project and groups within the project.
Policy Authority	makePolicyDecision	project:ProjectType action:ActionType target:TargetType saml:String	boolean	Determines whether the user's privileges are sufficient to perform the requested action on the specified target. The input is an XML document containing the project, action, target and SAML assertion.

AA services, the Java stubs are generated from the WSDL definition, and the messages are documents defined by an XML schema. The number and range of operations is too great to describe each one in detail, but selected examples are presented in Table 1.

WEB SERVICE IMPLEMENTATION EXPERIENCES

In this section we discuss three detailed aspects of implementing our Web services together with the lessons learned.

Using WS-Security

Our original intention was to use the WSS4J implementation of the WS-Security standard for retrieving, transporting and verifying the signed SAML assertions used in the access control system. Whilst the transport and verification functionality met our needs, retrieval from the Attribute Authority required that the data collection project identifier be included in the request. There was no easy way to communicate this from the client application through the Axis Web service stack to the WSS4J handler that implemented the SAML retrieval function. The only way we found to do this was to set the information on the Axis `Call` object, but this required that the Web service client API stubs were extended to pass this information through to Axis, which in turn meant that they could not be generated automatically from the WSDL definition of the service. Consequently, we moved the SAML functionality into the application layer and added the SAML assertion as a parameter to each operation of the repository Web service that required access control.

Document vs. RPC Encoding

Within PsyGrid we have used both styles of encoding for our services. The Authentication and the Repository services use RPC encoding, and are generated using Java2WSDL, directly from the Java class definitions. Conversely the Policy Authority and the Attribute Authority use document-literal encoding. An XML schema defines the content of the documents, which is used in the WSDL definitions of the services. WSDL2Java was then used to generate the Java classes for the implementation. The Repository object model was implemented making full use of Java's complex types for managing collections. Java2WSDL was unable to map these complex types into the corresponding XML representation and so a Data Transfer Object layer was implemented which mapped the complex collections types into simple arrays. We then used the DTO classes to generate the WSDL for the Repository. In a similar way we also implemented a DTO layer for the Policy Authority and Attribute Authority. The WSDL2Java implementation was unable to handle abstract types in the XSD definition, however they were required for the natural implementation of the object models of these two services. Consequently the XSD was restructured to eliminate the use of abstract types and the DTO layer was implemented to restore this in the Policy Authority and Attribute Authority services.

SSL Implementation

By default Axis provides a default SSL socket factory for the management of SSL connections, but allows a custom socket factory to be configured. The default factory was unsuitable for our needs as the DCCA had to refresh the PKI credential each time it expires and so we implemented a PsyGrid specific socket factory. The PsyGridClientSocketFactory Java class provides a specialised key manager and a certificate manager that is capable of updating the Java SSL subsystem with the new PKI credentials.

Logging

All Web services implement logging of each attempt to access the operations they provide. This provides a full audit trail of the usage of the system. Each log entry records the entity invoking the operation (from the X.509 certificate), the user requesting access (from the SAML assertion), the source IP address from which the request originated, a time stamp, and the name of the operation being invoked. If a request to access a Web service is denied, this is also logged. A common logging API was implemented for use by all Web services, which uses the Apache Log4J framework.

IMPLEMENTATION DETAILS

We describe the implementation of three of the four core components of the data collection system architecture. The project manager client application has not been implemented at the time of writing.

Data Repository

The PsyGrid data repository has two principal objectives; to store the definition and structure of the data to be stored in the repository, and to store the data itself. It also performs a number of ancillary functions, which will be described below.

To define the definition of the data to be collected a three-level hierarchy has been created. At the top-level of the hierarchy a single data repository may define a number of datasets, each dataset being equivalent to a single data collection project. At the next level of the hierarchy, within each dataset may be defined a number of documents, each document being equivalent to a single paper-based assessment form. The repository also caters for documents that are to be completed multiple times within a single study (for instance for longitudinal studies), and allows documents to be grouped into those that are intended to be completed together.

The final level of the hierarchy allows each document to contain a number of entries, with each entry being equivalent to a single item of data to be collected. Provision has been made for collecting a variety of different types of data, including, but not restricted to, numeric data, textual data, and selection from a list of options. The entries in a document may be logically grouped into sections, and it is possible to define sections that are to be completed multiple times, but with a different context each time. This simplifies the definition of documents with repetitive entries.

More advanced behaviour for entries is also supported. Each entry may have one or more validation rules defined for it to prevent illegal data from being entered. Also, an entry type is provided for which data is not entered directly, but the value is calculated by performing a calculation involving the values entered for other entries in the document; this can be used to calculate the overall value of an assessment scale involving multiple questions for example. Finally, for entries where the value is selected from a list of options it is possible to define the users flow through a document in response to the option they select, by enabling or disabling subsequent entries in the document.

For storing the data collected a similar three-level hierarchy is used, to match the hierarchy of the data definition. At the top level of the hierarchy a record represents a single instance of a dataset. At the bottom level of the hierarchy is a response to a single entry; each response maintains a collection of values which represent an add-only store of all entered data for the response, with accompanying provenance metadata.

The PsyGrid data repository is implemented as a set of persistable Java classes, mapped to a relational database using the open-source Hibernate object-relational mapping package. As well as reducing development time by eliminating the need to hand-craft SQL this approach also allows the PsyGrid data repository to be deployed using any of the back-end databases supported by Hibernate, which includes MySQL (used during development), IBM DB2 (used for the production deployment) and the majority of other RDBMS providers. The data repository also provides scheduling functionality whereby for a dataset representing a longitudinal study it can be configured to send e-mail reminders to the responsible persons when the next set of documents are ready to be completed for each record.

Data Collection Client

The PsyGrid data client is a rich-client application that automatically generates a visual representation of a data set from a definition provided by the data repository. It represents the data set by using a combination of simple user interface elements such as radio buttons, combo boxes and text fields, as well as more complex ones such as tables. The data client also has a specialised date widget that allows the user to select a date from a calendar-like interface. Finally, support for grouping is available in the form of sections.

The application presents a clean and simple interface using a wizard-like approach providing the user with a unique path to follow in order to complete an assessment. This approach was taken considering that users of the system are unlikely to have a technical background. In addition, context-sensitive help is available throughout the application. The data set definition is the source of the information used by the context-sensitive help, providing maximum flexibility. In keeping with making the usage of the application as intuitive as possible, the application is able to fully restore its state after being terminated. So, if for some reason an assessment had to be interrupted, the user could simply close the application without having to worry about losing data.

The accuracy of the data entered is of great importance, and humans entering data are likely to occasionally make mistakes. The data client attempts to mitigate this problem by providing instant ("as you type") feedback on the validity of the data inserted. If the data entered is not valid, an icon is displayed next to the entry to indicate it. An explanation is also available and it is shown as a tool tip when the mouse cursor hovers over the icon. Validation rules to govern this process can be defined per data set and can be applied to individual entries. A common source of errors is the incorrect usage of measurement units. In order to solve this issue, the client application displays the units defined for an entry in a combo box, allowing the user to choose from one of them. Finally, entries whose values are calculated from values in other entries do not require human intervention at all. The calculated score is updated in real-time as the values are entered in the relevant entries. Additional features provided by the client are a "review mode", and specific support for unanswered questions. The former allows a user to examine a completed assessment in order to correct any mistakes. Any correction requires an annotation describing the reason for the change. An audit trail is retained making it possible for a change to be reverted if necessary. Unanswered questions are common when collecting data, so special support is present to allow the user to select the reason for not answering the question. The set of options available is defined per repository. This information can be useful in the analysis stage and when revising the contents of an assessment.

The client has two modes of operation, "off-line" and "on-line". The former provides a subset of the functionality available in the latter mode. The ability to save records, retrieve data sets, and to review already submitted data are only available in the on-line mode. While in the off-line mode, the information collected is stored in an encrypted form and as soon as the application enters the on-line mode, the data is uploaded to the data repository and deleted from local storage.

Security

The Authentication service provides a Web service interface to the myProxy credential repository (Basney et al., 2005) that has been configured to act as an on-line certificate authority (CA). The Attribute Authority is implemented as a set of persistable Java classes, mapped to a relational database. The definition of a project includes the project name, a list of sub-groups which exist for this project, and a list of the valid roles which a user can be assigned. The AA knows each user by their Distinguished Name, and the AA stores the list of projects and privileges for the user. The AA provides a Web service interface, which exposes operations to query project information and user privileges. The other major function of the AA is the issuing of signed SAML assertions, which binds a users identity to their privileges. We have used the OpenSAML implementation for this. The Policy Authority makes access control decisions, based on the stored policy and the user privileges supplied in the SAML assertion. Policies are implemented as sets of persistable Java objects, and any number of policies may be stored; typically there will be one policy for each data collection project. The PA verifies the validity and integrity of a SAML assertion and confirms it comes from a trusted AA. It then checks the policy to determine if the supplied privileges are sufficient for the request target and action. In this context a target corresponds to an object or object group in the data repository and the action is the operation to be performed.

Client side APIs have been developed for both the PA and AA to hide the details of accessing the Web service. This means the services using the access control functionality need only call the makePolicyDecision() function provide by the API

DEPLOYMENT AND OPERATION

Third Party Software Components

All software components developed as part of PsyGrid are open source licensed under the GPL. We rely on a number of 3rd party software components, and to ensure that anyone can deploy their own PsyGrid, we have used only free, open source components. Most notably we have used Apache Tomcat (version 5.5.12) as our Web service container and Apache Axis (version 1.3) as our Web service stack. OpenLDAP was used for the directory, which provides user authentication, combined with myProxy on-line credential repository. Our backend database in the production system is IBM DB2, although in our development environment we use mySQL. Hibernate (version 3.1) provides an object relation mapping service (ORM) which enables any database supported by Hibernate to be used.

Hardware Architecture

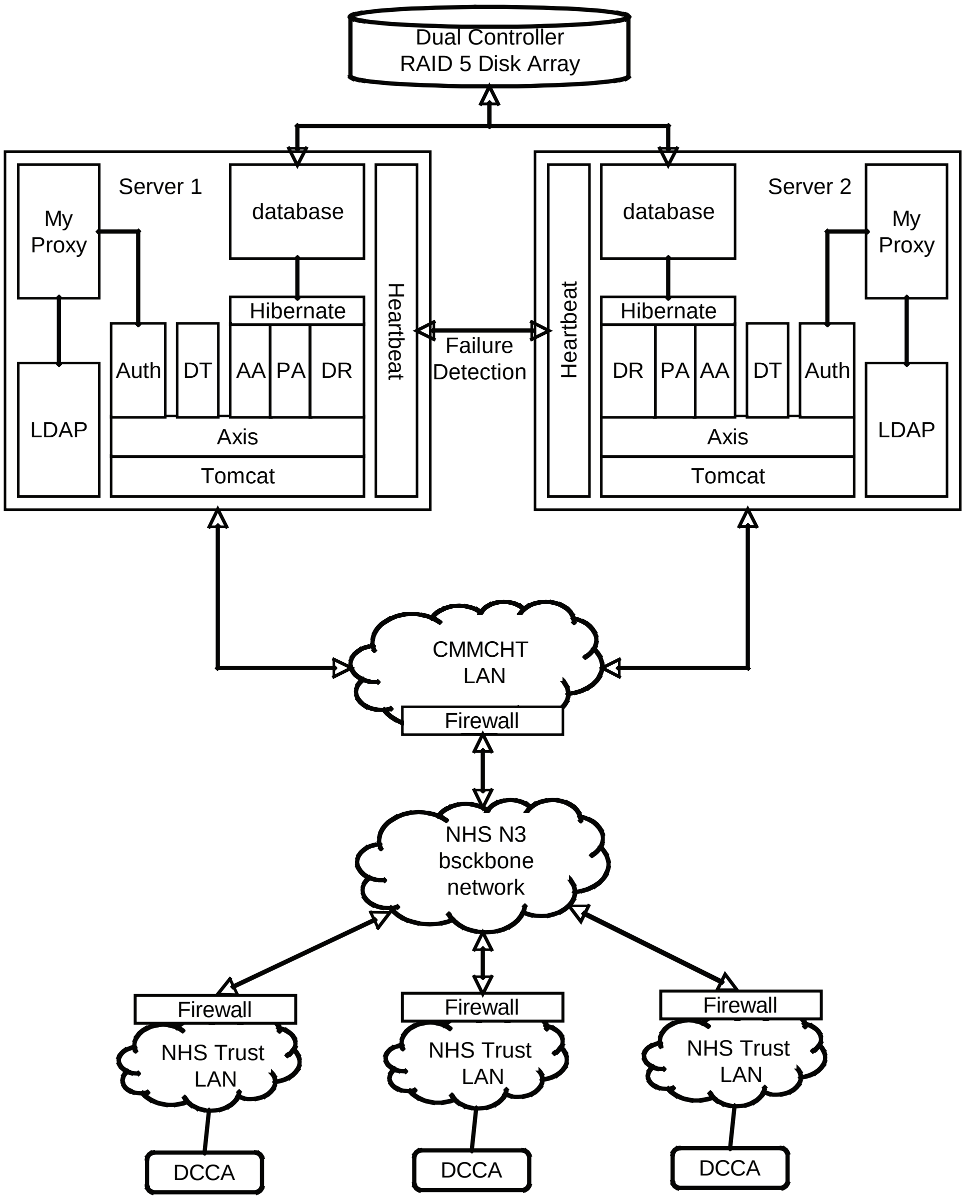
PsyGrid is a production system and so it was designed to be continuously available. The deployment configuration is shown in Figure 3. High availability was achieved by using a pair of servers to host the Web services and database. These two servers are deployed with an identical software stack, which consists of the Web service container (Apache Tomcat), the LDAP directory (OpenLDAP), the database, which is IBM DB2 in the production system, and the myProxy on line certificate authority. The disk used for storage of the data is a RAID 5 array. The servers are identical in every respect, and use the Linux Heartbeat application to manage high availability, using a shared "virtual" IP address. At any one time one server is active and the other is in standby. The active server is assigned the virtual IP address by Heartbeat. The URL of the Web services, that the client used to access the system, resolves in the NHS network DNS to this virtual IP address. The active server has the RAID 5 disk mounted, which stores the database. Heartbeat manages failure detection, and can switch activity for the active server to the standby under a number of conditions. Heartbeat will determine that the active server has failed if the active server does not respond to the heartbeat message sent by the standby server via both IP and a direct serial cable connection. It will also determine the active server to have failed if the Web services do not respond to a getVersion() request, which would indicate a Tomcat failure. The active server continually monitors the health of the DB2 and Tomcat processes. A switch of activity proceeds as follows:

1. The standby node forces the other node to shutdown, releasing the virtual IP address and unmounting the database disk.
2. The standby node mounts the database disk.
3. Tomcat is started and the Web services initialised. myProxy and the LDAP server are always running.
4. The virtual IP address is taken over by the standby server, which is now the active server.

Since all persistent data for the AA, PA and Repository services is stored on the shared disk, there is no need for replication of their data. The LDAP directory uses the openLDAP replication service to keep the two directories synchronised, and myProxy maintains no state data. A switch of activity has been measured to take 12 seconds on the production system, which causes minimal interruption to service.

There was no opportunity to perform scaling tests on the system before it had to be placed

Figure 3. PsyGrid Data Collection System deployed for FEP data collection (DCCA = Data Collection Client Application, AA=Atribute Authority, PA=Policy Authority, DT=Data Transformer, DR=Data Repository, Auth=Authentication Service)



into production due to time constraints, though for the initial deployment we were confident that our hardware platform would be more than powerful enough. We have long since passed the initial sizing requirements and we now support three independent data collection projects, with a combined user base of fifty distributed across 16 sites, and there is no degradation in performance. As and when performance becomes an issue, the use of Web services will enable us to seamlessly add new hardware and redistribute the services among the available hardware.

Software Upgrade

Our requirement was that a software upgrade should not cause an interruption to service. There are two different types of modification that need to be made to the running system.

The first of these is an upgrade to the Web services that occurs when bugs are fixed or new functionality has been added. The redundant hardware architecture employed enables the upgrade to be completed with minimal service interruption. The process we used is as follow:

1. Suspend the heartbeat application so that fail-over cannot occur.
2. On the standby machine remove the currently deployed Web services that are to be upgraded
3. Deploy the new versions of the Web services.
4. Restart heartbeat and manually switch activity. The upgraded Web services now become active.
5. Repeat steps 1, 2, and 3 so that the other server is upgraded.
6. Repeat step 4 to ensure that the deployment to the second server has been successful.

Changes to the interface of Web services require the data collection client to also be upgraded. The client is distributed using Java WebStart, and a new version is checked for each time it is launched by the user. If a newer version exists it is automatically updated. If the user is running the client during the upgrade, then any change to an existing Web service port type will cause problems. Consequently we maintain backwards compatibility between consecutive releases of a Web service. The addition of new port types to a Web service can be safely deployed without affecting running clients. To assist in debugging and user support, all Web services implement a `getVersion()` port type which return the version information set at compile time. We do not use Tomcat's Web service "hot deploy" capability as it proved to be unreliable, producing runtime errors. Instead the Web archive file and the deployment directory of the service are completely deleted.

The second type of modification we need to make to a running system is modification of the data sets. This is required when either a bug is found or a change request from the project clinicians is made. The repository Web service enables a live update to the data set to be made using a "patching" port type. This enables the complete object graph to be downloaded, modified by the patching client and then saved in the repository. Each data set records its own patch level.

The combination of Web services and the high availability hardware architecture enable us to upgrade and patch the system remotely and in a way that is largely transparent to our users.

Remote Connectivity

Deploying PsyGrid across the eight NHS Trusts proved to be very difficult. The NHS code of connection requires that only computers owned and managed by an NHS Trust may connect to the NHS network. Consequently, the purchase of hardware for the PsyGrid project and the configuration of those computers had to be devolved to the local NHS Trusts. We also required information from the Trusts about the source IP addresses the remote client computers would present to the firewall protecting the Central Manchester and Manchester Children's Hospital Trust (CMMCHT), which was the host for the central PsyGrid servers, so that this could be configured accordingly. The CMMCHT policy is to treat the NHS network link as hostile, and so connection was denied by default. The IP addresses were needed so that they could be added to the "white list" of trusted remote computers. Acquiring this information proved to be very difficult. The first hurdle that had to be overcome was to find the person who was able to release this information, and then to explain to them we were a legitimate research project, providing our supporting documentation. The response we often received to our first approach was one of suspicion, and that we were attempting to obtain restricted security information using social engineering techniques. The situation was further compounded by the fact the NHS IT departments are often overloaded with the work of keeping the IT systems used for the provision of clinical care running. Whilst we encountered no refusals to help support PsyGrid, research projects such as ours have a low priority and so help was provided on a best effort basis. We concluded that we needed to minimise as much as possible our dependence on NHS Trust IT staff as each dependency could turn into a project delay. We were fortunate in that we were able to negotiate with CMMCHT a relaxation in their firewall policy such that access to our servers would be possible from any computer on the NHS network. Without this relaxation it would be impossible to scale the deployment of PsyGrid across the NHS in the future.

Once the central system was deployed the next step was to test connectivity from the remote NHS Trusts. The initial assumptions we made about the NHS network infrastructure (in a previous section) were very quickly shown to be wrong. We discovered that there were three different classes of NHS Trusts differentiated by the way they allowed connections to be made to the NHS network. The first type of Trust caused no problems as they permitted direct connection to the NHS network with no restrictions on the ports that could be used. The second type of Trust employs HTTP proxy servers between the Trust's network and the NHS network. These proxy servers only permit HTTP traffic on port 80 and HTTPS traffic on port 443. The third type of Trust also employs proxy servers, but additionally requires the user to authenticate to the proxy server. All of the Trusts that require proxy authentication were not using the standard HTTP proxy server authentication method, but were using the proprietary Microsoft NTLM authentication protocol, which requires the user to supply their Windows domain credentials.

Whilst it may have been possible to negotiate with the individual Trust's to remove these restrictions so that a direct connection could be made in line with our initial assumptions, we had already concluded that relying on NHS Trusts for this was not viable or scaleable. Even if we could negotiate direct connection, there would be no guarantees that this would not be withdrawn in the future without out knowledge. Therefore we had to extend and refactor the system to take into account the reality of the different types of Trusts. The first task was to use only the standard ports for HTTP(S). We were using two ports for HTTPS traffic, 443 and 8443, with Tomcat configured to perform server side authentication and mutual authentication respectively. The data collection application Web start Web service and the authentication service were hosted on the server-authenticated connector and the other Web services required mutual authentication. We now only had one port we could use for HTTPS and one port for HTTP. The data collection application Web start Web service was moved to the HTTP connector on port 80. The authentication service could not be moved to an unecryptyed connection however, since the user name and password would be sent in the clear to the authentication service. However, the available HTTPS port had to employ mutual authentication, for the other Web services that use the distinguished name for the client certificate to make access control decisions (for Policy Authority and Attribute Authority) and log service invocation. We therefore had to place the authentication service on this connector and distribute with the data collection client a default certificate that would enable authentication to the server, so that login may proceed. The consequence of this is that the client application is distributed with a certificate with a long lifetime, which could be used to mutually authenticate with the PsyGrid servers. However, the actual access-control decision on all services requires either a valid signed SAML Assertion or a certificate issued by the PsyGrid Infrastructure CA. This default certificate has the distinguished name of "CN=nobody, O=user, O=psygrid, C=uk", and is issued by the PsyGrid Online CA and so can not be used to do anything more than complete the SSL mutual authentication. It cannot be used to retrieve a SAML Assertion (as there is no user with this Distinguished Name), nor are certificates from this CA accepted as valid for access without a SAML assertion.

The next task was to provide the ability to tunnel SSL through proxy servers. By default the version of Apache Axis (version 1.3) we were using did not provide support for SSL tunnelling through proxy servers. The existing PsyGridClientSocketFactory was extended to provide this behaviour and a configuration dialog was added to set the proxy server address and port number to the client application. If a proxy server were configured, then the PsyGridClientSocketFactory would setup an SSL tunnel through the proxy, using the HTTP Connect method. One negative consequence of this is that it requires the end user to perform this configuration, which often requires a call to PsyGrid technical support to guide then through this process. Finally we had to deal with authenticating proxies. Basic and digest authentication is a standard extension of the HTTP protocol, but the NTLM protocol is Microsoft proprietary, and the specification is not published. However, the protocol has been reverse engineered and the Apache Commons HTTP Client provides an implementation. It also implements basic and digest authentication methods. Axis 1.3 does not use the HTTP Client and so we had to again integrate this into the PsyGridClientSocketFactory, so that it would perform authentication if this was configured. We had to further extend the configuration dialog so that the authentication type could be selected, and a further user name and password dialog is displayed during login if proxy authentication is configured.

DISCUSSION

The operating environment in the NHS has undoubtedly provided the biggest challenge to PsyGrid. The high degree of heterogeneity between NHS Trusts, the use of propriety protocols and the firewall restrictions, which we only really began to understand when we tried to make the system operational, all caused the system to be reworked to overcome these restrictions. This is the single most important lesson learnt—a thorough understanding of the operating environment is required when the system is being design. However, these obstacles were not insurmountable and this shows that Web service implementations have reached a level of maturity that is able to cope with such a hostile environment. There is no doubt that a Web application based around off the shelf products would not have encountered these problems, but this would not have met our current or future requirements. We identified restrictions in the Web service tooling which defined a lowest common dominator as to the way we structured our data types whether the starting point was Java or XML. We anticipate our decision to use Web services will be fully justified in subsequent phases of the project as we begin to extend the system to include Web service orchestration and workflow techniques. Performance of the Web services, which we thought might cause problems, has not proved to be an issue to date, and we currently support three data collection projects encompassing 16 locations and 50 users.

The adoption of the PsyGrid system for mental health research in the UK has resulted in modification to the business processes of data collection in three ways. Firstly, data collection is now paperless, and data no longer needs to be transcribed from paper into an electronic format for analysis. This requires fewer resources, and the risk of introducing errors in the transcription is eliminated. Secondly, because PsyGrid can operate off-line, clinical researchers are able to enter data at any time in any location, including patient's homes, and consequently they have more flexibility in their working patterns. Thirdly, we believe that the system is easier to use, and the restrictions on data input imposed in the user interface will lead to higher data quality. An informal analysis indicates that this is likely to be true, and we will report on a thorough evaluation in the future.

Grid computing techniques and technology are spreading rapidly through the health informatics community. Web services underpin the majority of current Grid middleware and consequently the use of Web services is becoming wide spread. The caCORE project (Phillips et al., 2006) and the closely allied caBIG from the National Cancer Institute Centre for Bioinformatics, have developed a set of tools and applications that provide much of the functionality of PsyGrid for the cancer domain. We are evaluating whether the caCORE SDK can be used to implement the data set designer function in the Project Manager application. PsyGrid is closely allied with three sister projects, namely CancerGrid (Brenton et al., 2005), NeuroGrid (Geddes et al., 2006) and VOTES (Virtual Organisations for Trials and Epidemiological Studies) (Stell et al., 2006). All these projects have a common theme, which is to develop middleware that will enable distributed collaboration and resource in their specific domain. CancerGrid is focused on supporting clinical trials in cancer research. NeuroGrid is developing a medical imaging Grid, which provides a searchable, distributed file system for the curation brain images, and workflow tools for analysis. Finally VOTES is focused on the security aspects of dynamic virtual organisations in a clinical context, and their application to the integration and query of clinical data sets arising from routine care, which can be used to identify eligible clinical trial participants.

We have described the PsyGrid data collection system, which is being used to record the patient assessments in the longitudinal FEP cohort study. The next phase of the project will see the development of tools for epidemiology, which will enable the testing of hypotheses about the causes of schizophrenia, and identify environmental risk factors. This will require the core PsyGrid data set to be integrated with existing data sets such as census data from the Office of National Statistics and mental health service data from the Mental Health Minimum Data Set. We plan to expose these data sets using grid-computing techniques. The OGSA-DAI protocol (Chervenak et al., 2003) provides a uniform method of accessing data sources and OGSA-DQP (Alpdemir et al., 2004) provides the capability to perform distributed queries over multiple data sets. A range of Web services will be created to provide functionality for data cleaning and statistical analysis. To orchestrate the Web services developed for epidemiology we will use the myGrid e-science workbench that provides a creation and execution environment.

We will add a clinical trial manager component to the data collection system. This will be implemented as an additional Web service and will provided the ability to randomise treatments according to a user configurable algorithm.

In the long term we plan to further develop PsyGrid so that it can be used for predicting patients most at risk, and given their early symptoms from brain imaging and psychiatric assessments combined with genetic data, can be used to target the most effective treatments to prevent adverse outcomes. This will require integration with further data sources collected as part of clinical care and the development of a decision support system for clinicians. The longitudinal data collected during the FEP cohort study will provide the evidence base for making these predictions.

ACKNOWLEDGMENT

The PsyGrid project was funded by the UK Medical Research Council and the Department of Health under the core e-science program, grant number G0300610. We would like to thank Ismael Juma for implementing the Data Collection Client Application.

REFERENCES

Ainsworth, J.D., Harper, R.S., Juma, I., & Buchan, I. (2006) PsyGrid: Applying e-Science to epidemiology. Computer-Based Medical Systems, 2006. CBMS 2006. *19th IEEE International Symposium*. (pp. 727-732).

Alpdemir, M., Mukherjee, A., Gounaris, A., Paton, N., Watson, P., Fernandes, A., & et al. (2004). OGSA-DQP: A service for distributed querying on the grid. In *Proceedings of the Advances in Database Technology,* Crete, Greece (LNCS 2992, 858-861). Springer.

Atkinson, M., Chervenak, A., Kunszt, P., Narang, I., Paton, N., Pearson, D., & et al. (2003). Data access, integration and management. In I. Foster and C. Kesselman (Eds.) *The Grid 2: Blueprint for a new computing infrastructure*. Morgan Kaufman

Basney, J., Humphrey, M., & Welch, V. (2005) The MyProxy online credential repository. *Software: Practice and Experience 35*(9), 801-816.

Brenton, J., Caldas, C., Davies, J., Harris, S., & Maccallum, P. (2005) CancerGrid: Developing open standards for clinical cancer informatics. In *Proceedings of the UK e-Science All Hands Meeting 2005*,

Cocharne, A. (1972) *Effectiveness and efficiency: Random reflections on health services*. Re-published 1999 Royal Society Medicine Press, London.

Davies, L.M., & Drummond, M.F. (1994) Economics and schizophrenia: The real cost. *British Journal of Psychiatry, 165*(supp. 25), 18-21.

Geddes, J., Mackay, C., Lloyd, S., Simpson, A., Power, D., Russell, D., & et al. (2006) The challenges of developing a collaborative data and compute grid for neurosciences. Computer-Based Medical Systems, 2006. CBMS 2006. *19th IEEE International Symposium* (pp. 81-86).

Glover, G., Robin, E., Emami, J., & Arabscheibani, G. (1998) A needs index for mental health care. *Social Psychiatry and Psychiatric Epidemiology 33*, 89-96

Goble, C., Pettifer, S., Stevens, R., & Greenhalgh, C. (2003) Knowledge integration: In-silico experiments in bioinformatics. In I. Foster and C. Kesselman (Eds.) *The Grid 2: Blueprint for a new computing Infrastructure*. Morgan Kaufman

Marshall, M., Lewis, S., Lockwood, A., Drake, R., Jones, P., & Croudace, T. (2005) Association between duration of untreated psychosis and outcome in cohorts of first-episode patients: A systematic review. *Archive of General Psychiatry, 62*, 975-979

NCCMH. (2003) National Collaborating Centre for Mental Health: Core interventions in the treatment and management of schizophrenia in primary care. *Clinical Guideline 1*, NICE, London.

Oinn, T., Addis, M. J., Ferris, J., Marvin, D. J., Senger M., Carver, T., & et al. (2004) Taverna:

A tool for the composition and enactment of bioinformatics workflows. *Bioinformatics Journal 20*(17), 3045-3054

Pedersen, C.B., & Mortensen, P.B. (2001) Family history, place and season of birth as risk factors for schizophrenia in Denmark: a replication and reanalysis. *British Journal of Psychiatry 179*, 46-52.

Phillips, J., Chilukuri, R., Fragoso, G., Warzel, D., & Covitz, P. (2006) The caCORE software development kit: Streamlining construction of interoperable biomedical information services. *BMC Medical Informatics and Decision Making 6*(2).

Sandhu, R., Coyne, E., Feinstein, H., & Youman, C. (1996) Role based access control models. *IEEE Computer 29*, 38-43.

Shibboleth. (2006) http://shibboleth.internet2.edu/ [URL checked and verified 1st August 2006]

Stell, A., Sinnott, R., & Ajayi, O. (2006) Secure federated data retrieval in clinical trials. In *Proceedings of Telehealth 2006*.

Thornley, B., & Adams, C. (1998) Content and quality of 2000 controlled trials in schizophrenia over 50 years. *British Medical Journal* 317,1181-1184.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 2, Issue 2, edited by J. Tan, pp. 1-20, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 2.4
On The Development of Secure Service-Oriented Architectures to Support Medical Research

Andrew Simpson
Oxford University, UK

David Power
Oxford University, UK

Mark Slaymaker
Oxford University, UK

Douglas Russell
Oxford University, UK

Mila Katzarova
Oxford University, UK

ABSTRACT

In this article we report upon our experiences of developing Web-services based infrastructures within two e-health projects. The first—a small demonstrator project funded by the UK's National Cancer Research Institute (NCRI)—is concerned with facilitating the aggregation of different types of data (specifically, MRI scans and histopathology slides) to aid the treatment of colorectal cancer; the second—a rather larger project funded by the UK's Medical Research Council (MRC)—is concerned with the development of a virtual research environment to support neuro-imaging research. In both cases, the underlying infrastructures are being developed by a team that is based in Oxford; it is the experiences of this team that we report upon in this article. We also report upon how we have considered the future potential for our systems interoperating with other systems which are deployed within the UK's National Health Service (NHS).

INTRODUCTION

The potential for distributed, service-oriented architectures to support healthcare delivery, training and research has been acknowledged widely. In this article we report upon our experiences of developing Web-services based infrastructures within two projects. In both cases, the architectures have been influenced by our earlier experiences within the e-DiaMoND project (Brady, Gavaghan, Simpson, Parada, & Highnam, 2003), which explored the development of distributed architectures to support a number of breast-cancer related applications. Although the e-DiaMoND infrastructure was based on grid services—rather than Web services—we have used many of the lessons learnt and adapted many of our designs from that project.

The two projects that we report upon are rather different in nature, but the same broad principles underpin both. The first project—a small demonstrator project funded by the UK's National Cancer Research Institute (NCRI)—is concerned with facilitating the aggregation of different types of data to aid the treatment of colo-rectal cancer. The second project is concerned with the development of a virtual research environment to support neuro-imaging research. In both cases, the projects involve multi-disciplinary teams from a number of institutions, with the underlying infrastructures being developed by a team within Oxford. It is the experiences of this team that we report upon here.

In Power, Politou, Slaymaker, and Simpson (2005), the authors considered the information security requirements incumbent upon health grid architectures deployed within the United Kingdom, and presented an architecture for an idealised health grid that was informed by those requirements; in Power, Politou, Slaymaker, and Simpson (2006), the authors described requirements for, and an approach to, the facilitation of fine-grained access control within systems in which third party Web services are deployed. In this article, we describe how some of the ideas from those articles have been combined to produce designs for, and implementations of, secure infrastructures, which underpin the two aforementioned e-health projects. We also comment our future intentions—which involve building on the work undertaken thus far and considering the potential for inter-operating with systems deployed within the UK's National Health Service (NHS).

The structure of the remainder of the article is as follows. In the next section we describe the background and motivation for our work. Then we reprise the contributions of Power et al. (2005) and Power et al. (2006), which, together, provide a blueprint for our work. Next we report upon our experiences within two projects: an NCRI-funded demonstrator project and the MRC-funded NeuroGrid project. We first introduce the projects, and then present an overview of the technologies used within our solutions. We also briefly consider some of the challenges that we have faced. Finally, we summarise the contribution of this article and outline some areas for future work, much of which is being undertaken with the GIMI (Generic Infrastructure for Medical Informatics) project (Simpson, Power, Slaymaker, & Politou, 2005).

CONTEXT

A number of *e-health* projects have been undertaken in recent years, with the term *grid computing* (see, for example, Foster & Kesselman [1999] and Berman, Fox, & Hey [2003]) often being used within this context. Some interpretations of the term grid computing characterise it as the utilisation of a specific collection of services and toolkits to build a distributed architecture; other interpretations are rather looser and characterise it as the bringing together and sharing of compute and data resources from different administrative domains—in the form of a *virtual organisa-*

tion—to perform tasks that would otherwise be very difficult, if not impossible. In this respect, *compute grids* offer the opportunity to provide unparallelled processing power to facilitate, for example, analysis of 3D images, and *data grids* offer the opportunity to share information between sites to allow distributed data analysis.

The UK's national e-Science Programme (Hey & Trefethen, 2002)—the main aims of which were to build a computational infrastructure to support large-scale research and to identify potential applications for such an infrastructure—funded a number of *e-health* projects. Such projects, including the aforementioned e-DiaMoND, have sought to develop distributed infrastructures to facilitate healthcare research, training, and delivery. Other initiatives have been seen in other countries, with examples including Singapore and Australia.

The EU HealthGrid initiative, aims (amongst other things) to promote the concept of grid computing within the biomedical community, are being undertaken to ensure that relevant technological advances developed by the grid computing community benefit healthcare research and delivery.

Simultaneously, the UK government has invested significant amounts (the initial estimate was approximately six billion pounds; the latest is approximately 12 billion pounds) in a National Programme for Information Technology (NPfIT) (since renamed Connecting for Health) in the NHS, which promises to deliver electronic records, electronic prescription of drugs and electronic booking of appointments, all of which will be underpinned by an NHS Information Technology (IT) infrastructure (Humber, 2004). Similar schemes are being developed throughout Europe and in Australia, Canada, and the United States to provide "cradle-to-grave" views of patients via the linking of electronic information (Cornwall, 2002).

While the potential benefits of the system are significant in terms of increased quality of healthcare delivery, there are potential drawbacks, with many authors (see, for example, Collins [2004], Leyden [2004], Carvel [2005], Keighley [2005], and Mulholland [2005]) being critical. Critics have tended to focus on potential breaches of security and confidentiality. (The interested reader is referred to Anderson [1996] and Anderson [1999] for overviews of the relevant issues.)

It seems almost inevitable that the two paths of e-health research and systems such as Connecting for Health will converge in the near future: with real patient data stored in electronic patient records being used to support medical research. In this respect, it should be noted that research has been characterised as a "secondary use" for Connecting for Health. The following quote from the editors of the Journal of Medical Internet Research supports this view:

One aspect of electronic care records which has received little attention is the potential benefit to clinical research. Electronic records could facilitate new interfaces between care and research environments, leading to great improvements in the scope and efficiency of research. Benefits range from systematically generating hypotheses for research to undertaking entire studies based only on electronic record data ... Clinicians and patients must have confidence in the consent, confidentiality and security arrangements for the uses of secondary data. Provided that such initiatives establish adequate information governance arrangements, within a clear ethical framework, innovative clinical research should flourish. Major benefits to patient care could ensue given sufficient development of the care-research interface via electronic records. (Powell & Buchan, 2005)

Of course, it will first be necessary to consider the confidentiality and security of patient records, and, in particular, appropriate anonymisation and pseudonymisation of data before new interfaces between healthcare delivery and clinical research environments can be facilitated. Other issues include determining appropriate consent arrangements (whether "opt-in" or "opt-out"), ensuring that trust between practitioners and patients isn't compromised, and establishing workable governance arrangements.

Our work over the past three years has been concerned with the design, implementation, and deployment of distributed security solutions (or health grids) to facilitate medical research. Early work—such as that within e-DiaMoND—utilised Globus Toolkit 3 and enterprise-level commercial products. More recent work—for reasons of interoperability and extensibility—has been based on Web services and freely available software.

In Power et al. (2005), the authors considered the information security requirements incumbent upon health grid architectures deployed within the United Kingdom, and presented an architecture for an idealised health grid that was informed by those requirements. In Power et al. (2006), the authors described an approach to the facilitation of system-wide security that enables fine-grained access control within systems in which third party Web services are deployed. We provide a brief overview of the work of these articles in the next section before considering how the theory has been realised in practice.

TOWARDS SECURE HEALTH GRIDS

In this section, we provide brief overviews of the work of Power et al. (2005), which presented an architecture for a secure health grid, and Power et al. (2006), which described an approach to the securing of Web services.

A Secure Health Grid Architecture

Our discussion is necessarily focused on the situation within the United Kingdom: other countries will have their own concerns to address. Within the United States, for example, the Health Insurance Portability and Accountability Act of 1996 (HIPAA) is of concern: the privacy rule "sets forth what uses and disclosures are authorized or required and what rights patients have with respect to their health information" (Verhanneman, Jaco, & De Win, 2003); the security rule "specifies what implementation is obligatory for enforcement of this policy or what reasonable efforts should be [undertaken]'' (Verhanneman et al., 2003).

UK-based e-health projects must adhere to the principles of the Data Protection Act of 1998, which can be stated as follows.

- Personal data shall be processed fairly and lawfully (and in accordance with certain conditions).
- Personal data shall be obtained for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes.
- Personal data shall be adequate, relevant, and not excessive in relation to the purpose or purposes for which they are processed.
- Personal data shall be accurate and, where necessary, kept up to date.
- Personal data processed for any purpose or purposes shall not be kept for longer than is necessary for that purpose or those purposes.
- Personal data shall be processed in accordance with the rights of data subjects under the Data Protection Act.
- Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data.
- Personal data shall not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data.

The DPA is augmented by the Council of Europe's Recommendation on the Protection of Medical Data.

If any UK-based health grid were to be deployed to facilitate healthcare delivery, or if it were to receive data from systems deployed within the NHS, then it would have to consider additional requirements. The NHS comprises a number of independent legal entities—known as *hospital trusts*—with each hospital trust is being responsible for the data held at its sites. This data is released only with respect to the principles of the Caldicott Guardian, which can be stated thus.

- Justify the purpose(s): every proposed use or transfer of patient-identifiable information within or from an organisation should be clearly defined and scrutinised, with continuing uses regularly reviewed by an appropriate guardian.
- Don't use patient-identifiable information unless it is absolutely necessary.
- Use the minimum necessary patient-identifiable information.
- Access to patient-identifiable information should be on a strict need-to-know basis.
- Everyone should be aware of their responsibilities.
- Understand and comply with the law: every use of patient-identifiable information must be lawful.

It should be noted that each trust retains the ownership of all data located at its sites and each trust determines who can access its data (and under what circumstances): this is a model that we have adhered to when developing infrastructures to facilitate research. A goal of our work, then, is to design and deploy systems that allow data to be shared as ensuring that the data owner retains absolute control over who can access which data, when it can be accessed, and even where it can be accessed from.

To this end, a number of use cases—pertaining to information security requirements—were presented in Power et al. (2005). We provide an overview of these below.

- **Distributed queries of patient data.** "A user wishes to query the data held on a subset of the hospitals that form the health grid. Each hospital is allowed to decide its own policy for data access. The user should receive the combined results containing only data that they are permitted to access."
- **Working at a remote site.** "A doctor is working at a remote hospital, which is part of the health grid. The doctor should be able to access data from their home hospital, though their request may be subject to a policy that differs from the one used when they are at their home institution.''
- **Delegation of access permissions.** "A senior health professional would like to grant access to data to a colleague. This access should be temporary, and could be granted to either a named individual or a group of people.''
- **External access.** "Either a health professional working from home or an individual patient wishing to see their own records should be able to access data in accordance with the local hospital's access control policy. The hospital would use a different policy for such external access than would be used for requests from a remote hospital. This use case differs from the others as the request comes from outside of the current virtual organisation.''
- **Modification of data.** "Having made a clinical decision about a case, a doctor wishes to modify the data stored in the health grid. A doctor will only be able to modify data if a hospital's policy allows it. Each hospital is responsible for the data it stores and as such it should keep a record of all modifications made. This use case is similar to the

delegation of access permissions described above, with the only difference being that the data—rather than the policy—is being changed.''

- **Transferring patient records.** "In this use case a patient has moved and is now being treated at a new hospital. As the patient is likely to stay at the new hospital for some time, it would make sense to move their data. To be able to move the data it will first need to be read: this may involve a distributed query as data may already be present at other hospitals. The data will then need to be deleted from one hospital and copied to another—as the responsibility for it has transferred. This will involve the modification of data. Finally the access policies at both of the hospitals may need to be changed to reflect the change of ownership of the data.''

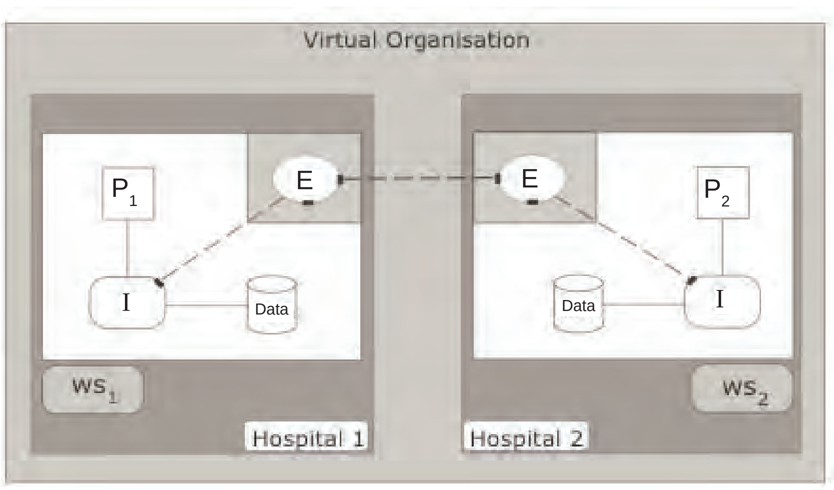
In the architecture of Power et al. (2005), each node contains a data store, externally facing services, internally facing services, access control policies, and workstations (see Figure 1). (This situation is visualised in terms of a health grid with two nodes above.) It is the externally facing services that allow different sites to communicate with each other. Importantly, this architecture allows each site to retain control of its data and access control policies. All access requests within this system are governed by the policies accessible to the internal service, with the policies governing who can access the data, when they can access it, where it can be accessed from, and what rights they have to delegate that access.

In this architecture, a local user can make a request to its local externally facing service, which would then direct that request to: the local internal service; another external service (or set of external services) at a remote site; or both the local internal service and other sites' external services.

Securing Web Services

In Power et al. (2006) an approach to the facilitation of system-wide security that enables fine-grained access control within systems in which third party Web services are deployed was described; in particular, a characterisation of security features required to enable existing Web services to be secured to fit in with a secure infrastructure was presented. We provide an overview of some of these requirements here.

Figure 1.



- **Authenticating clients.** Our assumption is that the system of interest already has an established authentication mechanism. The existing Web service, however, may utilise a different authentication mechanism. When writing a wrapper service, it is possible to authenticate clients using the system-wide authentication mechanism and then translate the system-wide client identifier into one that is understood by the existing Web service via the utilisation of a client mapping function. Once the mapping has been performed, the original SOAP message can be passed to the existing service using whatever authentication mechanism it supports.
- **Secure messaging.** If the content of the message is encrypted, then the message will need to be decrypted and possibly encrypted again. Furthermore, the keys used between the client and the wrapper will differ from those used between the wrapper and the Web service.
- **Access control.** The resource which we are trying to provide access to will typically have its own access control mechanism. Ideally, to be consistent with the requirements of Section 3.1, all access control for the resources at a single node should be determined by a single set of coordinated policies, with all requests for access to the existing Web service having to comply with these policies.

A PLATFORM FOR DATA AGGREGATION TO SUPPORT COLORECTAL CANCER APPLICATIONS

The National Cancer Research Institute (NCRI) Informatics Initiative has facilitated funding for a nine-month demonstrator project to bring together researchers from various disciplines to demonstrate the utility of a multi-scale, multi-disciplinary approach to enhancing the information that can be derived from data collected within a clinical trial. The intention is to develop a prototype system to relate MRI images to the consequence macro images, and to demonstrate the application of medical image analysis (developed for radiological scale images) to macro slides.

The development of a secure computing infrastructure to link disparate sites is being undertaken at Oxford University Computing Laboratory; the development of a prototype viewing application is the responsibility of the Department of Engineering Science at Oxford; the development of appropriate ontologies and meta-models is being undertaking by the Department of Computer Science at University College, London; and data collection is being undertaken by the Pathology Department at Leeds University and the Royal Marsden Hospital. Finally, the task of coordinating all of this activity is the responsibility of the NCRI Informatics Coordination Unit.

In Slaymaker et al., 2006), the development of the underlying computing infrastructure was reported. In this section, we characterise this development in terms of the requirements of a previous section.

Colorectal cancer is first diagnosed using endoscopy and confirmed by histopathology. It is then staged radiologically in order to determine the extent of local and distant disease—with this being best done using MRI. The primary tumour site is assessed to determine if the tumour has extended into the adjacent fat and involved the adjacent tissue planes, including the mesorectum; the detection of lymph nodes and their possible infiltration by the tumour is also crucial, with distant metastases being assessed separately using Computed Tomography (CT) or (if appropriate) Positron Emission Tomography (PET).

Typically, an oncologist sends a patient who is suspected of suffering from colorectal cancer to a radiologist for a CT scan: suspicious dense areas that could be cancer may be revealed; the scans may also give early evidence that, in the case of colorectal cancer, there is already strong evidence of metastasis. A number of options are available on the basis of the CT examination: "palliative care"; sending the patient for MRI scans for further information; providing a course of chemo- or radiotherapy; or surgery. The project is concerned with the last of these options.

During surgery, the surgeon cuts out the tumour. The extracted tumour, together with some flesh, is photographed from the front and the back. The tumour is divided into slices of three-mm thickness, with each slice being captured in slides with two different resolutions: a low-resolution image is taken at about x20 zoom (macroscopic resolution), and a high resolution is taken at about x140 zoom (microscopic resolution). The microscopic slides are analysed by histopathologists to assess type and stage of a cancer, with this analysis consisting of considering the distribution, shape variation, and staining of cells visible at the higher resolution.

The Royal Marsden Hospital has collected (anonymised) MRI volumes of colorectal tumours, with all MRI volumes being taken prior to chemotherapy/radiotherapy and prior to surgery. These volumes are also accompanied with descriptive metadata such as the MRI position, the extent of the tumour, the surgical plan, etc. The other data resource that the project draws upon is a collection of macroscopic and microscopic slides in digital form supplied by Leeds University, with each micro image being stored in a custom file format.

NEUROGRID

Currently, neuro-imaging research typically consists of small studies being carried out in single centres. The sharing of data between centres is limited and even when data is shared, it is common for researchers to be very guarded with their data and algorithms—which leads to much duplication of effort.

The NeuroGrid project (Geddes et al., 2005, Geddes et al., 2006), which is funded by the UK's Medical Research Council, is concerned with tackling problems that are currently holding back widespread data sharing, with the principal aim of the three-year project being to develop a distributed collaborative research environment to support the work of neuroscientists. The potential benefits of the NeuroGrid platform include the streamlining of data acquisition, the aiding of data analysis, and providing improvements to the power and applicability of studies. The project involves collaborators from the University of Oxford, University and Imperial Colleges, London, Nottingham University, Edinburgh University, Cambridge University, and Newcastle University.

As a means of ensuring that the technological solutions being developed within NeuroGrid are of clinical relevance, the project is focusing on a number of exemplars, which pertain to Dementia, Stroke, and Psychosis. The Dementia exemplar requires real-time transfer and processing of images with a view to assuring image quality prior to the patient leaving the examination. The Stroke exemplar is establishing and testing mechanisms for interpretation and curation of image data which are essential to the infrastructure of many multi-centre trials in common brain disorders. The Psychosis exemplar is testing the capabilities of NeuroGrid to: deal with retrospective data, assimilate material into databases, and facilitate data analysis.

SOLUTIONS

In this section we describe the solutions deployed within the projects described previously. Although the approaches taken in both cases are broadly similar we consider different aspects in our discussions of both projects to reflect the different natures of the two projects: first we concentrate on database aspects; and then we concentrate more on Web services.

An Infrastructure for Data Aggregation to Support Colorectal Cancer Applications

It is essential that it is verified that both MRI and pathology images come from the same patient cases, and that these two sets of data can be related (both technically and ethically). If this is the case, we can then relate and integrate pathology and radiology images of a case in two ways. The first means of relating these different types of data is via the integration of the full images as entities—which means having available the pathology image one can retrieve together with its associated MRI image/report. The second method is via the association of their corresponding region of interests (ROI)—which means if one has a ROI in an MRI volume, then one should be able to retrieve its associated area in the 3D pathology images.

Typically, each MRI volume can have more than one ROI, because individual lymph nodes could also be annotated as well as the main tumour (although only one of these annotations would apply to a lesion). Thus, a flexible database schema that can effectively describe all of the above data is necessary. The database underpinning the project is based on the work described in Power, Politou, Slaymaker, Harris, and Simpson (2004), in which a relational structure for images stored in the standard Digital Imaging and Communications in Medicine (DICOM) format is presented.

As the pathology images are not in DICOM format, a different method of dealing with them is required. The two formats that need to be handled are JPG for the macroscopic images and a custom file format for the microscopic images. We have chosen to employ a method that effectively automatically generates suitable DICOM wrappers for each image file. This has the benefit of allowing maximum flexibility while requiring minimal changes to the existing underlying schema.

Another advantage of using DICOM-style wrapping is that it handles collections of images well. A single set of macro images would form a DICOM series and hence be addressed together. This is also the case with the micro images. Furthermore, the two series—macro and micro—can be collected together as a single study. This provides a great deal of flexibility when processing queries.

All data is stored in a federated database that is distributed over several sites, with all of the data servers being exposed via Web service interfaces. The patient and image data are stored at the site at which they originated, with the data from each location reflecting the specialisation of that site. The selection of MRI and pathology data pertaining to a single case is achieved by federating data between multiple sites. The architecture is based on that of a prior section and utilises open source Web services technologies and standards.

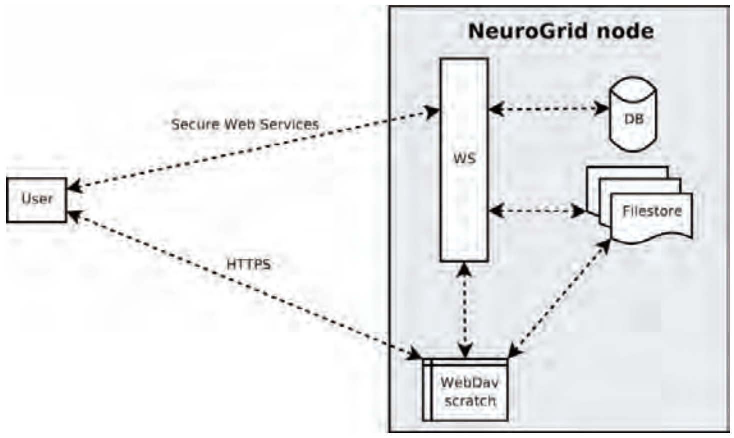
The NeuroGrid Infrastructure

While our developments within NeuroGrid have followed the sentiments of the architecture of Power et al. (2005), the system is by no means a direct implementation. For example, there has been no need to distinguish between internal and external services as both kinds of service have been written and deployed by the NeuroGrid core technology team. Should there be any need in the future to decouple the internal and external services (if, for example, a third party internal service were to be used), then this could be achieved via the existing services calling through to the internal services.

The configuration of a NeuroGrid server node is illustrated in Figure 2.

Ensuring secure messaging was our first challenge, and we use secure Web services to this end. In addition, we utilise WebDAV (Web-based Distributed Authoring and Versioning) (Whitehead, 1998) folders to provide scratch space for operations. The Web services and WebDAV folders are accessible only by using PKI mutual authentication using X.509 certificates. The Web

Figure 2.



services use an insecure channel to carry secure messages signed by the originator and encrypted for the recipient, ensuring both the identity of the sender and secrecy. The users communicate with the WebDAV server over HTTPS with mutual authentication enabled—which establishes a secure channel over which insecure messages can be passed.

WebDAV plays a key role in the NeuroGrid architecture. First, there is no standard high-speed and secure way of transferring binary data using SOAP messages; WebDAV allows us to both get files from and put files to Web servers in a secure way efficiently. Second, WebDAV allows the user to browse, upload and download files directly to their scratch space using a variety of third party WebDAV clients—this is especially useful when wrapping the Web services with a portal. The WebDAV folders also provide an ideal location for any intermediate results that might be generated by algorithms which do not belong in the file-store but may be of interest to the user.

A file-store and a database of meta-data provide part of the back-end of the system: this is where data is stored, with all interactions occurring through Web service calls. We separate image data from patient and meta-data (with appropriate references between the two to guarantee referential integrity). The use of the Apache Derby database—which is written entirely in Java, is easy to install, and has a minimal system footprint—was chosen as it enhances the interoperability and portability of the underlying infrastructure.

We are also using Sun's JWSDP (Java Web Services Development Pack) Web service implementation in the Sun customised version of the Apache Tomcat container. The intention behind using Java-based technologies is that it should allow us to deploy to any platform with a Java Virtual Machine; using Web services allows maximum interoperability with client implementations in a variety of languages. We are also using the Apache Web server which is available for most platforms.

The technologies we have used to create the NeuroGrid infrastructure were chosen to ensure that it would be as portable and interoperable as possible: we have attempted to choose solutions with excellent cross-platform support and which is at freely available—if not open source.

We now consider how the use cases of the third section are being realised.

- **Distributed queries of patient data.** This use case requires the ability to perform federated queries—an ability that the NeuroGrid infrastructure provides. To enable federated queries the system must support onward Web service calls—a situation in which one Web service makes a call to another Web service, which can be local or on a remote node. A request to a federated query Web service at one site will query local resources and, in addition, make Web service query calls to other known sites. On receipt of the results of its Web service queries the federated Web service assembles all the data into a single set of results and returns this to the caller of the federated Web service.

In order to identify the originator of the request, a *ticketing* system is being used. The ticketing system iteratively propagates information pertaining to authentication: each Web service call has a ticket associated with it and each Web service examines the tickets it receives. When a Web service creates a new ticket, the calling Web service's ticket (or a ticket issued to the calling user in the initial case) is embedded in the new ticket. Tickets are always signed by the sender ensuring the validity of the ticket can be verified by all subsequent receivers. The tickets constitute a verifiable audit trail which can be used to restrict access control when combined with a relevant access control policy.

- **Working at a remote hospital.** This simple use case again relies on the ticket system. At the points when access control policies are evaluated, the audit trail of the ticket can be consulted and that shows exactly the route the user has taken to make the request. Access can then be restricted by combining a policy which restricts access based on origin and/or route.

NeuroGrid already has the ability to onwardly call Web services (one Web service making a call to another Web service). Using this capability a user can make calls to a node at the hospital where they are currently located and that node will make appropriate requests to the user's home hospital on their behalf.

In a hospital situation it would be appropriate to have nodes which could only be communicated with locally—apart from node-to-node communication which is essential to the system's operation. This could be achieved by several means, such as IP address restrictions or the utilisation of a Virtual Private Network.

In the currently deployed NeuroGrid system, nodes are all public facing, which means that a user could bypass the step of calling the Web services on a local node and instead call directly to a remote node.

- **Delegation of access permissions.** In the original presentation of the use case of Power et al. (2005), the delegation of access permissions was similar to that pertaining to the modification of data; within NeuroGrid we are also looking at a model of delegation based on the aforementioned ticket system.

Using the ticket system a user can delegate their permissions to access data to another user. Using the ticket system, the originator of the request is known—and thus their identity can be used in making access control decisions. This information can also be used to delegate authority from one user to another.

- **External access.** In Power et al. (2005), external access referred to any access that was not initiated from within a site of the virtual organisation, such as a doctor accessing some medical records from home. From the perspective of the NeuroGrid infrastructure, all access is considered external: in the current system, physical location does not affect access rights.

There are two ways to access the NeuroGrid system: talking directly to a server using an application that utilises an API or through a Web portal (which itself is just an application using the API).

It is trivial to establish external access as any location not in a restricted list. This could then be used to deny direct access to NeuroGrid to any location not on the list. Locations not on the list could access the NeuroGrid through the Web portal—with more restricted rights than permitted to an internal user.

It is well understood that it is difficult to establish a definite location of a client on a network. As such, it could be argued that this "all or nothing" (without going through the portal) model represents a sensible approach: the users' rights would simply be restricted when accessing the system though a portal. Additionally, there is no reason why externally-facing NeuroGrid nodes could not be configured to give the same reduced rights to users directly through Web services.

The access control policies can take advantage of the tickets to accomplish this restricted access as they contain the audit trail of the request: any requests which have passed through a specific node can be given restricted access rights.

- **Modification of data.** Within the current implementation, it is perfectly possible to modify the data held in the system; all modifications are recorded as part of an audit system, so changes to the data can be tracked back to the responsible party if necessary. Again, the implementation of this use case is benefiting from the ticket system (combined with access control policies) to allow and disallow rights to modify data as necessary.
- **Transferring patient records.** It is entirely possible within NeuroGrid to copy some data (in this case, a patient record) to another node. Ideally, this process would take place within the scope of an atomic transaction ensuring that the patient record is neither duplicated, nor, more seriously, lost. Support for atomic transactions is part of the WS-Transaction specification, and support for this specification is under development as part of a future version of JWSDP. Once this becomes available we plan to add transactional support to the NeuroGrid system.

As any deployment of the NeuroGrid infrastructure evolves, modifications of access control policies—to reflect the change in location of data—will occur. For a standard record, this will typically mean that it is added to a list of records associated with a standard access control policy at the hospital it was transferred to—with any relevant specific policies for that patient also being transferred.

Challenges

Our goal is to develop secure distributed systems that will support a wide range of platforms: by using open standards we aim to support interaction with third party systems that may be developed in the future. By using a platform-independent programming language (Java) and open standards for communication (HTTP, SOAP) and security (TLS, WS-Security) we should—in theory—be able to meet these objectives. We also aim for our systems to be free to deploy and use—which prohibits the use of most commercially produced software.

Although large vendors provide free versions of commercially available software, they (typically) provide limited support for its use. In spite of this, we have chosen to use several software packages that fall into this category alongside some open source components. We discuss two of the challenges that we have faced in this regard in the following.

The distributed systems we have developed consist of multiple components, which communicate using a combination of WebDAV and Web services. The WebDAV services are provided using an Apache Httpd server with DAV and SSL extensions—which allows us to provide a mutually authenticated WebDAV service. The URLs used tend to be long—for technical and security reasons (and user convenience). This had led to several problems on the client-side, where ideally we would like to use the default WebDAV clients provided as part of the operating system.

Microsoft provides a WebDAV client built into Internet Explorer and also support the opening of WebDAV folders as a Network Connection. We have had two problems with these clients. The first is the lack of support for long URLs in Internet Explorer—a series of support calls over a period of several months led to a suggestion that we shorten the URLs. We also have had problems with Microsoft clients trying to write to the root of the server's file system as this would be normal behaviour on a Microsoft server. Apple also provides a WebDAV client which is used primarily for their Mac services. This client has the problem that it does not support mutual authentication.

As a means of providing Web services, we use the Sun Java Web Services Development Pack (JWSDP) running on top of a Sun provided version of Tomcat 5.0 (Sun do not currently support a version of Tomcat 5.5 although the reason for this is unclear). Part of our reason for choosing the Sun provided software instead of the open source Axis was because of its support for a large number of the Web service standards. Unfortunately, the WS-security libraries require extensions to the Java security libraries—which are shipped with all Sun-provided JVMs, but not IBM-provided JVMs. This has led to problems running our Web services on IBM servers.

There are a number of lessons to be learnt from this. First, the use of Web services to develop *real, interoperable* systems to support genuine virtual organisations still has some way to go. Second, it is arguable that the use of extensible standards to facilitate interoperability is doomed to failure as long as vendors are able to release product versions underpinned by their own closed-source extensions to these open standards.

DISCUSSION

In this article we have described how some of the requirements and approaches described in Power et al. (2005) and Power et al. (2006)—pertaining to the design of health grid architectures and the securing of Web services, respectively—have been realised in two e-health projects.

Developing robust distributed solutions using current technologies is a challenging task. One of the major issues that we have faced is that current toolkits for Web services are very much focused on a single server being contacted by multiple clients; as such, are merely a substitute for a traditional HTML-based solution. We would argue that inadequate thought has been given to building multiple-server applications. For example, we have encountered several problems when making onward calls from one Web service to another. Other challenges include "standard" extensions not being standard and a lack of a mechanism for transferring large binary objects that is both secure and efficient *simultaneously*.

Future work is being taken in a number of directions, much of which is being undertaken within the GIMI (Generic Infrastructure for Medical Informatics) project (Simpson et al., 2005). The main aim of GIMI is to develop a generic, dependable middleware layer capable of: (in the medium-term) supporting secure and ethical data sharing across disparate sources to facilitate healthcare research, delivery, and training; and (in the longer-term) interfacing with technological solutions deployed within the NHS.

The key to ensuring legal and ethical access to such data is that the mechanism should offer *fine-grained* and *dynamic* access control to resources: this is the key driver behind GIMI. The infrastructure will be developed so that it is sympathetic to the needs of clinical researchers, commercial organisations involved in the medical domain, healthcare providers, and those concerned with providing training facilities—all of which will be concerned with sharing confidential data. Our plan is to extend our existing designs and implementations to facilitate such means of authorisation.

First, we have made significant strides towards realising the architecture of Power et al. (2005). There is some way to go, however: the addressing of the "transferring of patient records" use case, for example, has yet to be addressed. Continuing to realise these use cases within the context of the NeuroGrid project shall continue to be a priority.

Second, our architectures have been designed with the opportunity for fine-grained and flexible access control policies in mind. Our medium-term goal is to realise this opportunity by utilising XACML to secure nodes in such a fashion. This—together with the long-term interoperability view—is our focus within GIMI.

Finally, to date, our techniques have been deployed exclusively within the healthcare domain; we intend, in the near future, to determine other relevant application domains in which our approaches may be validated.

REFERENCES

Anderson, R. (1996). Patient confidentiality – at risk from NHS-wide networking. In B. Richards and H. de Glanville (Ed.), *Current perspectives in healthcare computing* (pp. 687-692), BJHC Books.

Anderson, R. (1999). Information technology in medical practice: Safety and privacy lessons from the united kingdom. *Medical Journal of Australia*, 170, 181-185.

Berman, F., Fox, G.C., & Hey, A.J.G. (2003). *Grid computing: making the global infrastructure a reality*. Wiley.

Brady, J.M., Gavaghan, D.J., Simpson, A.C., Parada, M.-M., & Highnam, R.P. (2003). e-DiaMoND: A Grid-enabled federated database of annotated mammograms. In F. Berman, G.C. Fox, & A.J.G. Hey (Ed.). *Grid computing: Making the global infrastructure a reality,* (pp. 923-943),Wiley.

Carvel, J. (2005). Patients can stay off NHS database. *The Guardian* (14th January, 2005).

Collins, T. (2004). GPs vote to boycott patient record database. Retreived by www.computerweekly.com/Article131577.htm.

Cornwall, A. (2002). Electronic health records: An International perspective. *Health Issues*, 73.

Foster, I., & Kesselman, C. (1999). *The Grid: blueprint for a new computing infrastructure*. Morgan Kaufman.

Geddes, J., Lloyd, S., Simpson, A.C., Rossor, M., Fox, N., Hill, D., et al. (2005). NeuroGrid: using grid technology to advance neuroscience. In *Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems (CBMS*), IEEE Computer Society Press.

Geddes, J., Mackay, C., Lloyd, S., Simpson, A.C., Power, D.J., Russell, D. et al. (2006). The challenges of developing a collaborative data and compute grid for neurosciences. In *Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems (CBMS)*, IEEE Computer Society Press.

Hey, A.J.G. & Trefethen, A.E. (2002). The UK e-Science Programme and the Grid. In *Proceedings of Computational Science (ICCS 2002), Part I*, Springer-Verlag Lecture Notes in Computer Science, volume 2329, 3-21.

Humber, M. (2004). National programme for information technology. *British Medical Journal*, *328*(7449): 1145-1146.

Keighley, B. (2005). Confidentiality – the final betrayal. *British Medical Journal career focus*, *330(*259).

Leyden, J. (2004). US wins David Blunkett lifetime menace award. Retreived from www.theregister.co.uk/2004/07/29/big_brother_awards.

Mulholland, H. (2005). Patients fear online records threaten confidentiality. *The Guardian* (29th June, 2005).

Powell, J., & Buchan, I. (2005). Electronic Health Records Should Clinical Research. *Journal of Medical Internet Research*, *7*(1), e4.

Power, D.J., Politou, E.A., Slaymaker, M.A., Harris, S., & Simpson, A.C. (2004). An approach to the storage of DICOM files for Grid-enabled medical imaging databases. In *Proceedings of the Annual ACM Symposium on Applied Computing (SAC '04)*, ACM Press.

Power, D.J., Politou, E.A., Slaymaker, M.A., & Simpson, A.C. (2005). Towards secure Grid-enabled healthcare. *Software: Practice and Experience*, *35*(9), 857-871.

Power, D.J., Politou, E.A., Slaymaker, M.A., & Simpson, A.C. (2006). Securing Web services for deployment in health grids. *Future Generation Computer Systems*, *22*(5), 547-570.

Simpson, A.C., Power, D.J., Slaymaker, M.A., & Politou, E.A. (2005). GIMI: Generic Infrastructure for Medical Informatics. In *Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems (CBMS)*, IEEE Computer Society Press.

Slaymaker, M.A., Simpson, A.C., Brady, J.M., Gavaghan, D.J., Reddington, F., & Quirke, P. (2006). A prototype infrastructure for the secure aggregation of imaging and pathology data for colorectal cancer care. In *Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems (CBMS)*. IEEE Computer Society Press.

Verhanneman, T., Jaco, L., & De Win, B. (2003). Adaptable access control policies for medical information systems. In J.B. Stefani, I. Demeure, & D. Hagimont (Eds.) *Proceedings of DAIS 2003*, (LNCS, Vol. 2893, pp. 133-140). Springer-Verlag

Whitehead, J. (1998). Collaborative Authoring on the Web: Introducing WebDAV. *Bulletin of the American Society for Information Science, 25*(1), 25-29.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 2, Issue 2, edited by J. Tan, pp. 75-89, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 2.5
Decision Support With BPEL and Web Services

Iain Morrison
The University of Melbourne & Bond University, Australia

Sony Nugrahanto
The University of Melbourne, Australia

ABSTRACT

A review of difficulties experienced with proprietary guidelines languages and in clinical decision support has led us to the development of a decision support system (DSS) based on a service oriented architecture (SOA) using Business Process Execution Language (BPEL) and Web services. BPEL allows simple branching conditions and tests to be set based on the information artifacts in the workflow. For more complex tests, we preferred to place decision support in an external DSS Web service. An advantage of this approach is that the DSS could evolve independently of the clinical workflow. We have implemented an exemplar of the above using the Java expert system shell (JESS). The domain objects are sent by the BPEL engine into the DSS which consults external rules, reasons on the domain objects and rules, and transmits recommendations back. The article will discuss issues involved in implementation using as an example drug-drug adverse reactions.

INTRODUCTION

In an earlier article (Liaw, Deveny, Morris, Lewis, & Nugrahanto, 2004), we developed an approach to Clinical Decision Support (CDS) based on the BPEL and Web services. The process flow associated with the treatment (diagnosis, referrals, consolidation, and action), was expressed, following expert group analysis, in a series of blocked activities which could then be drilled down into specific actions. The expert analysis led to (i) a set of activities (tasks or services or actions) and decision points; (ii) a set of information requirements at each stage of the process flow, to be mapped onto information sources and sinks; and (iii) a "best-practice" sequence of activities and actions. The sources and sinks and activities could be resolved into local process variables, or a Clinical Information System (CIS) or a set of external service providers described through Web Service Description Language (WSDL) (W3C, 2001).

In a follow-up article (Morrison, Lewis, & Nugrahanto, 2006), we explored issues in interoperability between the workflow engine executing the decision support guideline and the endpoints supporting the external service providers (including manual intervention). We showed the pros and cons of using either SOAP-encoded RPC-style information transfer or document-style transfer and how these approaches were reflected in the design of Web services endpoints and the types of Information Models able to be used to transfer application data.

The examples we provided at that time performed decision support within the workflow engine, utilizing BPEL decision constructs. However, this approach is not suitable when the decision is complex or if flexibility is required in environment/language bindings as in rule-based approaches commonly used in clinical decision making.

In this article, we present and discuss implementation of an evolved architecture, focussing attention on the separation of concerns amongst clinical workflow, reasoning engines, and rules bases in drug-drug adverse reactions as an application scenario. We place complex decision support in an external DSS Web Service accessible to the workflow engine in the same way as the EHR and other service providers Thus, the detail of the DSS implementation is abstracted from the workflow description and its mechanisms hidden behind an interface façade. Different reasoning engines can be made available depending on the application context and rules bases can be updated following expert review independent of the CDS processes. The rules bases could be maintained centrally and the most current version loaded on-demand.

SERVICE ORIENTED ARCHITECTURE AND CLINICAL DECISION SUPPORT

Service Oriented Architecture

The basic principles of SOA consist in modularizing functions and exposing them as services. This idea of a software application as a service was recognized in the past (Sun's RPC, Microsoft's COM/DCOM), but it can now be fully realized using the Web services for systems interoperability.

A Web service is a software application available on the Web whose capabilities are described in XML and is able to communicate through XML messages over an Internet transport protocol. Basic elements of Web services are the standards for interoperability: XML for information representation), SOAP (W3C, 2000) for invoking, WSDL (W3C, 2001) for describing, and UDDI (http://www.uddi.org) for discovering.

On top of this basic interoperability protocol stack, new languages and specifications for defining the composition of Web services to form business processes have emerged, such as Business Process Execution Language (BPEL) (IBM, 2002). Characteristics of BPEL include a service orientation model that emphasizes standards compliance and XML encoding. As well as a formal basis to the language, services are loose coupled and can support synchronous and asynchronous activation.

When compared to related standards (such as XPDL and WSCI), it appears that BPEL is relatively expressive (Wohed, van der Aalst, Dumas, & ter Hofstede, 2003) and currently BPEL is the only standard that comes with execution engines such as Oracle BPEL Manager (Oracle, 2006), IBM's BPWS4J (IBM, 2004a) and ActiveBPEL (ActiveBPEL, 2006)

SOA and CDS

Benefits from CDS include improved patient safety, improved quality of care, and improved efficiency in health care delivery (Sinchenko, Westbrook, Tipper, Mathie, & Coiera, 2003). However, these goals are yet to be fully realized largely due to poor integration with clinical workflow and practice and poor integration with EHR and/or other clinical systems (Eccless et al., 2003).

In recent years, we have seen development in SOA-related standards and technologies that could potentially to provide answers to some of the technical challenges around integration and interoperability in the clinical and health care context. Research in this areas are quite fluid at the moment but initial indications are positive (Peleg, 2005; Kawamoto, 2005; Anzböck & Dustdar, 2004; Anzböck & Dustdar, 2005).

Recognizing the need to develop service interface specification standards to support clinical decision support in SOA, the Health Level 7 (HL7, http://www.hl7.org) and Object Management Group (OMG, http://www.omg.org) have agreed to collaborate in the Healthcare Services Specification Project (HSSP, http://hssp.wikispaces.com). The project has identified some of the services that may be useful for the implementation of a CDS system. These services concern with, for example, managing patient identity, locating records, verifying terminology, and drawing conclusion based on patients' data.

In Australia, NEHTA (The National E-Health Transition Authority) has recently recommended the use of SOA for the design of health applications and has produced an Interoperability Framework (NEHTA, 2005) and a Web Services Standards Profile (NEHTA, 2006) lending further weight to the adoption of these standards and approaches.

CDS with Web Services and BPEL

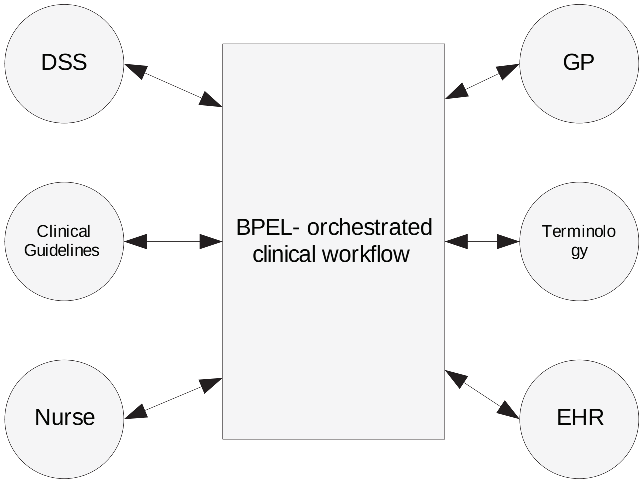
As clinical decision support requires a robust distributed architecture (Nealon & Moreno, 2003), Web services are a strong candidate as a delivery mechanism. BPEL has been developed in tandem with the development of Web services, firstly as a process description and execution language, and then, due to availability of Java-platform Application Programming Interface (API) and execution engines, as a means of automating (orchestrating) the coordination of Web services.

In using SOA (BPEL and Web services) as an architectural design, we noted (Morrison et al., 2006):

- Web services may not completely solve interoperability problems between clinical systems, however, they show great promise of reducing the complexity of interconnecting heterogeneous software systems distributed over the Internet partially by providing a level of service and information flow abstraction.
- The expressiveness of BPEL as a process description language to model business processes (Wohed et al., 2003) as well as providing execution engines supportive of web-services-based tasks/activities makes it a strong candidate for modelling coordination required in a clinical setting to assist decision making.
- BPEL cleanly separates transport, process definitions and information flows. Information flows can be defined by schemas (document-style) or mappings to classes/objects which are defined by SOAP-encoded RPC.

Coordination of decision-making process in BPEL can be done in a way that reinforces the centrality of the roles enacted (see Figure 1),

Figure 1. Representation of roles assignments using BPEL



whether they be by the General Practitioner (GP), EHR, or another entity, and specifies the interfaces to be implemented in a technology-neutral way. Although BPEL does not directly support a "role" at a language level, these can be affected through environmental bindings. BPEL then makes these roles and services explicit and transparent while hiding details of service implementations—in our earlier examples, we had a GP, practice nurse, and external service providers.

In the SOA architecture (with BPEL and WS), for example, roles can be assigned to a healthcare–provider, a patient (and patient health record), the DSS/Workflow engine and associated services, such as terminology or medicines information.

Decision Making Strategies with BPEL

BPEL allows simple workflow branching conditions and tests to be set based on the information artefacts provided by the services or held in the process execution context. For example, in the case of asthma (Liaw et al., 2004), we could check the patient's age and X-Ray consolidation (extracted from CIS or Provider reports as external services) before deciding on the next course of action.

For more complex tests, we could embed snippets of Java code into BPEL as proposed by IBM and BEA (IBM, 2004b). The BPELJ extension will allow BPEL to not only to orchestrate web services but will also allow orchestration of local resources which is more efficiently accessed with local means. However, this approach as noted earlier would be at a cost to the portability, language and platform neutrality of BPEL as the extension deals only with Java 2 Platform Enterprise Edition (J2EE) (Sun, 2001) resources.

Therefore, we will investigate placing complex decision support in an external DSS Web service accessible to the workflow/process engine in the same way as the EHR and other service providers.

An advantage of this approach is that the clinical workflow, reasoning engine, and rules bases are all separated. Furthermore:

- The approach places all services on an equal conceptual footing. In particular, the DSS as a service could itself call on other services independently and as required.

- In the Web Services context, the DSS could itself then be comprised of other complex flows and be independent of the workflow provided it met the required interface contract-for-service.
- The approach decouples the clinical workflow and timing from the decision tests. In the case of the introduction of new drugs or the publishing of an update or an alert on an existing medication, the clinical workflow would not be affected however the DSS would draw on the new information.

THE TEST APPLICATION SCENARIO

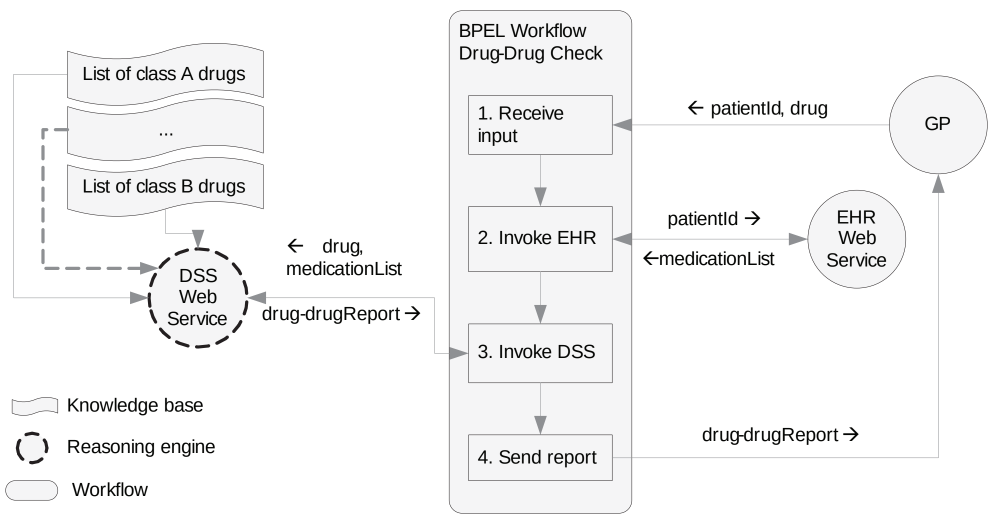
We will look at the example of drug-drug adverse reactions. Drug-drug interactions occur when two or more drugs react with each other with potentially adverse consequences to the patient to whom they are being administered. For example, mixing a sedative and an antihistamine can slow reactions and make driving a car dangerous.

We model a GP interacting with a patient and deciding on a course of medication (see Figure 2).

The proposed medication, determined by the GP after consultation with the Patient, is sent with the patient external ID to an external (to the workflow engine) EHR service which then calls the EHR (practice-based or externally-based) and extracts the current medication list (expressed in this case as a drug list modelled as extracted from a Medical Director (http://www.hcn.com.au/) application (Practice CIS).

The current medication list and the proposed medication goes to the DSS which has loaded (i) drug concepts and drug classes and (ii) decision support rules in terms of these classes. The current implementation (there could be many rule sets per supported CIS) maps the proposed medication into a drug class and checks for known adverse

Figure 2. BPEL scenario for drug-drug collisions checking



reactions with other drug classes. That is we did not map specific medicine preparations but more types of drug (e.g. determined as per active ingredient, not preparation). This, of course, could be easily generalised. The examples we used in testing were drawn from the Therapeutics Goods Administration (TGA) Adverse Drugs Reaction Bulletin (http://www.tga.gov.au).

The decision-making process implemented was choreographed in a synchronous BPEL process and all services were implemented as Web Services (EHR, DSS, and GP) that exchanged activity and information via the SOAP protocol over HTTP. Although we focus in this article on rule-based decision support, and implemented the test scenario in a simple Web Services Profile (equivalent to the NEHTA Basic v1.0 Profile), the addition of other services (e.g. WS-Reliable Messaging, WS-Security) and alternate communication models and transports is quite straightforward as shown in an earlier analysis of practice workflow and interoperability framework (Morrison, Lewis, Liaw, 2005).

Rule Engine

For the reasoning engine behind the DSS service facade we used Jess (Friedman-Hill, 2006), a rule engine developed at Sandia National Laboratories in the late 1990s and written in Java. Creating and invoking Java objects from Jess (or vice versa) is fairly transparent if the objects have JavaBeans (Sun, 1997) properties and reflect domain constructs able to be represented in XML schemas. As a result, domain-level attributes can be easily mapped into specific rules and facts and reasoned on.

The current Jess implementation does not support, however, automatic conversion of collection objects other than Java arrays but this restriction parallels those involved in the choice of Document-based rather than SOAP-encoded RPC-based content model transfer in Web services and is more an application-level design issue. It does not have practical limitations for the types of information models currently used in CIS Extracts and on Drug-Drug and Drug-Condition tests.

Architecturally, Jess is well suited to Web services DSS use as it complies with the JSR94 specification (JSR94, 1994) for “plug-and-play” rules engines for Java-based Decision Support and is the first engine to be endorsed under this community standard. This allows the DSS façade to be backed by an API that can select specific engines at run-time based on application requirements—facilitating multi-facetted DSS services. The concept is very similar to having multiple messaging providers and directory service providers behind a standardised service interface. Applications see the standard interface and providers can be switched.

One of interesting features of Jess is the notion of shadow facts. A shadow fact is an unordered fact whose slots/properties correspond to the properties of a JavaBeans. Therefore shadow facts serve as a connection between facts in Jess working memory and Java objects representing the facts.

Workflow Engine

For the workflow engine we used ActiveBPEL BPEL engine (ActiveBPEL, 2006), an open source and free-to-use implementation of BPEL engine with a strong community support. It has a visual tool to design workflows and thus minimizing the need to script them at textual level.

We should mention that as this engine is BPEL-compliant, we had no difficulty importing our examples in the treatment process for asthma (Liaw et al., 2004), previously developed in Oracle BPEL Process Manager (Oracle, 2006). However, we noted that one must develop in this case a library to accommodate the omission of user task management (a given, straight-out-of-the-box feature in Oracle PM). Also, there is no persistence mechanism by default but we could use any standard or public domain database as the linking mechanism has been provided as part of the suite.

EHR

For implementation of the EHR service we have extracted data from the Medical Director Practice CIS and organized them into a Clinical Document Architecture (CDA) (HL7, 2004) structure appropriate for Summary Clinical Data Extracts. Using the associated XML schema, we populated our EHR using XMLBeans (Apache, 2006). As noted, our approach is schema-driven throughout. The patient summary schema is not only useful for populating the EHR, but can easily be used to import the information models into our BPEL workflow. Using this approach, we are able to easily change patient record structures (and other required information structures associated with support services) as these are agreed to extend our prototype. In this way, clinical summaries and external services report content is made available to both the BPEL Process and the DSS. The approach and architecture are quite flexible and allow the data to be sent directly (encrypted, de-identified, and through choice of multiple communication channel—mail, http, Message Oriented Middleware [MOM], etc.) or by reference/delegation and extracted as required by the workflow engine or DSS services.

RESULTS AND SUMMARY

We applied the above architecture and test scenario to a number of drug-drug adverse reaction alerts published by the TGA. In the cases considered the Process Engine and DSS operated within specification in detecting specific collisions between drug classes. The drug codes (by preparation) as prescribed and listed in the CIS were mapped to drug classes and the Jess rules scripted (i) to encode class definitions and (ii) define adverse reaction rules. The latter were drawn from the expert reference and the former from application (CIS) mapping tables. Once the class concepts were defined (schema, Java, or facts), the rules for each adverse reaction alert were trivial (a few lines of text each). The application mappings from a given CIS to drug classes needs done only once per CIS, however we note that a move to standardize medication descriptions (especially in schema form) would largely eliminate this last step. The rules base itself transparently factors into three elements (i) the adverse reaction alerts in standardized form, (ii) the concept definitions (drug class etc.), and (iii) per-CIS drug-class mappings. This leads to easier maintenance.

The tests show, in extension to our earlier work, that an SOA, Web services and BPEL approach to complex DSS in clinical care is feasible and architecturally exhibits useful features in (i) support for service and DSS abstraction; (ii) separation of concerns of clinical workflow, information models and flow, services, and transports; (iii) factoring of DSS concerns (concepts, rules, and application mappings); and (iv) concurrent support for multiple DSS mechanisms behind a common service façade. Although this article focused only on one specific application, the approaches support scalability and generalization at several levels, including drug-condition and related condition/co-morbidity checks.

The HSSP project has identified several services required to support clinical decision support in SOA. With regard to the DSS service, the project seeks to identify a common service interface for how DSS services make their knowledge available to the decision support applications (CDSs). As more functional and technical details become available, we could incorporate work on this area into our propotype.

REFERENCES

ActiveBPEL (2006). *BPEL engine*. Retrieved from http://www.activebpel.org/

Anzböck, R., & Dustdar, S. (2004). *Medical services workflow with BPEL4WS*. Department of Information Systems, TU Wien.

Anzböck, R., & Dustdar, S. (2005). *Semi automatic generation of Web Services and BPEL Processes: A Model-Driven Approach.* (LNCS 3649, pp. 64–79).

Apache (2006). *XMLBeans specification.* Retrieved from http://xmlbeans.apache.org/

Ecless M., McColl E., Steen N., Rosseau N., Grimshaw J., Parkin D., & et al (2003). *Effect evidence based guidelines on management angina in adults in primary care: Cluster controlled trial*. BMJ 2002, Vol. 325.

Friedman-Hill, E. (2006). *Java Expert System Shell* (Jess). Retrieved from http://herzberg.ca.sandia.gov/jess

HL7 (2004). *Clinical Document Architecture (CDA)*. Retrieved from http://www.hl7.org/

IBM (2002). *BPEL 1.1 Specification.* Retrieved from http://www.ibm.com/developerworks/webservices/library/ws-bpel/

IBM (2004a). *IBM's BPWS4J 2.1.* Retrieved from http://www.alphaworks.ibm.com/tech/bpws4j

IBM (2004b). *BPELJ specification.* Retrieved from http://www.ibm.com/developerworks/webservices/library/ws-bpelj/

JSR94 (1994) *Java Rule Engine API.* Retrieved from http://jcp.org/aboutJava/communityprocess/review/jsr094/

Kawamoto, K. (2005). *HL7 clinical decision support (CDS) service framework proposal.* Retrieved from http://cslxinfmtcs.csmc.edu/hl7/arden/2005-09-SAN/cds-service-framework-proposal-v1.00-2005-09-14.doc

Liaw T., Deveny, E., Morrison, I., Lewis, B., & Nugrahanto, S. (2004). *Modeling the clinical processes of prescribing (MCPOP) project*. Department of Health and Ageing, Australia.

Morrison, I., Lewis, B., & Nugrahanto, S. (2006). Modelling in clinical practice with Web Services and BPEL. *International Journal of E-Business Research*, *2*(4), 45-57.

Morrison, I, Lewis, B., & Liaw, T. (2005). *Application interoperability in general practice*. Report to Department of Health and Ageing and the RACGP, June 2005 (60 pages).

Nealon, J. L., & Moreno, A. (2003). Agent-based applications in health care. In J.L. Nealon & A. Moreno, (Eds.) *Applications of software agent technology in the health care domain* (pp. 3-18). Whitestein Series in Software Agent Technologies.

NEHTA (2005). *Towards a health interoperability framework v1.8.* Retrieved from http://www.nehta.gov.au/index.php?option=com_docman&task=doc_details&gid=26&Itemid=53

NEHTA (2006). *Web Services standards profile.* Retrieved from http://www.nehta.gov.au/component/option,com_docman/task,doc_download/gid,67/

Oracle (2006). *Oracle BPEL process manager.* Retrieved from http://www.oracle.com/technology/bpel/index.html

Peleg, (2005). *Open source publishing of medical knowledge for creation of computerinterpretable guidelines.* Retrieved from http://mis.hevra.haifa.ac.il/~morpeleg/pubs/AIME_2005.pdf

Sinchenko, V., Westbrook, J., Tipper, S., Mathie, M, & Coiera, E. (2003). Electronic decision support activities in different healthcare settings in Australia. Report for National Electronic Decision

Support Taskforce, Commonwealth of Australia. In E. Coiera (Ed.) *Guide to health informatics, Chapter 25: Clinical decision support systems*. Retrieved from http://www.coiera.com/aimd.htm

Sun (1997). *JavaBeans specification*. Retrieved from http://java.sun.com/products/javabeans

Sun (2001). Java 2 Enterprise Edition (J2EE) Specification. Retrieved from http://java.sun.com/javaee/

Wohed, P., van der Aalst, W.M.P., Dumas, M., & ter Hofstede, A.H.M.(2003). *Pattern based analysis of BPEL4WS,* Department of Computer and Systems Sciences, Stockholm University/The Royal Institute of Technology, Sweden.

W3C (2000). *SOAP Specification*. Retreived from http://www.w3.org/TR/soap/

W3C (2001). *Web Services Description Language 1.1 specification*. Retrieved from http://www.w3.org/TR/wsdl

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 2, Issue 2, edited by J. Tan, pp. 75-89, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 2.6
A Framework for the Design of Web Service Based Clinical Management Systems to Support Inter and Intra Organizational Patient Journeys

Carolyn McGregor
University of Western Sydney, Australia

ABSTRACT

The clinical management of premature and ill term babies is challenged by the necessity of several inter and intra organizational patient journeys. Premature and ill-term babies born in regional Australia and Canada must be moved to another hospital with Neonatal Intensive Care Unit (NICU) facilities. NICU babies requiring surgery must be moved to a Level IV NICU for surgery. Current clinical management supports the transfer of limited patient data via paper or telephone exchange. In this article a framework for the design of Web-service-based clinical management systems to support inter and intra organizational patient journeys is presented. A series of Web services are described and integrated and coordinated through BPEL processes enabling greater support for inter- and intra-organizational transfer of patient data. This framework is demonstrated through a NICU case study. A key benefit of this framework is that it enables the establishment of "on demand" patient journeys eliminating the need to establish permanent point-to-point connections.

INTRODUCTION AND MOTIVATION

Healthcare and specifically the allocation of funds towards improved provision of healthcare via improved patient journeys represents a significant portion of the Australian, USA, United Kingdom, and Canadian National Budgets (Curry, McGregor, & Tracy, 2006).

Patient Journeys document a pathway of Clinical Management guidelines that represent the patient's journey through treatment and care.

Clinical Management systems are designed to assist care providers in diagnosis and treatment using existing already established methods of diagnosis and accepted treatments (Gross-Portney & Watkins, 2000).

Health care redesign and in particular patient journey redesign (also sometimes referred to as patient pathway or flow) involves analyzing the overall processes involved with the intra- and inter- organizational movement of a patient and then identifying how this journey can be improved via the removal of wasted and excessive activities, process duplication and improved communications between the patient, their carers and the clinicians involved with the journey itself. Process redesign techniques have been used extensively in the business, manufacturing, and computing domains for many years but it is only recently that process improvements in healthcare have been achieved through similar reengineering approaches (Curry, McGregor, & Tracy, 2006).

In general, the research, development and adoption of new information technologies (IT) and information systems (IS) within the healthcare domain is currently lagging behind other industries (McGregor, 2006; Wu, Wang, & Lin, 2005).

Web services enable standardized information exchange through XML messaging. Business Process Execution Language for Web services (BPEL) (Andrews et al., 2003) provides a mechanism to integrate and coordinate the invocation of several Web services within the structure of a business process. Web-service-enabled Patient Journeys have the potential to significantly impact the effective expenditure of these funds by enabling paradigm shifts in Clinical Management.

This article presents case study based research supporting the development of a framework for the design of Web-service-based Clinical Management systems to support intra- and inter-organizational patient journey workflows. The framework enables the modeling of existing patient journeys and supports the reengineering of these patient journeys to improve communication between clinicians and remove duplication, wasted, and excessive activities. Web services are designed that enable greater support for inter- and intra- organizational transfer of patient data. This research builds on existing model-driven Web services integration and development research (Bordbar & Staikopoulos, 2004) by further applying these concepts researched within the business context to the domain of clinical management activities within a patient journey. To facilitate the coordination of the complete inter and intra organizational patient journeys, Business Process Execution Language for Web services (BPEL) is used as BPEL enables the generation of executable workflows built using Web services. A key benefit of this framework is that it enables the establishment of "on demand" patient journeys eliminating the need to establish permanent point to point connections.

The case study used for this research is the patient journey of neonatal and ill-term babies between (1) special care nurseries in urban, rural, or remote hospitals to tertiary or children's hospital Neonatal Intensive Care Units (NICUs) and (2) tertiary NICUs and children's hospital NICUs, when a baby is transferred for surgery.

The article is organized as follows. First a discussion is presented on related recent research in the areas of health informatics to support ICUs together with recent Web services and BPEL research. The next section presents an overview of the framework for the design of Web service based clinical management patient journeys. The following section introduces the Neonatal Intensive Care Unit (NICU) case study environment. Afterwards, the application of the framework to generate BPEL process to support neonatal patient journeys is presented. The article is concluded where future work is presented.

RELATED WORK

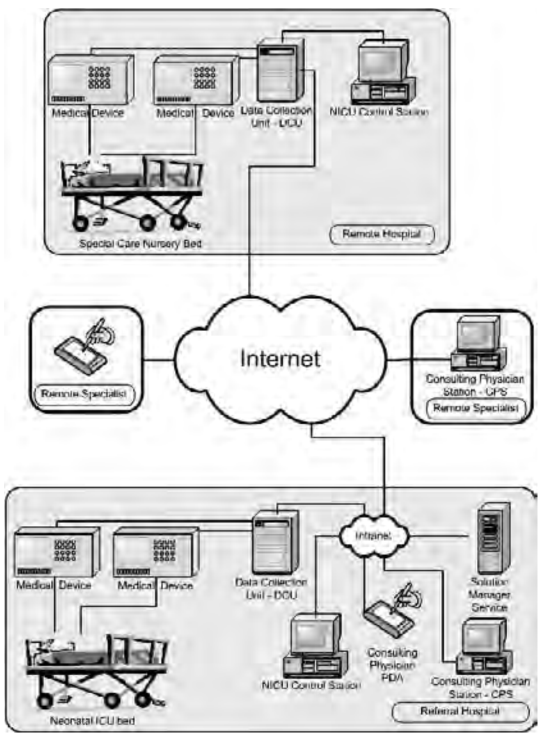
Much of the recent computing and IT related research to support intensive care units (ICUs) has focused on clinical alerts (Catley & Frize, 2003; Catley et al., 2003; Shabot, LoBue, & Chen, 2000; Sukuvaara, Makivirta, Kari, & Koski, 1989; van der Kouwe & Burgess, 2003). The information made available to these systems is limited to a small set of physiological data and/or clinical data from patients located within their ICUs. Clinician access to these systems is limited to the receipt of alerts with minimal content via email and in some cases pagers.

An integrated XML-based health care framework for a neonatal intensive care unit is described by Catley and Frize (2003) and Catley, Frize, Walker, and StGermain (2003). That research enables the capture of local real-time patient device data within a data repository located within a data collection unit (DCU). It utilises XML as a means to transfer data from a NICU data repository to clinical decision support system (CDSS). The transmission of the data from the medical devices to the DCU is not in XML.

Kregar (2001) defines a Web service as *an* interface that describes a collection of operations that are network-accessible through standardised XML messaging. Web Service Definition Language (WSDL) supports the implementation of Web services by providing a standard XML format for describing network services as a set of endpoints operating on messages containing either document-oriented or procedure-oriented information (Christensen, Curbera, Meredith, & Weerawarana, 2001).

Business Process Execution Language for Web Services (BPEL) defines a notation for specifying business process behaviour based on Web services (Andrews et al., 2003). There has been recent focus on linking process modeling with Web services development resulting in model driven Web services integration and development (Bordbar & Staikopoulos, 2004).

Figure 1. e-Baby architecture (McGregor, Heath et al., 2005)



This research is part of the larger "e-Baby" research project (Foster & McGregor, 2006; McGregor, 2005; McGregor, Bryan, Curry, & Tracy, 2002; McGregor, Heath, & Wei, 2005; McGregor, Kneale, & Tracy, 2005, 2006). The broad aims of the e-Baby project are to develop:

- Generic computer interface technologies to allow the integration of data emanating from several disparate medical monitoring systems and a clinical information database.
- Data management technologies to allow for the detection of trends and patterns in clinical and real-time physiological data.
- An Internet-based data distribution mechanism which is capable of integrating data from a number of cooperating data sources while supporting access security controls down to a specific item within a specific individual's medical record.

The e-Baby architecture defines architectural components located at Referral Hospitals containing NICUs, Remote Hospitals with Special Care Nurseries and supports remote and Referral Hospital access by Neonatologists.

The Solution Manager Service (SMS) architecture within the context of its use for business performance measurement is introduced in (McGregor & Schiefer, 2003, 2004; McGregor, Schiefer, & zur Muehlen, 2006; Schiefer & McGregor, 2004). The SMS is an agent-based intelligent decision support system that allows organizations to collect information about their business processes in a centralized repository, and share them among authorized parties, such as supply chain partners, clients, or government agencies. A key contribution of that research is that the interaction with this IDSS is via a set of Web services. In addition, the infrastructure boasts several innovative agent-based modules that support different components, providing the intelligence within this IDSS (McGregor, Schiefer et al., 2006). Within that research business process audit stream data is transmitted to the SMS via the use of a *log Web service*. The e-Baby research reapplies the principles of the SMS within the context of receiving and processing physiological stream data via a *physiological log Web service*.

This article focuses on the framework that enables the coupling of various previously defined e-Baby components (Foster & McGregor, 2006; McGregor, 2005; McGregor et al., 2002; McGregor, Heath et al., 2005; McGregor, Kneale et al., 2005, 2006) using Web-services-based Clinical Management systems to support intra and inter organizational patient journey workflows.

FRAMEWORK FOR THE DESIGN OF WEB SERVICE BASED CLINICAL MANAGEMENT PATIENT JOURNEYS

In this section the framework to enable the establishment of Web-service-based clinical management patient journeys is introduced. The framework builds on previous model driven Web service integration and development research (Bordbar & Staikopoulos, 2004) and proposes the following phases:

1. Model the existing patient journey
2. Reengineer the patient journey
3. Determine inter and intra information flows
4. Web Services Design
5. BPEL Design
6. Development/Testing/Implementation Phase
7. Post Implementation Evaluation

Each phase within the framework is briefly described in the following subsections. An NICU case study is used to demonstrate this framework in the following section.

Existing Patient Journey Model

During this phase information is gathered to enable the construction of patient journey model. Participants may include management, IT support staff, healthcare workers, patients, and their carers. This framework enables the patient journey to be modelled within a chosen patient journey/business process modelling tool. In this article, cross-functional flowcharts have been used to diagrammatically represent the existing patient journey. Further research has commenced on a patient journey modelling and communication tool for use during this phase that better supports the domain of healthcare (Curry, McGregor, & Tracy, 2006). Output from this phase is the documented existing patient journey model.

Reengineer the Patient Journey

The patient journey is then reengineered during this phase. To enable this reengineering focus sessions are run by an experienced facilitator whose role it is to encourage discussion and involvement amongst the participants and translate what is said into a resulting diagrammatic.

Output from this phase is the documented reengineered patient journey model.

Determine Inter and Intra Information Flows

Once the reengineered patient journey has been established, information flows are defined for each communication between the participating organization units. XML schemas are then constructed to represent information flows. The outputs from this phase are the associated XML Schemas.

Web Services Design

For each information flow between organization units that has been defined by an XML schema, Web services are designed to support the intra- and inter-organizational transfer of information. Output from this phase is the design of the Web services.

BPEL Design

During this phase the patient journey model is utilized to provide the information to enable the translation of the patient journey into a Business Process Execution Language for Web Services (BPEL) design. Organizational participants in the patient journey become partners in the BPEL definition. Web Services are invoked at the appropriate place in the patient journey. The output from this phase is the BPEL representation of the patient journey.

Development/Testing/Implementation Phase

The Web services and controlling BPEL are developed, tested, and implemented using the chosen development lifecycle during this phase. The BPEL representation of the BPEL journey and the associated Web services are implemented during this phase.

Post Implementation Evaluation

Once implemented, the patient journey is evaluated. Evaluation within this context is performed via clinical trial to evaluate performance based on the clinical management requirements. In addition, further performance evaluation is performed based on system performance measurement to assess system performance. Output from this evaluation will drive future iterations of this framework.

Figure 2. Neonate or ill-term baby patient journey—remote and referral hospital

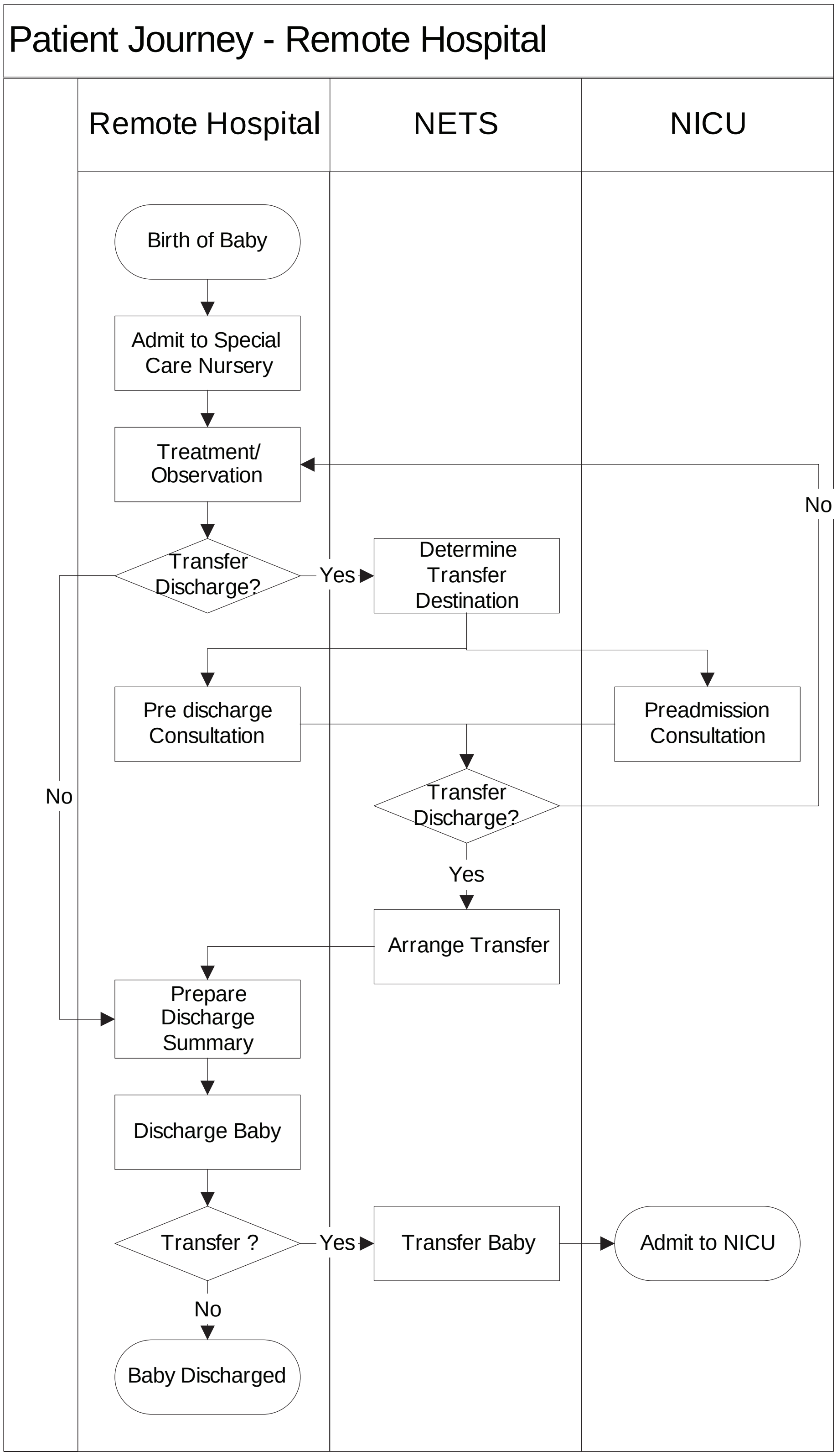
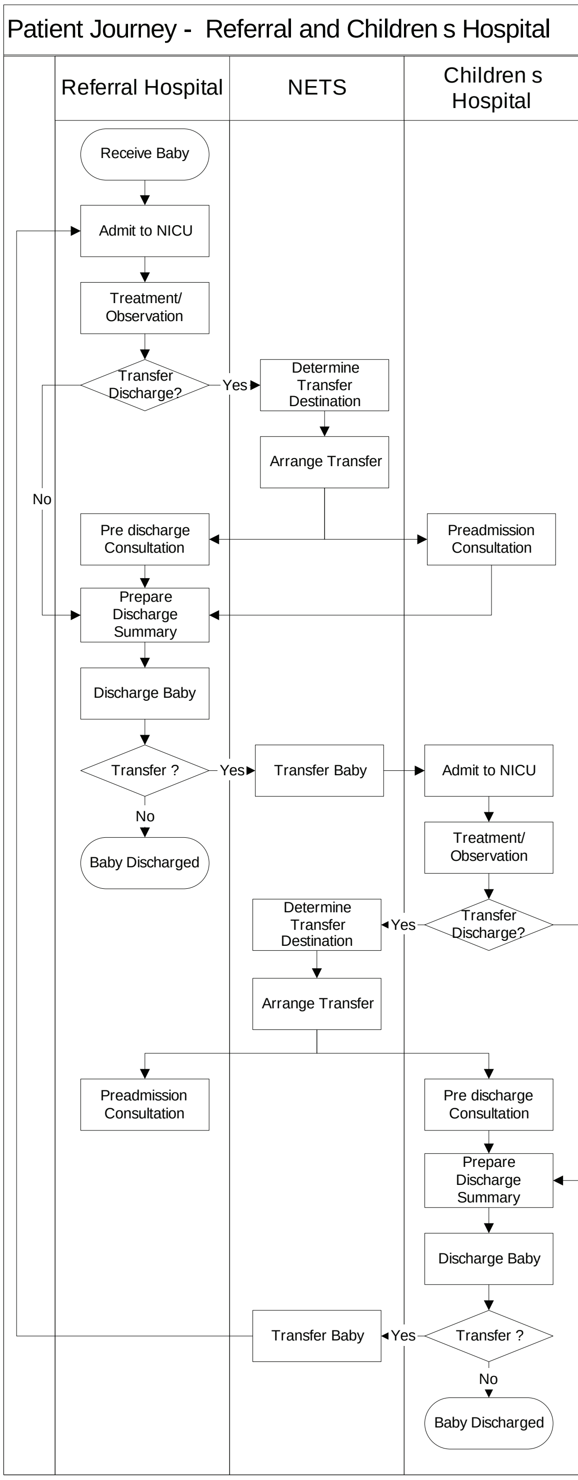


Figure 3. Neonate or ill term baby patient journey—referral and Children's Hospital



NEONATAL INTER AND INTRA ORGANIZATIONAL PATIENT JOURNEYS

The framework to enable the establishment of Web service based clinical management patient journeys is now demonstrated using neonatal inter- and intra-organization patient journeys.

Intensive care units (ICUs) worldwide offer support for patients in need of critical care. They boast a range of state of the art medical monitoring devices to monitor a patient's physiological parameters such as blood oxygen, blood pressure, and heart rate. Other devices such as ventilators offer mechanical life support.

Broadly, there are three types of intensive care namely, adult, pediatric, and neonatal. While the age of the patient is the differentiator between adult and pediatric ICUs, the clinical management differs greatly from these ICUs to the neonatal ICUs (NICUs)—where gestational age greatly impacts clinical management.

From a total of 86,414 babies born in New South Wales (NSW), Australia during 2003, 15 percent (12,296) required admission to a special care ward and 2.6 percent (2,277) were admitted to neonatal intensive care units (NICUs). These figures are similar to previous years (NSW Midwives, 2004). Premature babies can be up to 17 weeks early and may only weigh 450gms; they can spend three or four months in intensive care and have dozens of specific diseases before discharge. Ten percent of severely ill babies will die and 15 percent of survivors will have significant life-long disability. Premature babies, by the time they are discharged, can increase in body mass by as much as six times, and have had dozens of medical diagnoses and treatments. Many of these may have long-term implications for the future health of the individual. In addition, 15 percent of neonatal intensive care admissions are transferred after delivery from smaller regional or remote hospitals without intensive care facilities to larger tertiary referral or children's hospitals with Neonatal Intensive Care Units (NICUs). Similar conditions apply within Australia, New Zealand, Canada, and the USA where small non-tertiary units are spread throughout the country (McGregor, Kneale et al., 2006).

The NSW newborn and paediatric Emergency Transport Service (NETS) is the state wide clinical service for medical retrieval of critically ill newborns, infants, and children in NSW, Australia. Since 1995 NETS neonatal retrievals from regional hospitals to tertiary referral or children's hospitals has averaged 70 per month. There are only three children's hospitals in NSW, (two children's hospitals in Sydney and the John Hunter Children's Hospital in Newcastle). Children's Hospitals contain NICUs and in addition, are able to perform surgery on Neonates. Another seven hospitals are tertiary referral hospitals providing NICUs onsite for newborn babies. All of these hospitals are in major metropolitan areas on the eastern coast of NSW. There are approximately 200 other hospitals in NSW without specialist NICU care that must refer to Neonatologists within the NICUs at either the children's or tertiary referral hospitals for the care of critically ill newborns and infants.

Current treatment practices for premature babies and ill-term babies born in geographically rural and remote hospitals are impacted by the lack of NICU support onsite. Regional hospitals have equipment to provide limited NICU support, but without the ability for a Neonatologist to receive information from this equipment, the baby must be moved to another hospital with Neonatologist support. Given the critical requirement to maintain a consistent environment, moving a baby at this time can be life threatening. Transferred pre-term and critically-ill-term babies have higher mortality rates and much higher rates of long-term disability than similar babies born in hospitals with NICU facilities (McGregor et al., 2002). A major limitation is that the attending physician must contact a Neonatologist via telephone, who may or may not be located at the NICU at that time, to describe the baby's condition and, where possible, relay any physiological information verbally. The consulting Neonatologist must then make decisions based on this verbal exchange.

It is very common for critically ill babies to have significantly abnormal variation in the measured parameters minute by minute and not all of these variations are made available to the consulting Neonatologist. Frequent transient falls in blood pressure and blood oxygen content, often with swings into the high range, may be of critical importance in survival and quality of survival free of significant disability (Lister, Bryan, & Tracy, 2000).

Adding to the complexity of this neonatal support function, the NICUs located at the children's hospitals and tertiary referral hospitals each have a limited number of humidicribs. As a result, the remote hospital attending physician must contact NETS who utilise their referral database to determine which children's or tertiary referral hospital NICUs the remote hospital should contact for support and/or transfer, based on humidicrib availability. Thus the NICU patient journey support that the approximate 200 regional and tertiary referral metropolitan hospitals require is an adhoc connection to any of the Children's or Tertiary Referral hospital NICUs (10) for which they are referred to by NETS, on a needs basis, at a total rate of 70 per month (McGregor, Kneale et al., 2006).

Typical medical records for the most premature babies consist of manually prepared paper notes that detail enormous data collection and can weigh up to three to four kilograms. The preparation of these notes is dependent on the hand-annotated records of nursing staff, usually at 60-minute intervals, and 30-minute intervals at best. Also, within the NICUs, the collection of the information contained within the baby's notes to form a discharge letter represents a manually intensive analysis of physiological statistics, together with the replication of clinical data that can take the Neonatologist many hours to produce.

The following subsections present and describe the use of the framework proposed within this research on two high-level patient journeys namely: neonate or ill-term baby patient journey—remote hospital; and neonate or ill-term baby patient journey—referral hospital. The example commences with the second phase where the improved patient journey model has been developed and concludes at the BPEL Design.

Reengineer the Patient Journey

Neonate or Ill Term Baby Patient Journey—Remote Hospital

The neonatal or ill-term baby patient journey where the baby is born in a remote hospital is summarised in Figure 2. A neonate or ill-term baby born within a *Remote Hospital* firstly goes through the activity of *Admit to Special Care Nursery* (ie. placed within a humidicrib in the Special Care Nursery). They then begin to undergo *Treatment/Observation*. During the *Treatment/Observation,* the attending paediatrician will constantly access the baby's condition. While the care of neonates on a daily basis is quite complex and requires in simple terms nutrition planning, observation, and treatment, of relevance to this research is the point in the treatment/observation activity determining whether the baby can be discharged, or whether a discharge via transfer to a NICU may be required. If a transfer may be required, they will contact NETS to determine the transfer destination. Once NETS has allocated a humidicrib within a receiving NICU, the pediatrician at the remote hospital will contact the receiving on-call neonatologist from the NICU (who may or may not be located within the NICU at that time) for a predischarge/preadmission consultation and to determine whether a transfer discharge is required. If they decide not to transfer the baby the paediatrician continues the treatment/observation activity. If they decide a transfer discharge is required, NETS is contacted again to arrange the transfer. The pediatrician will then prepare the discharge summary. If the discharge is a transfer, NETS perform the activity of transferring the baby.

Neonate or Ill Term Baby Patient Journey—Referral Hospital

The neonatal or ill-term baby patient journey for a baby admitted into a tertiary hospital's NICU is summarised in Figure 3. Once admitted to the *NICU*, they then, similarly to that defined earlier, begin to undergo the complex Treatment/Observation activity. Of relevance to this research is the point(s) during the treatment/observation activity that the attending neonatologist will assess whether the baby can be discharged, or whether surgery may be required. If surgery may be required, then a transfer will be required. They will contact NETS to determine the transfer destination to confirm which children's hospital the baby will be transferred to. Once NETS has allocated a humidicrib within a receiving NICU, the Neonatologist at the tertiary hospital will contact the receiving on-call neonatologist from the children's hospital NICU (who may or may not be located within the NICU at that time) for a predischarge/preadmission consultation prior to surgery and to arrange the transfer. The neonatologist will then prepare the discharge summary. If the discharge is a transfer, NETS perform the activity of transferring the baby. Within the children's hospital, the baby undergoes the same patient journey as that of the referral hospital. The surgery is performed during the treatment/observation. Once the baby is stable post surgery, NETS are contacted to arrange the return of the baby to the tertiary hospital NICU and a similar transfer process proceeds.

Determine Inter and Intra Information Flows

As can be seen from Figure 2 and Figure 3, interorganization information flows are required to support the following activities:

1. *Determine Transfer Destination*
2. *Predischarge/Preadmission Consultation*
3. *Transfer Baby*

Determine Transfer Destination

The pediatriciation/neonatologist require the ability to contact NETS to determine where a humidicrib is available to receive the transferred baby.

Predischarge/Preadmission Consultation

The neonatologists located at both the children's hospitals and tertiary referral hospitals require the ability to obtain information from the monitors attached to babies that are not physically located at that specific hospital, but perhaps at another regional hospital where only limited support for premature babies exists. Similarly, a Neonatologist need not be located within the hospital to view patient data, but should be free to view this information through any secure Internet/Intranet connection.

Transfer Baby

Remote hospitals require the ability to send discharge information through to the receiving NICU in a standardised format to enable the receiving NICU to integrate the information directly within the receiving hospitals database.

In addition, there is a need to extract information from NICU databases to automate the preparation of discharge information to be sent to the receiving NICU.

Finally, the pediatriciation/neonatologist requires the ability to contact NETS to confirm the execution of the transfer.

Web Services Design

To support the information flows required as detailed in the previous section, five Web services are required to satisfy each of the listed information flow functions namely the:

1. Availability Web Service established by NETS
2. Physiological Log Web Service established by each NICU
3. Allocation Web Service established by NETS
4. Clinical Web Service established by each NICU
5. Analyse Web Service established by each NICU

Each Web service is detailed briefly below.

Availability Web Service

A database of all receiving neonatologists, together with their hospital's details are contained within the Referral Database. This information provides details based on humidicrib availability. The Referral Database is located within NETS and accessible via a set of Web services. The *Availability Web Service* forwards details of the receiving Neonatologist together with the details for the receiving Neonatologist's *Physiological Web Service*. Further details on Web services by NETS are contained in (McGregor, Kneale et al., 2005, 2006).

Physiological Web Service

To address the need for standardising physiological data transmission, a Web service based framework for the transmission, display, and storage of physiological data for local and remote neonatal intensive care has been established known as the *physiological Web service*. The Web service enables transmission of time interval packets of physiological data through the *physiological Web service* by sending *Physiological Log Packets* and receiving *Acknowledgements*. Further details on the *physiological Web service* are contained in (McGregor, Heath et al., 2005; McGregor, Purdy, & Kneale, 2005).

Allocation Web Service

This Web service enables NETS to be advised of the execution of the transfer.

Clinical Web Service

This Web service enables the receiving NICU to advertise the standardised format to receive discharge information from other hospitals

Analyse Web Service

This Web service enables NICUs to extract discharge summary information from their systems to send to receiving NICUs via their Clinical Web Service.

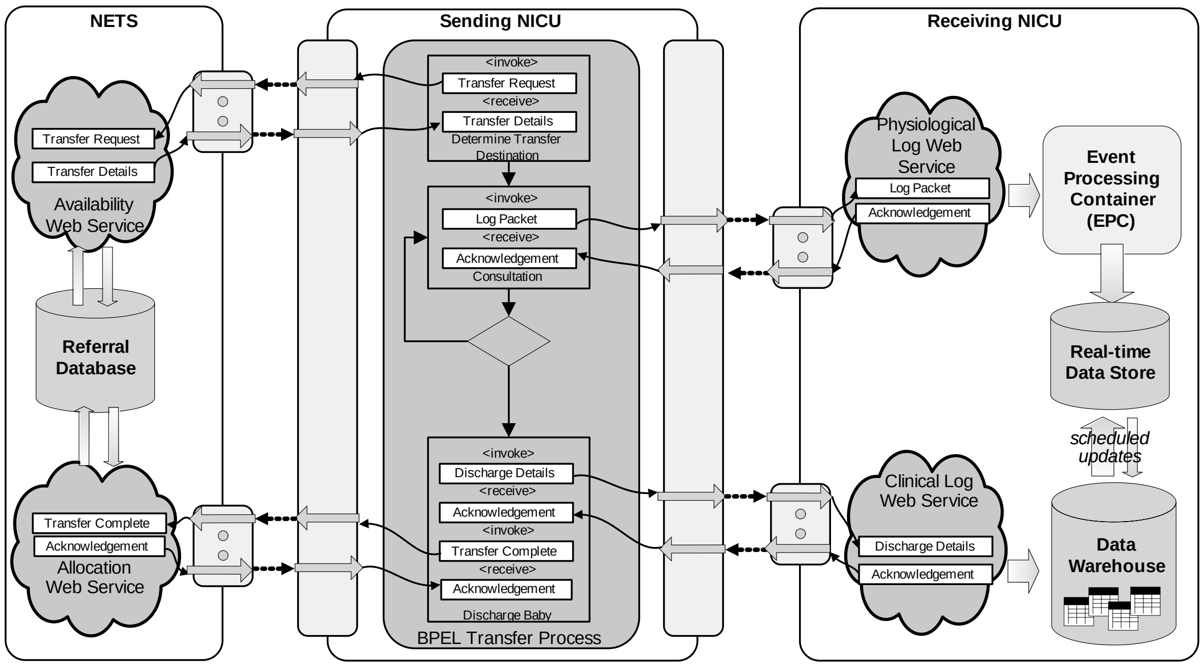
BPEL Design

The BPEL design of the patient journey as detailed previously are now presented in the following two subsections. The Web services introduced in the previous section are invoked at the appropriate place in the patient journey.

BPEL Transfer Patient Journey—Remote Hospital

The BPEL Transfer Patient Journey from a remote hospital is summarized in Figure 4 highlighting calls to the *Availability Web Service*, *Physiological Web Service*, *Clinical Web Service* and finally the *Allocation Web Service*.

Figure 4. BPEL Transfer Process—Remote to NICU



BPEL Transfer Patient Journey—NICU to NICU

The BPEL Transfer Patient Journey between NICUs is summarized in Figures 4 and 5, highlighting calls to the *Availability Web Service*, *Physiological Web Service*, *Clinical Web Service,* and finally the *Allocation Web Service*. In addition during the Discharge Baby activity a local call to the local *Analyse Web Service* is made to gather the information to be forwarded via the *Clinical Log Web Service*.

PATIENT JOURNEY PROTOTYPES

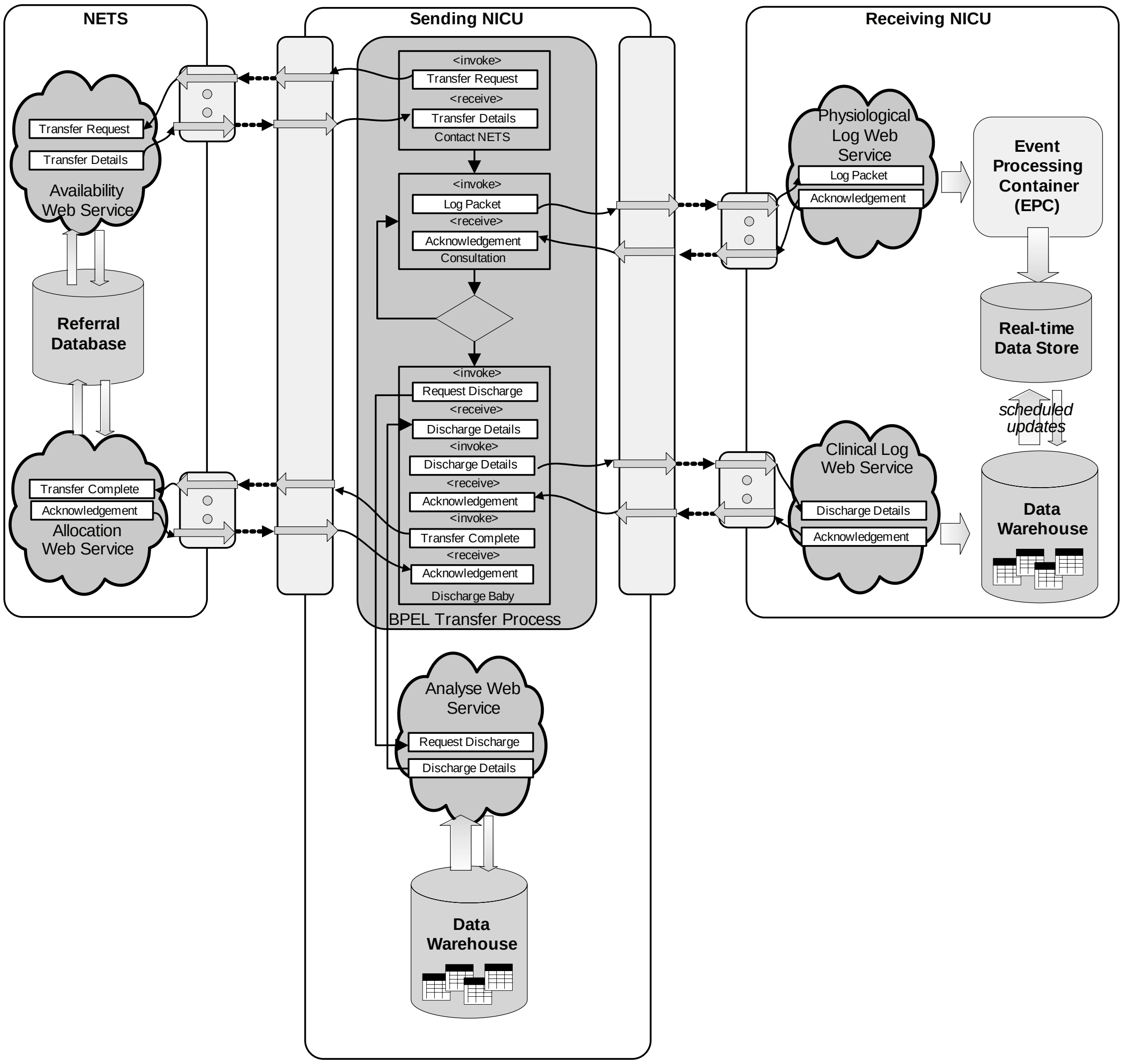
Several of the Web service components within this research have been or are currently being prototyped within or in association with the Nepean Hospital NICU.

The *physiological log Web service* and Data Warehouse components of the Solution Manager Service have been demonstrated via a pilot development using physiological data supplied by a Dragar Ventilator and the Phillips Component Monitoring System (CMS) monitor (McGregor, Heath et al., 2005). Initial testing of compression of XML encoded physiological data streams to improve Web service performance is detailed in (McGregor, Purdy et al., 2005).

The NETS *availability Web service* and *Referral Database* has been developed and are currently being prototyped as part of the Bush Babies on Broadband component of the e-Baby collaboration (McGregor, Kneale et al., 2005, 2006).

XML encoding of the discharge summary has been completed and further testing of the generation of this XML coded discharge summary will form part of our future research.

Figure 5. BPEL transfer process—NICU to NICU



CONCLUSION AND FUTURE WORK

This article has presented case study based research supporting the development of a framework for the design of Web-service-based Clinical Management systems to support intra- and inter-organizational patient journey workflows. The framework enables the modeling of existing patient journeys and supports the reengineering of these patient journeys to improve communication between clinicians and remove duplication, wasted and excessive activities. Web Services are designed that enable greater support for inter- and intra-organizational transfer of patient data. To facilitate the coordination of the complete inter and intra organizational patient journeys, Business Process Execution Language for Web services (BPEL) is used. A key benefit of this framework is that it enables the establishment of "on demand" patient journeys eliminating the need to establish permanent point to point connections.

This article further describes the application of that architecture to a specific pilot in association with the NICU at Nepean Hospital, Penrith Australia where several of the Web services have been or are currently being tested and trialed. The set of Web services described here have been found to enable mobility, interoperability, alert delivery, and to improve data capture for later research and analysis. While the paradigm shift to using Web services within BPEL processes to support NICUs offers the potential to significantly improve the speed, efficiency, and effectiveness of critical care within NICUs, there are currently several factors impacting its mainstream adoption. These include issues relating to the use of wireless networks in addition to security, privacy, cost, and a lack of standards for physiological data capture and exchange and user interface design. In addition, management issues associated with inter-organization collaboration impact embracing such approaches.

While this research has used cross-functional flowcharts to model the patient journey, research has commenced on an alternate patient journey modeling/communication tool to more richly describe patient journeys (Curry, McGregor, & Tracy, 2006).

Data security and privacy are significant issues within the healthcare domain. Future research has commenced to incorporate privacy policy information and security models during all phases of this framework.

ACKNOWLEDGMENTS

This research is supported by an Australian Research Council (ARC) Linkage Project Grant (LP0561518). In addition, the researchers wish to thank the staff of the Neonatal Intensive Care Unit, Nepean Hospital, Sydney West Area Health Services, NSW, for their continued support and collaboration in this research.

REFERENCES

Andrews, T., Curbera, F., Dholakia, H., Goland, Y., Klein, J., Leymann, F., et al. (2003). *Business process execution language for Web services, version 1.1*: OASIS.

Bordbar, B., & Staikopoulos, A. (2004). *On behavioural model transformation in Web services.* Paper presented at the Conceptual Modelling for Advanced Application Domain (eCOMO), China.

Catley, C., & Frize, M. (2003). *A prototype XML-based implementation of an integrated 'intelligent' neonatal intensive care unit.* Paper presented at the 4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine, 322-325.

Catley, C., Frize, M., Walker, C. R., & StGermain, L. (2003). *Integrating clinical alerts into an XML-based health care framework for the neonatal intensive care unit.* Paper presented at the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2003, 1276-1279, Vol.1272.

Christensen, E., Curbera, F., Meredith, G., & Weerawarana, S. (2001, 15/3/2001). *Web Services Description Language (WSDL) 1.1 W3C Note.* Retrieved March 15, 2001 from http://www.w3.org/TR/wsdl

Curry, J., McGregor, C., & Tracy, S. (2006, August 30 – September, 3). *A communication tool to improve the patient journey modeling process.* Paper presented at the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC06), New York, 4726-4730.

Foster, D., & McGregor, C. (2006). *Design of an agent server for neonatal analysis and trend detection.* Paper presented at the Workshop on Multi-Agent Systems for Medicine, Computational Biology and Bioinformatics (MAS*BIOMED), Hakodate, Japan, 47-59.

Gross-Portney, L., & Watkins, M. (2000). *Foundations of clinical research, applications to practice* (2nd Edition ed.): Prentice Hall.

Kregar, H. (2001). Web Services Conceptual Architecture (WSCA 1.0): IBM Software Group.

Lister, R., Bryan, G., & Tracy, M. (2000). *The e-babies project: Integrated data monitoring and decision making in neonatal intensive care.* Paper presented at the European Conference of Information Systems (ECIS 2000).

McGregor, C. (2005). *e-baby Web services to support local and remote neonatal intensive care.* Paper presented at the Health Informatics Conference, Melbourne, Australia, CD-ROM, 6 pages.

McGregor, C. (2006). Mobility in healthcare for remote intensive care unit clinical management. In B. Unhelkar (Ed.). *Handbook of research in mobile business: Technical, methodological and social perspective,* (pp. 83-95). Hershey, PA: Information Science Reference.

McGregor, C., Bryan, G., Curry, J., & Tracy, M. (2002). *The e-baby data warehouse: A case study.* Paper presented at the 35th Annual Hawaii International Conference on System Sciences (HICSS), 3018-3024.

McGregor, C., Heath, J., & Wei, M. (2005). *A Web service based framework for the transmission of physiological data for local and remote neonatal intensive care.* Paper presented at the IEEE International Conference on e-Technology, e-Commerce and e-Service, Hong Kong, 496-501.

McGregor, C., Kneale, B., & Tracy, M. (2005). *Bush babies broadband: On-demand virtual neonatal intensive care unit support for regional Australia.* Paper presented at the 3rd International Conference on Information Technology and Applications (ICITA), Sydney, Australia, 113-117.

McGregor, C., Kneale, B., & Tracy, M. (2006). On-demand virtual neonatal intensive care units supporting rural, remote and urban healthcare with bush babies broadband. *Journal of Network and Computer Applications*.

McGregor, C., Purdy, M., & Kneale, B. (2005). *Compression of XML physiological data streams to support neonatal intensive care unit Web services.* Paper presented at the IEEE International Conference on e-Technology, e-Commerce and e-Service, Hong Kong, 486-489.

McGregor, C., & Schiefer, J. (2003). *A framework for analyzing and measuring business performance with Web services.* Paper presented at the IEEE International Conference on E-Commerce, 2003. CEC 2003, 405-412.

McGregor, C., & Schiefer, J. (2004). A Web-Service based framework for analyzing and measuring business performance. *Information Systems and e-Business Management, 2*(1), 89-110.

McGregor, C., Schiefer, J., & zur Muehlen, M. (2006). A shareable Web service based idss infrastructure for on-demand business process management. *International Journal of Business Process Integration and Management (IJBPIM), 1*(3), 156-174.

NSW Midwives, a. A. (2004). NSW mothers and babies. *NSW Public Health Bulletin, 15*(S-5), 26.

Schiefer, J., & McGregor, C. (2004, April 14-17, 2004). *Correlating events for monitoring business processes.* Paper presented at the 6th International Conference on Enterprise Information Systems, Porto, Portugal, 320-327.

Shabot, M. M., LoBue, M., & Chen, J. (2000). *Wireless clinical alerts for critical medication, laboratory and physiologic data.* Paper presented at the 33rd Annual Hawaii International Conference on System Sciences, 2000. p. 6, Vol.1.

Sukuvaara, T., Makivirta, A., Kari, A., & Koski, E. (1989). *An intelligent intensive care alarming system.* Paper presented at the Computers in Cardiology 1989. Proceedings (pp. 225-228).

van der Kouwe, A. J. W., & Burgess, R. C. (2003). Neurointensive care unit system for continuous electrophysiological monitoring with remote Web-based review. *Information Technology in Biomedicine, IEEE Transactions on, 7*(2), 130-140.

Wu, J.-H., Wang, S.-C., & Lin, L.-M. (2005). *What drives mobile health care? An empirical evaluation of technology acceptance.* Paper presented at the 38th Annual Hawaii International Conference on System Sciences, 2005. HICSS '05., 150a.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 2, Issue 2, edited by J. Tan, pp. 21-35, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 2.7
Conceptual Framework for Mobile-Based Application in Healthcare

Matthew W. Guah
Erasmus University Rotterdam, The Netherlands

ABSTRACT

The significance of aligning IT with corporate strategy is widely recognised, but the lack of an appropriate framework often prevents practitioners from integrating emerging Internet technologies (like Web services and mobile technologies) within organisations' strategies effectively. This chapter introduces a framework that addresses the issue of deploying Web services strategically within a mobile-based healthcare setting. A framework is developed to match potential benefits of Web services with corporate strategy in four business dimensions: innovation, internal healthcare process, patients' pathway, and management of the healthcare institution. The author argues that the strategic benefits of implementing Web services in a healthcare organisation can only be realized if the Web-services initiatives are planned and implemented within the framework of an IT strategy that is designed to support the business strategy of that healthcare organisation.The chapter will use case studies to answer several questions relating to wireless and mobile technologies and how they offer vast opportunity to enhance Web services. It also investigates what challenges are faced if this solution is to be delivered successfully in healthcare. The healthcare industry globally, with specific emphasis on the USA and United Kingdom, has been extremely slow in adopting emerging technologies that focus on better practice management and administrative needs. The chapter elaborates on certain emerging information technologies that are currently available to aid the smooth process of implementing mobile-based technologies into healthcare industry.

INTRODUCTION

This chapter is based on research—using a longitudinal case study—into the National Programme for Information Technology (NPfIT). NPfIT is an initiative that has been budgeted to cost the UK government £6.3 billion for the purpose of improving the information systems in the National Health Service (NHS), with emphasis on IT infrastructure and the creation of a nationwide patient database.

The significance of aligning IT with corporate strategy in healthcare organisations is widely recognised, but the lack of an appropriate framework often prevents medical practitioners from integrating emerging Internet technologies (like Web services and mobile technologies) within healthcare organisations' strategies effectively. This chapter introduces a framework that addresses the issue of deploying Web services strategically within a mobile-based healthcare setting. A framework is developed to match potential benefits of Web services with corporate strategy in four business dimensions: innovation, internal healthcare process, patients' pathway, and the management of healthcare institution. The author argues that the strategic benefits of implementing Web services in a healthcare organisation can only be realized if the Web-services initiatives are planned and implemented within the framework of an IT strategy that is designed to support the business strategy of that healthcare organisation.

The chapter will also consider certain essential issues regarding the deployment of any mobile data solution (i.e., reliability, efficiency, and security) in the healthcare industry and how such deployment can support healthcare professionals in saving patients' lives. Using case studies, the chapter will answer the following questions:

- Wireless and mobile technologies offer vast opportunities to enhance services, but what challenges are faced if this solution is to be delivered successfully in healthcare?
- Why has the global healthcare industry, with specific emphasis on the USA and United Kingdom, been extremely slow in adopting technologies that focus on better practice management and administrative needs?
- How complacent can IS strategists be to the productivity paradox in the wake of HIPAA (Health Insurance Portability and Accountability Act in USA) and NPfIT (in UK)?
- What emerging information technologies are there to aid the smooth process of implementing mobile-based technologies into the healthcare industry?

The existing economics and IS literature on information-technology adoption often considers network externalities as one of the main factors that affect adoption decisions (Brown & Venkatesh, 2003). It is generally assumed that potential adopters achieve a certain level of expectations about network externalities when they have to decide whether to adopt a particular technology. However, there has been little discussion on how the potential adopters reach their expectations. This chapter attempts to fill a gap in the literature on the adoption of mobile healthcare technology by offering an optimal control perspective motivated by the rational expectations hypothesis and exploring the process dynamics associated with the actions of decision makers in the healthcare industry. They must adjust their expectations about the benefits of a mobile healthcare technology over time due to bounded rationality. The model posed in this chapter addresses mobile healthcare technologies that exhibit strong network externalities. It stresses adaptive learning to show why different healthcare organisations that initially have heterogeneous expectations about the potential value of a mobile healthcare technology eventually are able to arrive at contemporaneous decisions to adopt the same technology, creating the desired network externalities. This further allows these organisations to become catalysts to facilitate processes that lead to healthcare industry-wide adoption.

BACKGROUND

The NHS has been responsible for the provision of healthcare and services in the United Kingdom for the past 56 years on the basis of being free for all at the point of delivery. The traditional perception of the NHS is one of a healthcare system organised as a professional guild, with unlimited finance from the government. This type of NHS is experiencing an irrevocable change as taxpayers are no longer complaisant and paternalistic employers are reacting against inflating costs and escalating complaints from the patients. The employer is reacting to the continuous massive flow of subsidies for inefficient physician practices, fragmented delivery systems, and cost-unconscious consumer demand. The patients are increasingly assertive as to their preferences and few have expressed their willingness to make additional contributions for particular health benefits and medical interventions.

Web services are technologies with roots in the Application Service Provision (ASP) business model that are used mostly to automate linkages among applications (Hagel, 2002). They are generally anticipated to make critical system connections not only possible but also easy and cheap (Kreger, 2003; Sleeper & Robins, 2001). One of the perceived benefits of Web services is that organisations would be able to concentrate on their core competencies (Perseid Software Limited, 2003). Service providers argued that the remote delivery of software applications would release managers from the perennial problems of running in-house IT departments, allowing more time to develop IT and e-business strategy rather than the day-to-day operations (Currie, Desai, & Kahn, 2004). This justification has been used in traditional forms of outsourcing over many years (Willcocks & Lacity, 1998).

The NHS is experiencing massive changes in the structure of information systems provision markets and organisations. The local service provision (LSP) and national service provision (NSP) models in use by the NPfIT are in a state of ferment. The payment methods borrow from both capitation and fee for service, and methods of utilisation management are compromised between arm's-length review and full delegation (Guah & Currie, 2006). LSP and NSP consist of large and more complex entities. These are the result of merger, acquisition, and product diversification. The service providers involved have had to take on a visible feature of ceaseless acquisition and divestiture, integration and outsourcing, and combination and recombination. Providers of medical systems, hospital administration systems, and health plans are coming together and then coming apart. They are substituting contracts for joint ownership, creating diversified conglomerates and refocused facilities, and experimenting with ever-new structures of ownership, finance, governance, and management (Robinson, 2000). These would give the NHSIA (National Health Service Information Authority) the benefits not only of a middle ground between the extremes of vertical integration and spot contracting, but also a balance of coordinated and autonomous adaptation in the face of its ever-new challenges.

The general assumption is that expenditures in the nation's health will outpace the overall growth in the economy (Collins, 2003; Pencheon, 1998). This is reflected in the percentage of the GDP (gross domestic product) of USA (13%), Germany (10.7%), France (9.6%), and the United Kingdom (7.6%) being devoted to the total cost of healthcare resources (Brown, 2002). Unlike the United Kingdom, however, some of these countries are faced with limitations in social willingness to pay. It has been documented that millions of U.S. residents currently lack the most basic insurance coverage (Institute of Medicine, 2002).

Response to Emerging Technologies in the NHS

Over the years, nontechnologists in the NHS have managed to muddle through one powerful new system after another. Generational strategy is one continuously being used to deal with some of the pressures induced by IS. Adopting such innovations as PCs (personal computers) and the Internet requires the personal and organisational costs of unfreezing deeply ingrained old habits. Many workforces ignore, deny, or deal awkwardly with such technologies.

Srinivasan, Lilien, and Rangaswamy (2004) found several reasons why an organisation should respond to new technology development. Two major reasons are listed below:

- Technological change is a principal driver of competition. This is principally because it destroys monopolies, creates new industries, and renders products and markets obsolete.
- Additional sources—both within and outside the organisation or industry—are increasingly complementing in-house technology development efforts.

A common response to new systems is the "not invented here" (NIH) syndrome (Collins, 2003; Guah & Currie, 2004; Haines, 2002). This often leads to certain organisations rejecting a perfectly useful system based on an implicit assumption that the system does not fully recognise or accommodate their own needs and idiosyncrasies (Brown & Venkatesh, 2003; Davis, 1989). Davis sees this as a likely result of a decline in communication with external sources. NIH syndrome could also result from competences that can be proven to be outdated and inefficient in comparison to an existing technology. One Trust, which places a central role in the direction of regional IS strategy, had to reject a system promoted by the Department of Health because the system was not as familiar as another bespoke system (Haines).

The common characteristics of new systems in the NHS are uniformity in products and prices in the face of great variability in consumer preferences and the actual costs of providing service (Collins, 2003). This one-size-fits-all approach usually leads to services that are of excessive costs for some users and insufficient quality for others, impeding the use of price flexibility to enhance capacity utilisation (Robinson, 2000). Also of concern is a combination of overcapacity and low load factors in some regional trusts with undercapacity and shortages elsewhere. Concerns are growing in the NHS that this may generate cross-subsidies from trusts for which the cost of service will be low to trusts for which the cost of service will be high (McGauran, 2002). Additionally, deregulation of healthcare costs has spurred an outpouring of new services. Consequently, several of these services are the following (Collins; Pencheon, 1998):

- A different cost structure.
- An impact on IS budgets.
- A better match between supply and demand.

Incomplete information has been a fascinating attribute of the NHS's unusual system's organisational and normative characteristics. The asymmetry of NHS information between patients and medical practitioners has changed in an exogenous fashion over its 56 years. The amount of healthcare information available to patients is usually the result rather than the cause of changes in the economic and political environment (Robinson, 2000).

PROJECT DESCRIPTION: NATIONAL PROGRAMME FOR INFORMATION TECHNOLOGY

The NPfIT is an initiative by the National Health Service Information Authority, born as a result of several plans to devise a workable IS strategy for the NHS (NHSIA, 2003; Wanless, 2002). The NPfIT was designed to connect the capabilities of modern IT to the delivery of the NHS plan devised in 1998. The core of this strategy is to take greater control of the specification, procurement, resource management, performance management, and delivery of the information and IT agenda (NHSIA).

The NPfIT is an essential element in delivering the NHS plan. It has created £6 billion information infrastructure, which could improve patient care by increasing the efficiency and effectiveness of clinicians and other NHS staff. The intention of the plan is to address the following (http://www.npfit.nhs.uk):

- Create an NHS Care Records Service to improve the sharing of consenting patients' records across the NHS.
- Make it easier and faster for GPs (general practitioners) and other primary care staff to book hospital appointments for patients.
- Provide a system for electronic transmission of prescriptions.
- Ensure that the IT infrastructure can meet NHS needs now and in the future.

The decision to implement a national programme for IT into the NHS system complexity is only the first step in the IS modernisation journey for a multifaceted organisation. There are many examples of new technologies disrupting organisational routines and relationships in the NHS (Atkinson & Peel, 1998; Majeed, 2003; Metters, Abrams, Greenfield, Parmar, & Venn, 1997). These usually require both medical professionals and NHS regional trusts managers to relearn how to work together. Orlikowski (1993) and Edmondson (2003) suggest that one technology can be seen differently by two groups of people in an organisation. Findings from Barney and Griffin (1992) and Orlikowski have showed how this could result in the elicitation of different responses for members of that organisation.

Scope of Project Work

The chapter takes a more in-depth look at the role of the NHSIA (seen as the project leader for the NPfIT), currently the most visible spokesperson and translator for the potential implications of the resulting new technologies. Research has shown NHS IS staff to pay particular attention to what the NHSIA says and does in regard to information systems (Collins, 2003; Ferlie & Shortell, 2001). This research builds on a framework that identifies the key dimensions of the NHSIA tactic that is situation specific for NPfIT assumptions. The work looks at the NHSIA goal and roles for the NPfIT, as well as the role of the private-sector service providers in the implementation of NPfIT.

Here are a few objectives the NPfIT hopes to accomplish:

- To have a series of tightly specified and priced framework contracts on a short list (of about five) primary service providers (PSPs) who can work at the regional and local Strategic Health Authority (StHA) level. This should enforce the integration and implementation partnership—at a national level—during all aspects of the NPfIT project. Each PSP will have an aligned consortium of service providers and vendors for the integrated care resource service element of the NPfIT, and will be mandated to work with the domain PSP for electronic booking, the infrastructure providers, and healthcare providers. StHA PSPs may not make their products exclusive or mandatory to their StHA.
- To create priced packages of national services and applications that the PSPs and StHAs can together implement locally. This activity will include managing the creation of a single Human Resource Information Systems (HRIS) and other national services to access and move health-record information as required.
- To create service-level agreements for the national services and other services out of the scope of the PSP consortium that the PSPs can work toward in providing an integrated service to the StHA
- To develop and maintain the national standards and specifications that all vendors must use. It is also anticipated to create the national business cases required for the

Department of Health governance (required by the National Treasury), and to support the local decision-making business cases required at the StHA level.

- To procure, under national contracts, a backbone network infrastructure

Such an arrangement provides the greatest clarity in respect to the appropriate allocation of responsibilities and should be well understood in the public and private sectors (see Table 1). Services will be procured on a long-term basis so the combination of local and central funding will be required for at least 5, and preferably 10, years at guaranteed levels.

THE RESEARCH STUDY

This study intended to address the gap in the existing literature with regard to the complex issues surrounding the adoption of mobile technology in healthcare. I define mobile healthcare as the use of all kinds of wireless devices (cell phones, personal digital assistants, mobile e-mail devices, handheld computers, etc.) to provide health information and patient-care records to healthcare practitioners, patients, and their caregivers, employers, and employees of health service providers and public regulars of healthcare and services.

Table 1. PSP implementation timetable (as of July 2002)

Activity/Output	***Target Date***
Agree on procurement strategy (Department of Health (DoH) & local health authorities)	End Jul 2002
Service requirement finalized and approved	End Sep 2002
Outline business case developed and approved	End Sep 2002
Official Journal of European Communities (OJEC) advert	Oct 2002
Procurement of systems and implementation services for electronic booking begins	Oct 2002
National long list of PSPs created	Dec 2002
Invitation to negotiate issued	Jan 2003
National short list of PSPs created	Apr 2003
First local health authorities begin detailed planning with PSPs	Aug 2003
PSP framework contract finalized	Oct 2003
Infrastructure provider(s) contract agreed	Oct 2003
First local health authorities begin implementation	Nov 2003
Infrastructure migration begins	Mar 2004

The findings reported in this chapter are part of a larger 5-year research study that was developed to investigate the deployment, hosting, and integration of the ASP and Web-services technologies from both a supply-side and demand-side perspective. The overall research was in two phases. The first phase, comprising of a pilot study, was conducted in the USA and United Kingdom (Currie et al., 2004). An exploratory-descriptive case-study methodology (Yin, 1994) was used to investigate 28 ASP vendors and seven customer sites in the United Kingdom. The dual focus upon supply side and demand side was critical for obtaining a balanced view between vendor aspirations about the value of their business models and customer experiences, which may suggest a less optimistic picture. The unit of analysis was the business model (Amit & Zott, 2001), not the firm or industry level, so a case-study methodology was anticipated to provide a rich data set for analysing firm activities and behaviour (Currie et al.).

The result from the pilot study led to the funding of two additional research studies by the Engineering and Physical Sciences Research Council (EPSRC) and Economic and Social Research Council (ESRC) respectively. Industrial collaborators were selected for the roles of technology partners, service providers, and potential or existing customers. These studies were concerned with identifying sources of value creation from the ASP business model and Web-service technologies in different vertical sectors (including health).

Research Methodology

The research followed a number of stages involving the use of both qualitative and quantitative data-collection techniques and approaches (Walsham, 1993). A questionnaire survey was distributed by e-mail to businesses and healthcare organisations all over the United Kingdom. These organisations were listed on a national database maintained by the NHSIA, plus those maintained by the university. To ensure relevant managers and practitioners responded, the covering note clearly stated the purpose of the questionnaire and requested that it be passed on to the person(s) with responsibility for managing healthcare e-business strategy. Scales to address the research questions were not available from the literature, so the questionnaire was developed based on the theory of strategic value (Banker & Kauffman, 1988). It included a checklist, open-ended questions, and a section seeking organisational data. Research questions under Part 1 required respondents to answer yes or no if the application of Internet technologies in healthcare were bringing value to patients. Data in Part 2 of the questionnaire were collected by open-ended questions seeking respondents' views on the best approach to healthcare performance improvement and Web-service value creation. This line of questioning was used to increase the reliability of data since all respondents were asked the same questions, but some added additional information. The purpose was to impose uniformity across the sample of representation rather than to replicate the data obtained from each participant (Yin, 1994).

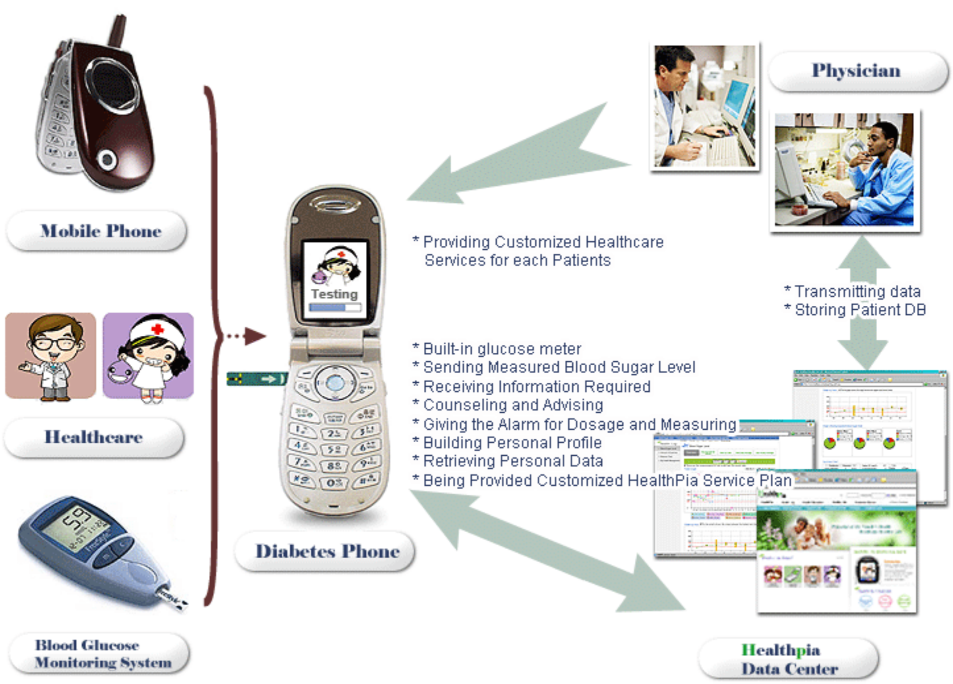
PATIENT-INFORMATION MANAGEMENT

Healthcare organisations are showing a clear interest in accelerating the transformation of clinical care through the routine use of appropriate emerging technologies by clinicians when diagnosing problems and subsequently planning and administering patient care. To support such noble efforts toward delivering better healthcare to the public, President George Bush and other national leaders have publicly called for the development of a national health information infrastructure. The U.S. government (in 2005) has over the past 2 years published plans for all Americans to have an electronic health record by 2014. Similar to the NPfIT in the United Kingdom, the plans call for a hierarchical set of local, regional, state, and national networks that facilitate peer-to-peer sharing of patient records.

When considering the deployment of any mobile data solution, reliability, efficiency, and security are essential, none more so than in the emergency services if lives are to be saved. To support such communication of critical and personal information, there has been an increased demand for the creation of electronic methods for storing and tracking clinical information (see Figure 1). This requires the solution for some fundamental architectural problems within the healthcare environment: scalability, reliability, recoverability, interchangeable vocabularies, and integration:

- Most service providers can support several thousands of simultaneous log-ons. Many are finding it difficult to demonstrate scalability in thin-client, rules-based order entry or structured clinical information.
- Service providers need to show evidence that they have appropriate schedules for downtime because healthcare organisations require reliability of 99.999% due to the critical nature of the information in the healthcare industry.
- There has to be a recognisable solution for a very quick data-recoverable plan in the event that downtime or a system failure occurs. Healthcare organisations have to ensure there are fault-proof backup plans to provide medical practitioners with information that is only available in an electronic form in the event that systems become suddenly unavailable. Certain work processes in the healthcare industry (including emergency care, scheduling, registration, order entry, and clinical procedure recording) would need to continue seamlessly, even with a primary-system interruption (see Figure 1). After there has been an interruption, recovery must be complete with no loss of

Figure 1. Flexible and independent patient care (http://www.healthpia.us/services.asp)



information. Backup, therefore, must prevent any IT failure from making care to the patient impossible. Where mobile healthcare technology is involved, adequate hardware, infrastructure, and tested processes should exist as part of a complete implementation to guarantee this recoverability.

Due to a previous lack of harmonized acquisition, healthcare organisations in both the United Kingdom and USA are frequently challenged by a variety of code sets and files that have proliferated across various healthcare institutions. HIPAA attachment transactions (in the United States) and the NPfIT (in the United Kingdom) are beginning to dictate that the future exchange of patient information be carried out electronically between healthcare organisations. To facilitate this portable and interoperable mobile healthcare technology, certain local vocabularies need to be replaced and the use of government-specified code sets should be synchronized. The way forward is to maintain current systems and historic data through mapping infrastructures that manage the correct translation, giving semantic meaning to patient data with the hope that one day soon there will be a complete migration to common vocabularies.

Many applications currently in the NHS have been designed with the assumption that that the approach and architecture does not need to co-exist and interoperate. While some of these may support integration with other applications that also have significantly different philosophical stances, they do not fully recognise that the need for a healthcare organisation to implement a total solution involves the practicalities of many different dimensions of time, scope, economics, and service providers' organisational politics. Toward this goal, all interoperability for a mobile healthcare technology should require that all features and functions work across all applications. The NPfIT project is proving that all service providers have to significantly alter their current approach to internal and external integration, security, and nomenclatures during the life of the project.

Case 1: MotoHealth

Motorola, along with its partners, initiated a telemedicine service at Harvard Teaching Hospital called MotoHealth (http://www.motorola.com/mediacenter/news/detial) late 2004. The Motorola solution uses mobile phones to help healthcare professionals to monitor chronically ill patients during their normal daily routines. This product was designed to meet the customer's convenience, and as a discreet way of monitoring patients in the mobile environment, can replace in-home hospital and home monitoring devices. As a result, it gives the patient more independence to continue daily activities virtually anywhere they like. This method to providing healthcare pushes healthcare and services out of high-cost health institutions. It enables the patient's body to become the point of care and the mobile healthcare technology becomes the bridge to the patient's body, thus, enabling the delivery of care, educational advice, and support remotely and transparently.

This case has proven that when a mobile healthcare technology is implemented as part of a comprehensive healthcare program, it can give healthcare providers useful daily updates on a patient's physiological levels such as blood pressure, glucose level, and weight control. Such a method of healthcare facilitates proactive treatment action, resulting in fewer hospitalizations and visits to emergency rooms, potentially lowering the increasing demand on the costs of providing healthcare and services to the public.

Policy Issues

Arguably, the most viable techniques for successful mobile healthcare technology implementation are practical guidelines and good management practices. Policies established by a healthcare organisation are the first steps toward this goal.

There are, however, few steps for establishing a policy for mobile healthcare technology:

Guidelines must be developed for the acquisition of mobile healthcare technologies. This would balance the need to encourage innovative applications against wasteful spending, which can be seen by certain members of the staff as duplication of effort. This in turn makes it the responsibility of every medical practitioner who may need a particular type of application to strictly adhere to this policy.

There must be regular inventory taken. This helps to identify and evaluate all installed hardware and software before setting or changing the standards. These would certainly affect policy decisions and future acquisitions. While standards can sometimes be looked upon as restrictive in the IT industry, medical practitioners actually see these to offer benefits for the care providers and the patients alike. Nearly all healthcare organisations have standards that cover many aspects of their operations within the healthcare process. Generally, standards in the NHS are recorded in formal standards and procedures manuals, but in certain cases, we came across informal handwritten notes (i.e., "this is the way we do things here") that are also considered to be standards.

Ruyter, Wetzels, and Kleijnen (2001) show how organisations implementing e-commerce first learn to exploit the Internet for information transfer before supporting transactions, and then finally use it for commercial trading and collaboration among various actors. Considering mobile healthcare is still in its infancy, borrowing from the e-commerce experience will mean that healthcare organisations will probably adopt wireless e-business methodologies first to support their existing healthcare processes and improve efficiency before they come up with new business models to transform the competitive landscape in the healthcare industry.

In the case of the NHS, wireless enterprise implementation issues frequently extend well beyond the technology domain and into areas of medical practices and organisational culture (see Figure 2). Nearly all the regional healthcare trusts that are actively pursuing wireless enterprise strategies at the moment are handcrafting solutions around their own local IT infrastructures and their own homegrown healthcare processes partly because there are very few packaged mobile healthcare solutions on the market.

The focus, currently, is on accessing information via wireless mobile healthcare messaging. However, the future should hold more applications like mobile access, telemedicine, and alerts for facilitating better disease management and controls. Given the emerging state of mobile

Figure 2. Home visits and general-practice consultation (http://www.bmj.com, 2004)

technology and its potential impact on healthcare, mobile healthcare can be seen as truly radical. Mobile healthcare has the potential to remake this entire industry and obsolete established strategies. Most healthcare organisations in Western Europe and North America feel they must participate in this emerging healthcare technology in order to survive the increasing demand to service a continuously evolving patient environment.

The research found two reasons why healthcare organisations are beginning to pay keen attention to mobile healthcare:

- These organisations are being defensive. There is a general belief that newcomers in the healthcare-provision market may be plotting to use new functionalities available through the use of mobile healthcare to attack the incumbents' core providers.
- Converse to the previous reason is the understanding that mobile healthcare could realize its potential. If this happens successfully, mobile healthcare would be too attractive a proposition to ignore, and joining in at a later date may prove too expensive.

The author poses a conceptual framework for the research upon which this chapter is based, showing the different stages of the technology-adoption process for healthcare organisations as well as the main factors operating at each stage (see Figure 3). In the preadoption stage, healthcare organisations take an internal perspective and analyse the fitness of the mobile healthcare technology for the contemplated task as well as the value of this technology. These should be the drivers of the adoption decision. The phase after that shows the healthcare organisation analysing whether organisational and environmental factors are favourable for continuing with this novel technology. This may uncover inhibitors that could slow down the adoption process. In situations where the healthcare organisation decides to implement the mobile healthcare technology in the next stage, it should find adoption facilitators that can help in the diffusion of mobile technology within the healthcare environment. Readers should note that implementation is beyond the scope of this framework.

This framework borrows from the technology-organisation-environment framework put forward by Tornatzky and Fleischer (1990) as well as the deinstitutionalisation framework by Tolbert and Zucker (1994). Tornatzky and Fleischer's framework identifies three aspects of an organisation's context that influence the process by which it adopts and implements innovation: technological context, organisational context, and environmental context. The omission of environmental context here is due to the fact that this chapter is about a single industry in which the environment is held constant.

The author considers innovation in healthcare as an idea, medical practice, or any material artefact in the healthcare process that is perceived to be new by the relevant unit of adoption in medical

Figure 3. Conceptual framework of mobile healthcare technology adoption

treatment. The relative advantage, compatibility, complexity, trial, and observation of such innovation can usually be used to determine the tendency for its adoption in the healthcare environment.

By relative advantage, the author means the degree to which an innovation within the health industry is perceived as better than the idea it supersedes. Compatibility, on the other hand, is seen as the degree to which an innovation within the health industry is perceived as being consistent with the existing values given by a particular healthcare community, or with past experiences and needs of potential adopters within the healthcare process. Compatibility of mobile healthcare technology can also be explained in terms of a combination of what healthcare practitioners feel or think about a particular innovation. This would also involve a critical look at the practical and operational compatibility with what healthcare practitioners are doing in the ongoing healthcare process. This interpretation of compatible innovation is in conjunction with perceived usefulness, the degree to which an end user believes a certain system can help perform a certain medical task. Complexity is the degree to which an innovation within the health industry is seen to be difficult to understand and use within the healthcare industry.

The trial of an innovation within the healthcare industry is an important part of evaluating new technologies within this critical industry. It is the degree to which an innovation within the health industry is experimented with on a limited basis. Given an opportunity to experiment with a new mobile healthcare technology before decisions are made about the adoption is an important benefit, especially for emerging technologies. This is an industry where practitioners take very highly the availability of information, while learning from experiences with previously disappointing IT projects.

Observation is a reliable means by which the healthcare industry evaluates innovations. This process identifies the degree to which the performance of a mobile healthcare technology and related benefits to the patients are visible to the medical practitioners and not only the service providers.

The determinants of mobile healthcare technology adoption are the benefits to the patients vs. the cost of such adoption. Most often, the NHS measures this in terms of the difference in costs for the shift from a previous technology to a mobile healthcare technology. Also worth considering are several factors that are important to the health service, such as the improvements made to the healthcare process as a result of a mobile healthcare technology after its introduction. There might even be a discovery of new uses for the mobile healthcare technology and the development of certain complementary inputs.

Hartmann and Sifonis (2000) relates to some of these features of mobile healthcare technology application by the identification of four dimensions within an organisation: leadership, governance, technology, and operational competencies. By leadership, they referred to the process of managing the initiatives and how the host organisation should stay motivated throughout the adoption process. By governance, they referred to the process of organising the innovation as regards the structure and operating procedures. Technology is where Hartmann and Sifonis looked at the organisation's ability to rapidly develop and implement new applications. They finally explained operational competencies as the way the host organisation manages the coordination of the relationship between leadership, governance, and technology as well as exploiting the available resources.

Levy and Powell (2005), on the other hand, presented evidence—from their study of small and medium-sized businesses in the United Kingdom—that the adoption of emerging technologies posits a similar framework as adoption related to the readiness of organisations taking into consideration the perceived ease of use and perceived usefulness. The readiness of the NHS to adopt mobile healthcare technology can be determined by reviewing the financial and technological resources available as well as various other factors dealing with the compatibility and consistency of emerging technologies with organisational culture and values.

Case 2: Pervasive Monitoring System

Oracle, along with its London partners, piloted a wireless sensor interface technology platform in mid-2005 (http://www.toumaz.com/news.php?act). It used advanced transactional database capabilities and offered the potential to transform the treatment and management of chronic diseases for millions of patients. This project was meant to bring the economies of scale of semiconductors into the healthcare industry with its advantages of real-time, personalised care and the delivery of some form of breakthrough. The system was based on a low-cost, disposable, integrated sensor interface chip.

Due to the chip's ultra low power and very small battery size, it could be worn on the body with complete freedom of movement, or it could be implanted. Such a mobile healthcare technology is compatible with a wide range of sensors (see Figure 4) and can therefore be configured to detect vital signs such as ECG (electrocardiogram), blood oxygen and glucose, body temperature, and mobility. The device can also dynamically process and filter event data (including irregularities in heartbeat or blood pressure) and send the details to a mobile phone or PC via an ultra low-power, short-range radio telemetry link.

Further improvements to this kind of mobile healthcare technology could enhance the quality and efficiency of the healthcare patients of the NHS receive in the future. It could permit the following in future healthcare:

- Provide more timely and personalised care.
- Deliver unprecedented freedom, flexibility, and control for patients.
- Include exciting possibilities for medical practitioners to ultimately offer consumer items for which selection is based on quality and efficiency.

Figure 4. Mobile healthcare technology with built-in sensor (http://www.healthpia.us)

CONCLUSION

This chapter has provided examples of mobile healthcare technologies (case studies) for which the successful delivery of mobile solutions can help with certain kinds of emergency services. When considering the deployment of any mobile data solution, reliability, efficiency, and security are essential, none more so than in the emergency services if lives are to be saved. Wireless and mobile technology offers the opportunity to vastly enhance services, but there are still challenges to be faced if a cost-efficient solution is to be delivered.

Although some of these initiatives described as e-health can deliver certain benefits (including increased productivity and effectiveness of healthcare personnel and improved delivery of information and services), they will be faced with a number of challenges. Service providers in the private sector are looking to government for more leadership on identification issues and, as such, mobile healthcare technology should be a welcomed measure.

IT is seen as a key driver in the delivery of an efficient public sector, but how can departments justify further expenditure and eventually provide a clear road map to return on investment whilst delivering what has been promised? Also, what are the key short-term issues and, more importantly, the solutions that government departments can focus on? This panel will examine the savings that ICT investment is expected to deliver in the public sector, and how to serve more people by making things more efficient.

The United Kingdom has certainly increased its uptake on open-source software since the Office of Government Commerce's (OGC's) announcement that open source is a viable desktop alternative for the majority of users. However, many government organisations have chosen to remain with their existing proprietary software. This panel will examine the advantages and the drawbacks of both software solutions.

While this chapter is not intended to give specific guidelines for using mobile healthcare technologies, the author finds it useful to mention the following two points:

- Have clear objectives. Mobile healthcare technology is only a means to an end. It is advisable for managers of healthcare organisations to not be dazzled by the technology.
- Mobile healthcare technology is a unique medium, requiring management to capitalize on its uniqueness. The information being transmitted by mobile healthcare technology is the same, but is just delivered in a different way.

A conceptual framework has been posed from the research upon which this chapter is based, showing the different stages of the mobile healthcare technology-adoption process for healthcare organisations as well as the main factors operating at each stage.

In conclusion, the author has argued that Web services can aid the strategic planning of a healthcare organisation and can be used for competitive advantage. Web services can also contribute to improving our understanding and management of the critical issues surrounding mobile-based healthcare. Such understanding not only avoids disastrous consequences during the adoption of information systems, but also proves essential in supporting healthcare professionals to effectively manage the current trend of rapid increase in healthcare costs.

REFERENCES

Amit, R., & Zott, C. (2001). Value creation in e-business. *Strategic Management Journal, 22*,493-520.

Atkinson, C. J., & Peel, V. J. (1998). Transforming a hospital through growing, not building, an electronic patient record system. *Methods of Information in Medicine, 37*, 285-293.

Banker, R., & Kauffman, R. (1988). Strategic contributions of information technology: An empirical study of ATM networks. In *Proceedings of the Nineth International Conference on Information Systems*, Minneapolis.

Barney, J. B., & Griffin, R. W. (1992). *The management of organizations: Strategy, structure, behavior.* Boston: Houghton Mifflin.

Brown, S. (2001). NHS finance: The issue explained. *The Guardian*, 30 May (pp. 21).

Brown, S. A., & Venkatesh, V. (2003). Bringing non-adopter along: The challenge facing the PC industry. *Communications of the ACM, 46*(4), 76-80.

Collins, T. (2003). Doctors attack health IT codes. *ComputerWeekly,* February 6, (pp. 21).

Currie, W., Desai, B., & Khan, N. (2004). Customer evaluation of application services provisioning in five vertical sectors. *Journal of Information Technology, 19*, 39-58.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use and user acceptance of information technology. *Management Information Systems Quarterly, 13*(4), 982-1003.

Edmondson, A. C. (2003). Framing for learning: Lessons in successful technology implementation. *California Management Review, 45*(2), 34-54.

Ferlie, E. B., & Shortell, S. M. (2001). Improving the quality of healthcare in the United Kingdom and the United States: A framework for change. *Milbank Quarterly, 79*, 281-315.

Guah, M. W., & Currie, W. L. (2004). Application service provision: A technology and working tool for healthcare organisation in the knowledge age. *International Journal of Healthcare Technology and Management, 6*(1/2), 84-98.

Guah, M. W., & Currie, W. L. (2006). *Internet strategy: The road to Web services solutions.* PA: IRM Press.

Hagel, J. (2002). *Out of the box: Strategies for achieving profits today and growth tomorrow through Web services.* Boston: Harvard Business School Press.

Haines, M. (2002). *Knowledge management in the NHS: Platform for change.* Department of Health. Retrieved November 2002 from http://www.healthknowledge.org.uk

Hartmann, A., & Sifonis, J. (2000). *NetReady-strategies for the success in the e-economy.* New York: McGraw-Hill.

Institute of Medicine. (2002). *Crossing the quality chasm: A new health system for the 21st century.* Washington, DC: Committee on Quality Healthcare in America, National Academy Press.

Kreger, H. (2003). Fulfilling the Web services promise. *Communications of the ACM, 46*(6), 29-34.

Levy, M., & Powell, P. (2005). *Strategies for growth in SMEs: The role of information and information systems.* Oxford: Butterworth Heinemann.

Majeed, A. (2003). Ten ways to improve information technology in the NHS. *British Medical Journal, 326*, 202-206.

McGauran, A. (2002). Foundation hospitals: Freeing the best or dividing the NHS? *British Medical Journal, 324*(1), 1298.

Metters, J., Abrams, M., Greenfield, P. R., Parmar, J. M., & Venn, C. E. (1997). *Report to the Secretary of State for Health of the professional committee on the appeal of Mr. D. R. Walker under paragraph 190 of the terms and conditions of service of hospital medical and dental staff (England and Wales).* London: Department of Health.

National Health Service Information Authority

(NHSIA). (2003). *Annual operating plan: To be the national provider of information and infrastructure services.* London: UK, Department of Health.

Orlikowski, W.J. (1993). CASE tools as organizational change: Investigating incremental and radical changes in systems development. *MIS Quarterly, 17*(3).

Orlikowski, W. J., & Tyre, M. J. (1994). Windows of opportunity: Temporal patterns of technological adaptation in organisations. *Organisation Science*, May, 98-118.

Pencheon, D. (1998). Matching demand and supply fairly and efficiently. *British Medical Journal, 316*, 1665-1667.

Perseid Software Limited. (2003). *The strategic value of Web services for healthcare and the life sciences.* Retrieved August 2003 from http://www.perseudsiftware.com

Robinson, J. C. (2000). Deregulation and regulatory backlash in healthcare. *California Management Review, 43*(1), 13-33.

Ruyter, K. D., Wetzels, M., & Kleijnen, M. (2001). Customer adoption of e-service: An experimental study. *International Journal of Service Industry Management*, *2*, 184-206.

Sleeper, B., & Robins, B. (2001). *Defining Web services.* Retrieved April 2002 from http://www.stencilgroup.com

Srinivasan, R., Lilien, G. L., & Rangaswamy, A. (2004). Technological opportunism and radical technology adoption: An application to e-business. *Journal of Marketing, 66*(3), 47-60.

Tolbert & Zucker. (1994). *Institutional Analysis of Organizations: Legitmate but not Institutionalized*. Institute for Social Science Research working paper, University of California, Los Angeles, Vol. 6, No. 5.

Tornatzky, L. G., & Fleischer, M. (1990). *The process of technology innovation.* Lexington: Lexington Books.

Walsham, G. (1993). *Interpreting information systems in organisations.* Chichester, United Kingdom: Wiley.

Wanless, D. (2002). *Securing our future health: Taking a long-term view. Final report of an independent review of the long-term resource requirement for the NHS.* London: Department of Health.

Willcocks, L., & Lacity, M. (1998). *Strategic sourcing of information systems.* John Wiley & Sons, New York.

Yin, R. K. (1994). *Case study research: Design and methods.* CA: Sage Publications.

This work was previously published in Web Mobile-Based Applications for Healthcare Management, edited by L. Al-Hakim, pp. 355-375, copyright 2007 by IRM Press (an imprint of IGI Global).

Chapter 2.8
Design of an Enhanced 3G-Based Mobile Healthcare System

José Ruiz Mas
University of Zaragoza, Spain

Eduardo Antonio Viruete Navarro
University of Zaragoza, Spain

Carolina Hernández Ramos
University of Zaragoza, Spain

Álvaro Alesanco Iglesias
University of Zaragoza, Spain

Julián Fernández Navajas
University of Zaragoza, Spain

Antonio Valdovinos Bardají
University of Zaragoza, Spain

Robert S. H. Istepanian
Kingston University, UK

José García Moros
University of Zaragoza, Spain

ABSTRACT

An enhanced mobile healthcare multi-collaborative system operating over Third Generation (3G) mobile networks is presented. This chapter describes the design and use of this system in different medical and critical emergency scenarios provided with universal mobile telecommunications system (UMTS) accesses. In these environments, it is designed to communicate healthcare personnel with medical specialists in a remote hospital. The system architecture is based on advanced signalling protocols that allow multimedia multi-collaborative conferences in IPv4/IPv6 3G scenarios. The system offers real-time transmission of medical data and videoconference, together with other non real-time services. It has been optimized specifically to operate over 3G mobile networks using the most appropriate codecs. Evaluation results show a reliable performance over IPv4 UMTS accesses (64 Kbps in the uplink). In the future, advances in m-Health systems will make easier for mobile patients to interactively get the medical attention and advice they need.

INTRODUCTION

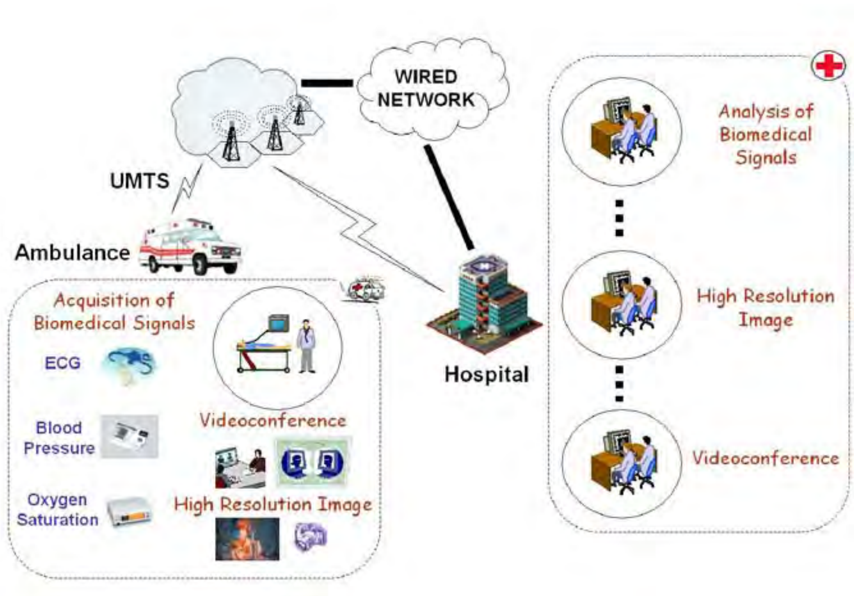
Mobile health (m-health) is an emerging area of telemedicine in which the recent development in mobile networks and telemedicine applications converge. m-health involves the exploitation of mobile telecommunication and multimedia technologies and their integration into new mobile healthcare delivery systems (Istepanian & Lacal, 2003). Wireless and mobile networks have brought about new possibilities in the field of telemedicine thanks to the wide coverage provided by cellular networks and the possibility of serving moving vehicles. One of the first wireless telemedical systems that utilized second-generation (2G) global system for mobile communications (GSM) networks addressed the electrocardiogram (ECG) transmission issues (Istepanian, 2001a). In recent years, several m-health and wireless telemedical systems based on GSM were reported (Istepanian, 2001b), allowing the accomplishment of remote diagnosis in mobile environments, as well as communication to geographic zones inaccessible by wired networks. The recent developments in digital mobile telephonic technologies (and their impact on mobility issues in different telemedical and telecare applications) are clearly reflected in the fast growing commercial domain of mobile telemedical services. A comprehensive review of wireless telemedicine applications and more recent advances on m-health systems is presented in Istepanian, Laxminarayan, and Pattichis (2005).

However, 2G-based systems lack the necessary bandwidth to transmit bandwidth-demanding medical data. The third-generation (3G) universal mobile telecommunications system (UMTS) overcomes limitations of first and second mobile network generations supporting a large variety of services with different quality of service (QoS) requirements. However, this fact makes network design and management much more complex. New applications require networks to be able to handle services with variable traffic conditions keeping the efficiency in the network resources utilization. The UMTS air interface is able to cope with variable and asymmetric bit rates, up to 2 Mbps and 384 kbps in indoor and outdoor environments, respectively, with different QoS requirements such as multimedia services with bandwidth on demand (Laiho, Wacker, & Novosad, 2000). In this kind of scenario, the emergence of 3G mobile wireless networks will permit to extend the use of m-health applications thanks to the provided higher transmission rates and flexibility over previous mobile technologies.

UMTS introduces the IP multimedia core network subsystem (IMS) (3GPP, 2005a), an IPv6 network domain designed to provide appropriate support for real-time multimedia services, independence from the access technologies and flexibility via a separation of access, transport and control. The fundamental reason for using IPv6 is the exhaustion of IPv4 addresses. Support for IPv4 is optional, but since network components require backward compatibility, it is clear that a dual stack configuration (IPv4 and IPv6) must be provided. The IMS uses the session initiation protocol (SIP) as signalling and session control protocol (Rosenberg et al., 2002). SIP allows operators to integrate real-time multimedia services over multiple access technologies such as general packet radio service (GPRS), UMTS or, ultimately, other wireless or even fixed network technologies (interworking multimedia domains). This chapter presents a 3G-based m-health system designed for different critical and emergency medical scenarios, as shown in Figure 1. Several medical specialists in the hospital take part in a multipoint conference with the ambulance personnel, receiving compressed and coded biomedical information from the patient, making it possible for them to assist in the diagnosis prior to its reception.

The 3G system software architecture includes intelligent modules such as information compression and coding, and QoS control to significantly improve transmission efficiency, thus optimizing the use of the scarce and variable wireless channel bandwidth compared to previous systems (Chu & Ganz, 2004; Curry & Harrop, 1998). Finally, unlike Chu and Ganz (2004), this m-health system follows a multi-collaborative design which supports IPv6/IPv4 interworking, uses SIP as

Figure 1. Typical medical mobility scenario



the service control protocol and integrates new real-time multimedia features intended for 3G wireless networks.

3G M-HEALTH SYSTEM ARCHITECTURE

In this section, the 3G m-health system structure is described in detail. The system (see the system components in Figure 2 and the main application graphical user interface (GUI) in Figure 3) has been built using standard off-the-shelf hardware, instead of developing propriety hardware as in Cullen, Gaasch, Gagliano, Goins, and Gunawardane (2001), uses free software and commercially available 3G wireless UMTS cellular data services. In addition, it provides simultaneous transfer of, among other services, videoconference, high-resolution still medical images and medical data, rather than only one media at a time (Kyriacou, 2003; Pavlopoulos, Kyriacou, Berler, Dembeyiotis, & Koutsouris, 1998).

The 3G m-health system consists of different modules that allow the acquisition, treatment, representation and transmission of multimedia information. The modular design of each medical user service allows great flexibility. In addition, there exist other medical user services like chat and electronic whiteboard that allow data exchange in order to guide the operations performed by remote users.

The medical signals module acquires, compresses, codes, represents, and transmits medical signals in real time. The medical signals acquisition devices included are: a portable electrocardiograph that allows the acquisition of 8 real and 4 interpolated leads of the ECG signal, and that follows the standard communication protocol-ECG (SCP-ECG); a tensiometer that provides systolic and diastolic blood pressure values; and a pulsioximeter that offers the blood oxygen saturation level (SpO_2) and the cardiac pulse.

The details of the 3G system architecture are shown in Figure 4 and Figure 5. As it can

Figure 2. 3G m-Health system

Figure 3. m-Health application GUI

be seen, the system comprises of the signalling and session control, medical user services and application control sub-systems, which will be described later.

This architecture allows the 3G system to offer real-time services such as medical data transmission (ECG, blood pressure, heart rate and oxygen saturation), full-duplex videoconference, high-resolution still images, chat, electronic whiteboard, and remote database Web access.

Communication between the remote ambulance personnel and medical specialists is established by means of multipoint multi-collaborative sessions through several network environments capable of supporting the different types of multimedia traffic (Figure 4). The selected conference model (tightly coupled conference model (Rosenberg, 2004)) requires the existence of a MCU (multipoint control unit) to facilitate multipoint operation. The MCU maintains a dialog with each participant in the conference and is responsible for ensuring that the media streams which constitute the conference are available to the appropriate participants. The MCU can belong to mobile's home network (SIP application server) or to a external service platform. Furthermore, the m-health system is developed to support IPv4/IPv6 interworking (Wiljakka & Soinien, 2003) and can be integrated in a 3G network scenario (see Figure 4).

The MCU receives the information generated by each participant in the multipoint conference, processes and forwards it to the appropriate destinations. System users and the MCU exchange information associated with the different services provided (medical user services) and its presentation (application control). Moreover, they exchange information related to communication and service quality management (signalling). Next, a detailed description of the sub-systems is presented.

Signalling and Session Control

The developed signalling allows the exchange of the characteristics associated to the different information flows between system elements and is based on standard protocols that favour interoperability. Signalling tasks, performed by the SIP protocol, begin with the establishment of a SIP dialog with the MCU in which, by means of session description protocol (SDP) messages, the different services are described. In order to do that, each element in the system has a SIP user agent (UA), slightly modified in the MCU to allow the use of multiple simultaneous dialogs.

Figure 4. 3G network scenario

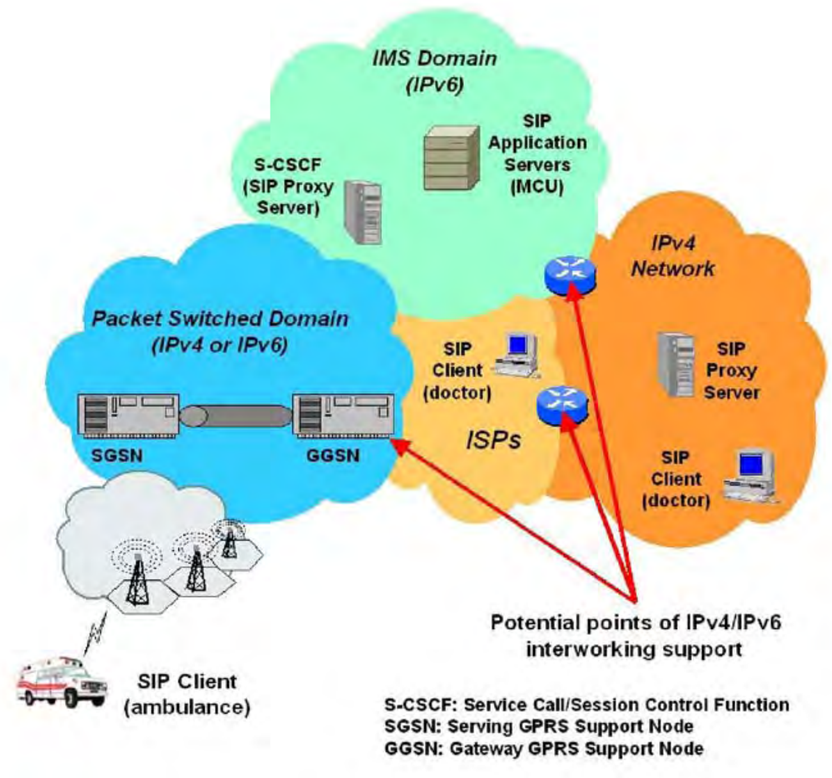
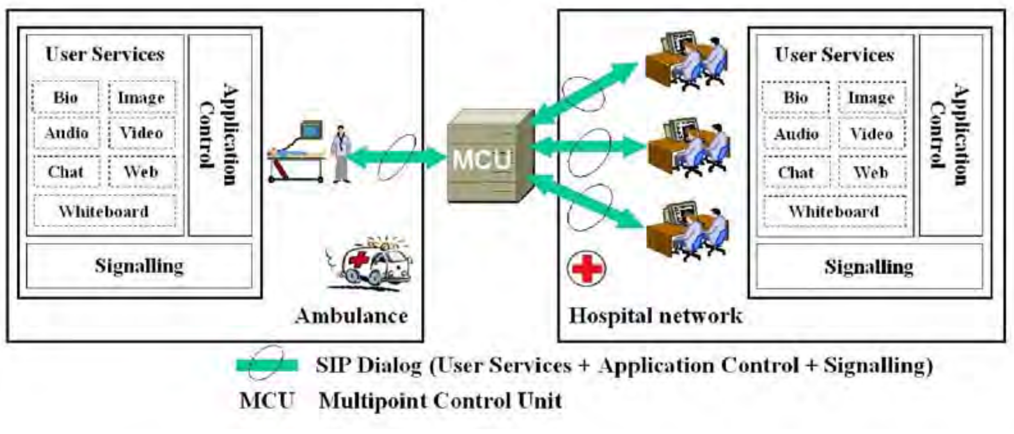


Figure 5. Block diagram of the m-Health system architecture and sub-systems



In addition to session control functions (establishment, management and termination of the multipoint conference), the SIP protocol is also useful for user mobility purposes inside the IMS environment.

Multipoint conference establishment, management and termination is performed by exchanging SIP messages between the different users. When a user connects, he creates a SIP dialog with the MCU, joining the conference. During the conference, SIP messages are exchanged between users and the MCU, varying conference characteristics and therefore allowing its management. In a similar process to that of conference joining, when a user wants to leave it, this fact must be communicated to the MCU with the necessary SIP messages. SIP messages also serve as the mean of transport of SDP messages with the description of medical user services.

The QoS in this system is mainly determined by the characteristics of the UMTS link. Mobile links are very variable, therefore a QoS-monitoring process is required in order to obtain a good system performance. This process is especially important in the MCU because it is there where the QoS-related decisions are taken. When the MCU detects that a particular conference participant needs to modify the characteristics of its multimedia session in order to improve QoS, it renegotiates the corresponding session by sending SIP/SDP messages. Hence, conference participants can modify certain upper-level protocol parameters (codecs used, transmission rates, compression ratios, etc.) in order to adapt the transmitted information to network performance.

The QoS monitoring process is possible thanks to a transport library that provides a uniform interface to send the information generated by medical user services and different QoS estimation tools developed for several types of links. This transport library offers different queuing policies and tools designed to measure the following QoS-related parameters: delay, bandwidth and packet loss rate. Due to the variable nature of wireless links, reception buffers have been properly dimensioned to minimize jitter, delay and packet loss.

Wireless Medical User Services

The medical user services included in the m-health system are associated with information shared in a multi-collaborative environment. Specifically, the system has services to share audio, video, medical data information, high-resolution still images, and graphical and textual information, as well as a Web service that allows remote access to clinical information databases. In addition to these services, there is a service designed to exchange control information (application control), which is discussed later.

Each kind of information is associated with a medical user service and uses a transport protocol and a codec according to its characteristics (see Table 1). Hence, real-time services (audio, video, and medical data information) use the real-time transport protocol (RTP), whereas the rest of the services use the transmission control protocol (TCP). Furthermore, the exchanged information can be very sensitive and requires a secure communication. The 3G m-health system uses an IP security protocol (IPSec) implementation supporting public key certificates in tunnel mode. This protocol ensures private communication of selected services.

The ECG signal is stored both in transmission and reception following the SCP-ECG standard. It is well known that for an efficient transmission an ECG compression technique has to be used. In our implementation, a real-time ECG compression technique based on the wavelet transform is used (Alesanco, Olmos, Istepanian, & García, 2003). This is a lossy compression technique, therefore the higher the compression ratio (lower the transmission rate), the higher the distortion at reception. It is clear that there is a trade-off between transmission rate and received ECG signal quality. From the transmission efficiency

Table 1. Codec operation modes for 3G real-time wireless medical user services

	CODEC	CODEC RATE
Audio	AMR*	4.74 5.15 5.9 6.7 7.4 7.95 10.0 12.2 (Kbps)
Video	H.263	5 10 (Frames per second)
Biomedical Signals	WT**	5 10 20 (Kbps)

* *Adaptive Multi-rate*
** *Wavelet Transform*

point of view, a very low transmission rate is desired but from the clinical point of view, a very distorted ECG is useless. Therefore, there exists a minimum transmission rate to be used so as the transmitted ECG is useful for clinical purposes, which was selected in collaboration with cardiologists after different evaluation tests. The minimum transmission rate used in our implementation (625 bits per second and per ECG lead) leads to a clinically acceptable received ECG signal.

Regarding blood pressure, oxygen saturation, and heart rate, these signals have low bandwidth requirements and, therefore, are not compressed.

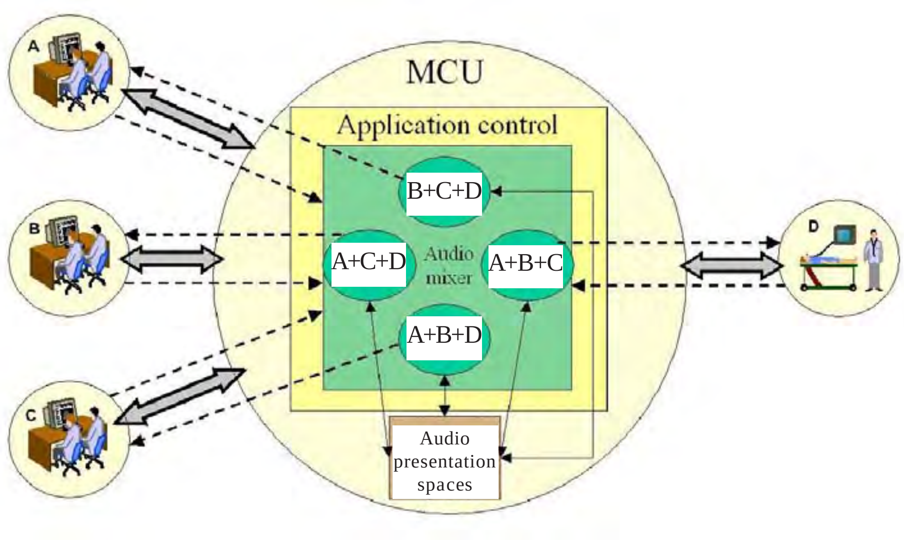
The videoconference module captures, sends, and plays audio and video information obtained by means of a Web camera and a microphone. In order to reduce the bandwidth, these data are compressed and coded. The video signal is compressed following the H.263 standard, whereas the audio signal uses the adaptive multi-rate (AMR) codec, recommended for UMTS by the 3G Partnership Program (3GPP) (3GPP 2005b). This module provides the basic functionality for starting, pausing and stopping video signals acquisition and representation, as well as volume control for the microphone (capture) and the speakers (reproduction). Due to the fact that each participant in the conference receives a unique video signal, the system allows the user to select the particular video signal among all the users connected.

The high-resolution still image module obtains high quality images with a charge copled device (CCD) colour camera connected to the computer through an image acquisition card. This module includes options to preview the captured images and modify their main characteristics in real time: brightness, contrast, hue, etc. Captured images can be stored and transmitted in different formats, with various qualities and compression levels. These images are sent automatically to the electronic whiteboard module of the remote users, allowing to select and mark fixed areas in a multi-collaborative fashion to facilitate a diagnostic clinical procedure.

Application Control

The MCU forwards the information generated by each medical service according to the presentation spaces defined using the control service. Each medical service has a presentation space associated with it that defines the way in which the information has to be transferred and its destination. The MCU simply forwards the information it receives, but has a special treatment for the audio, video, medical data, and control services. Regarding the Audio service, the MCU decodes the signal of each user, mixes it with the decoded signal of the other conference participants and codes the result in order to transfer a unique audio signal to each user (Figure 6). On the other hand, the MCU only forwards one video signal to each conference participant. The particular video

Figure 6. Application control (audio presentation spaces)



signal forwarded to each user is selected by using the control service. Finally, the medical data service is similar to the video service: medical data are only generated in the remote location, whereas the other conference participants can only receive them.

The 3G m-health system was adapted to the critical and emergency medical scenarios. In the first stages of its design, user requirements and functional specifications were established in collaboration with medical specialists, in order to create a portable and modular m-health system that could be easily integrated in any environment, using any underlying network technology capable of supporting IP multimedia services

3G M-HEALTH SYSTEM PERFORMANCE

In order to measure the 3G m-health system performance, several tests have been carried out using the system over 64 Kbps (at IP level) UMTS accesses. Table 2 presents the results about average IP-level bandwidth used by real-time network services.

As it can be observed, considering more audio samples per network packet reduces the used bandwidth, since transmission efficiency (information carried by each packet to total packet size ratio) is increased. However, there is a limit in the number of audio samples per packet that can be used because more audio samples per packet yield more audio delay. For example, a more efficient transmission mode including four audio samples per packet every 80 ms causes four times the delay including only one audio sample per packet every 20 ms. Moreover, if an audio packet is lost, all the audio samples carried by it are lost and, therefore, a reduced number of audio samples per packet is more suitable to error-prone environments. Regarding the video user service, it is worth noting that the bandwidth shown in Table 2 can vary substantially with the movement of the video scene captured. Finally, the medical data service adapts well to the codec rate specified because medical data frame sizes are long enough to obtain a good transmission efficiency.

As it can be checked, the total bandwidth consumed by all real-time medical user services fits in a 64 Kbps UMTS channel, even when the

Table 2. Average IP-level bandwidth used by real-time user services

	Operation Mode		IP Bandwidth (Kbps)
Audio	Samples/packet	Codec rate (Kbps)	
	1	4.75	21.2
	1	12.2	28.8
	3	4.75	10.5
	3	12.2	18.1
Video	Frames per Second		
	5		16
	10		24
Biomedical Signals	Bit Rate		
	5		5.3
	10		10.3

most bandwidth-consuming codec rates and the lowest transmission efficiencies are used. If all medical user services are used (including non real-time services), lower codec rates should be selected. Thus, according to the previous discussions, the codec operation modes selected in this m-health system have been those highlighted in Table 2, achieving a reasonable trade-off between bandwidth, transmission efficiency, delay and loss ratio.

NEXT-GENERATION OF M-HEALTH SYSTEMS

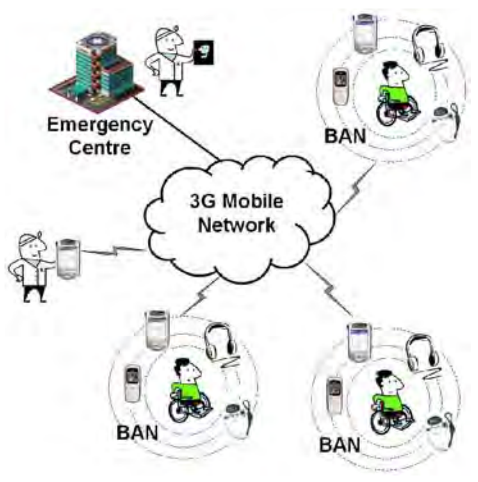
It is evident that organizations and the delivery of health care are being underpinned by the advances in m-health technologies. In the future, home medical care and remote diagnosis will become common, check-up by specialists and prescription of drugs will be enabled at home and virtual hospitals with no resident doctors will be realized.

Hence, the deployment of emerging mobile and wireless technologies will face new challenges: inter-operability between heterogeneous networks (fourth-generation, 4G) and smart medical sensor design integrating sensing, processing, communications, computing, and networking.

With the aid of wireless intelligent medical sensor technologies, m-health can offer healthcare services far beyond what the traditional telemedical systems can possibly provide. The individual sensors can be connected wirelessly to a personal monitoring system using a wireless body area network (WBAN) and can be integrated into the user's clothing, providing wearable and ubiquitous m-health systems. A typical scenario comprises of a WBAN system that communicates with cardiac implantable devices (pacemakers, defibrillators, etc.) and that is linked to the existing 3G infrastructure, achieving "Health Anytime, Anywhere" (Figure 7).

It is expected that 4G will integrate existing wireless technologies including UMTS, wireless LAN, Bluetooth, ZigBee, Ultrawideband, and other newly developed wireless technologies into a seamless system. Some expected key features of 4G networks are: high usability, support for multimedia services at low transmission cost and facilities for integrating services. 4G advances will make easier for mobile patients to interac-

Figure 7. Typical WBAN-3G scenario



tively get the medical attention and advice they need. When and where is required and how they want it regardless of any geographical barriers or mobility constraints.

The concept of including high-speed data and other services integrated with voice services is emerging as one of the main points of future telecommunication and multimedia priorities with the relevant benefits to citizen-centered healthcare systems. The new wireless technologies will allow both physicians and patients to roam freely, while maintaining access to critical patient data and medical knowledge.

CONCLUSION

This chapter has presented a feasible 3G-based m-health system targeted specifically for critical and emergency medical scenarios. The system architecture is based on 3G networks and advanced signalling protocols (SIP/SDP) that allow the integration of real-time multimedia services over multiple access channels that support IPv4 and IPv6 interworking depending on current commercial UMTS releases.

The system has the following features: simultaneous transmission of real-time clinical data (including ECG signals, blood pressure, and blood oxygen saturation), videoconference, high-resolution still image transmission, and other facilities such as multi-collaborative whiteboard, chat, and Web access to remote databases. The system has been optimized specifically to operate over 3G mobile networks using the most appropriate codecs. Evaluation results show a reliable performance over UMTS accesses (64 Kbps in the uplink).

Home telecare and chronic patient telemonitoring are other application areas in which this m-health system can be used, thus further work is currently undergone to adapt it and to evaluate its performance in each particular scenario.

REFERENCES

3GPP TS 23.228 V6.8.0. (2005). IP Multimedia Subsystem (IMS); Stage 2 (Release 6).

3GPP TS 26.235 V6.3.0. (2005). Packet switched conversational multimedia applications; Default codecs (Release 6).

Alesanco, A., Olmos, S., Istepanian, R. S. H., & García, J. (2003). A novel real-time multilead ECG compression and de-noising method based on the wavelet transform. *Proceedings of IEEE Computers Cardiology* (pp. 593-596). Los Alamitos, CA: IEEE Comput. Soc. Press.

Chu, Y., & Ganz, A. (2004). A mobile teletrauma system using 3G networks. *IEEE Transactional Information Technology in Biomedicine*, *8*(4), 456-462.

Cullen, J., Gaasch, W., Gagliano, D., Goins, J., & Gunawardane, R. (2001). *Wireless mobile telemedicine: En-route transmission with dynamic quality-of-service management*. National Library of Medicine Symposium on Telemedicine and Telecommunications: Options for the New Century.

Curry, G. R., & Harrop, N. (1998). The Lancashire telemedicine ambulance. *Journal Of Telemedicine Telecare*, *4*(4), 231-238.

Istepanian, R. S. H., Kyriacou, E., Pavlopoulos, S., & Koutsouris, D. (2001). Wavelet compression methodologies for efficient medical data transmission in wireless telemedicine system. *Journal of Telemedicine and Telecare*, *7*(1), 14-16.

Istepanian, R. S. H., Laxminarayan, S., & Pattichis, C. S. (2005). *M-Health: Emerging mobile health systems*. New York: Springer. To be published.

Istepanian, R. S. H., & Lacal, J. C. (2003). Emerging mobile communication technologies for health: Some imperative notes on m-health. *Proceedings of the 25th Silver Anniversary International Conference of the IEEE Engineering in Medicine and Biology Society 2* (pp. 1414-1416).

Istepanian, R. S. H., Woodward, B., & Richards, C. I. (2001). Advances in telemedicine using mobile communications. *Proceedings of the IEEE Engineering Medicine and Biology Society 4* (pp. 3556-3558).

Kyriacou, E., Pavlopoulos, S., Berler, A., Neophytou, M., Bourka, A., Georgoulas, A., Anagnostaki, A., Karayiannis, D., Schizas, C., Pattichis, C., Andreou, A., & Koutsuris, D. (2003). Multipurpose healthcare telemedicine systems with mobile communication link support. *BioMedical Engineering OnLine*, *2*(7).

Laiho, J., Wacker, A., & Novosad, T. (2000). *Radio network planning and optimization for UMTS*. New York: Wiley.

Pavlopoulos, S., Kyriacou, E., Berler, A., Dembeyiotis, S., & Koutsouris, D. (1998). A novel emergency telemedicine system based on wireless communication technology–AMBULANCE. *IEEE Trans. Inform. Technol. Biomed*, *2*, 261-267.

Rosenberg, J. (2004). *A framework for conferencing with the session initiation protocol*. Internet draft. Work in progress.

Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., & Schooler, E. (2002). *SIP: Session initiation protocol* (IETF RFC 3261).

Wiljakka, J., & Soinien, J. (2003). Managing IPv4-to-IPv6 transition process in cellular networks and introducing new peer-to-peer services. *Proceedings of IEEE Workshop on IP Operations and Management* (pp. 31-37).

KEY TERMS

4G: The fourth-generation (4G) is the continuation of the first, second, and third generations of mobile networks. It is a wireless access technology that provides high-speed mobile wireless access with a very high data transmission speed (2-20 Mbps) and enables users to be simultaneously connected to several wireless access technologies and seamlessly move between them in an all-IP environment. These access technologies can be any existing or future access technology. Smart antennas, low power consumption, and software-defined radio terminals will also be used to achieve even more flexibility for the user of 4G systems.

IMS: The IP multimedia subsystem (IMS) is a new open and standardized framework, basically specified for mobile networks, for providing Internet protocol (IP) telecommunication services. It offers a next generation network (NGN) multimedia architecture for mobile and fixed services, based on the session initiation protocol (SIP), and runs over the standard IP. It is used by telecom operators in NGN networks (combining voice and data in a single packet switched network), to offer network controlled multimedia services. The aim of IMS is not only to provide new services but to provide all the services, current and future, that the Internet provides. In addition, users have to be able to execute all their services when roaming as well as from their home networks. To achieve these goals the IMS uses IETF protocols. This is why the IMS merges the Internet with the cellular world; it uses cellular technologies to provide ubiquitous access and Internet technologies to provide new services.

IPv6: Internet protocol version 6 (IPv6) is the next generation protocol designed by the Internet Engineering Task Force (IETF) to replace the current version of the Internet protocol, IP version 4 (IPv4). Today's Internet has been using IPv4 for twenty years, but this protocol is beginning to become outdated. The most important problem of IPv4 is that there is a growing shortage of IPv4 addresses, which are needed by all new machines added to the Internet. IPv6 is the solution to several problems in IPv4, such as the limited number of available IPv4 addresses, and also adds many improvements to IPv4 in other areas. IPv6 is expected to gradually replace IPv4, with the two coexisting for a number of years during a transition period.

QoS: ITU-T recommendation E.800 defines the term quality of service (QoS) as "the collective effect of service performance which determines the degree of satisfaction of a user of the service." Service performance comprises of very different parts (security, operability, etc), so the meaning of this term is very broad. In telecommunications, the term QoS is commonly used in assessing whether a service satisfies the user's expectations. QoS evaluation, however, depends on functional components and is related to network performance via measurable technical parameters. A QoS-enabled network has the ability to provide better service (priority, dedicated bandwidth, controlled jitter, latency, and improved loss characteristics) to selected network traffic over various technologies.

SIP: The session initiation protocol (SIP) is a signalling protocol developed by the IETF intended for setting up multimedia communication sessions between one or multiple clients. It is currently the leading signalling protocol for voice over IP, and is one of the key components of the IMS multimedia architecture. The most important functions of SIP include name mapping and redirection (user location), capabilities negotiation during session setup, management of session participants and capabilities management during the session.

Telemedicine: Telemedicine can be defined as the rapid access to shared and remote medical expertise by means of telecommunications and information technologies, no matter where the patient or the relevant information is located. Any application of information and communications technologies which removes or mitigates the effect of distance in healthcare is telemedicine. The terms e-health and tele-health are terms often interchanged with telemedicine.

This work was previously published in Handbook of Research on Mobile Multimedia, edited by I. K. Ibrahim, pp. 521-533, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 2.9
The M-Health Reference Model:
An Organizing Framework for Conceptualizing Mobile Health Systems

Phillip Olla
Madonna University, USA

Joseph Tan
Wayne State University, USA

ABSTRACT

The reference model presented in this article encourages the breakdown of M-Health systems into the following five key dimensions: (1) Communication Infrastructure: a description of mobile telecommunication technologies and networks; (2) Device Type: the type of device being used, such as PDA, sensor, or tablet PC; (3) Data Display: describes how the data will be displayed to the user and transmitted, such as images, e-mail, and textual data; (4) Application Purpose: identification of the objective for the M-Health system; (5) Application Domain: definition of the area in which the system will be implemented. Healthcare stakeholders and system implementer can use the reference model presented in this article to understand the security implications of the proposed system and to identify the technological infrastructure, business requirements, and operational needs of the M-Health systems being implemented. A reference model that encapsulates the emerging M-Health field is needed for cumulative progress in this field. Currently, the M-Health field is disjointed, and it is often unclear what constitutes an M-Health system. In the future, M-Health applications will take advantage of technological advances such as device miniaturizations, device convergence, high-speed mobile networks, and improved medical sensors. This will lead to the increased diffusion of clinical M-Health systems, which will require better understanding of the components that constitute the M-Health system.

INTRODUCTION

M-Health is defined as "mobile computing, medical sensor, and communications technologies for healthcare" (Istepanian, Jovanov, & Zhang, 2004, p. 405). The first occurrence of the term *M-Health* in the literature was in the "Unwired E-Med" special issue on Wireless Telemedicine Systems (Istepanian & Laxminaryan, 2000). Since then, there has been an increased use of the term, encapsulating various types of healthcare systems. The use of the M-Health terminology relates to applications and systems such as telemedicine (Istepanian & Wang, 2003), telehealth (Istepanian & Lacal, 2003), and biomedical sensing system (Budinger, 2003). Until now, there have been considerable confusion and overlap with the use of these terms (Tulu & Chatterjee, 2005).

Rapid advances in Information Communication Technology (ICT) (Godoe, 2000), nanotechnology, biomonitoring (Budinger, 2003), mobile networks (Olla, 2005a), pervasive computing (Akyildiz & Rudin, 2001), wearable systems, and drug delivery approaches (Amy et al., 2004) are transforming the healthcare sector. The insurgence of innovative technology into the healthcare practice not only is blurring the boundaries of the various technologies and fields but also is causing a paradigm shift that is blurring the boundaries among public health, acute care, and preventative health (Hatcher & Heetebry, 2004). These developments not only have had a significant impact on current e-health and telemedical systems (Istepanian, Jovanov, & Zhang, 2004), but they also are leading to the creation of a new generation of M-Health systems with a convergence of devices, technologies, and networks at the forefront of the innovation.

This article proposes the use of a five-dimensional reference model in order to assist system implementers and business stakeholders in understanding the various components of an M-Health system. The approach used by the this article focuses on identifying different dimensions of a Mobile Healthcare Delivery System (MHDS) (Wickramasinghe & Misra, 2005), which then can be used to identify user security requirements for different categories in an organized manner. These dimensions were driven from our literature review (Bashshur, 2002; Bashshur, Reardon, & Shannon, 2000; Raskovic & Jovanov, 2004; Istepanian, Laxminaryan, & Pattichis, 2006; Jovanov, Milenkovic, Otto, & Groen, 2005; Field, 1996; Moore, 2002; Olla & Patel, 2003), and the model reflects a combination of various classification schemes proposed in earlier studies in order to classify telemedicine and telehealth systems.

Based on the previous definition, M-Health is a broad area that transcends multiple disciplines and utilizes a broad range of technologies. There is a variety of applications, devices, and communication technologies that are emerging in the M-Health arena and that can be combined to create the M-Health system. The dimensions consist of the following:

1. **Communication Infrastructure.** Description of the mobile telecommunication technologies that will be used, such as Bluetooth, wireless local area networks, or third-generation technologies (Olla, 2005a).
2. **Device Type.** Relates to the type of device being used to collect the medical data, such as Personal Digital Assistance (PDA), sensor, or tablet PC (Parmanto, Saptono, Ferrydiansyah, & Sugiantara, 2005).
3. **Data Display.** Describes how the data will be displayed and transmitted to the user through images, e-mail, textual data, and other types of data presentation languages (Tulu & Chatterjee, 2005).
4. **Application Purpose.** Identification of the objective for the M-Health system (Field, 1996).
5. **Application Domain.** Definition of the area in which the system will be implemented, such as clinical (e.g., dermatology, radiology, etc.) or non-clinical (e.g., billing, maintenance, etc.) domains (Bashshur et al., 2000).

Selecting different alternatives in the five dimensions will have implications on the functionality of the system; however, the key focus of this discussion will be the emphasis on how the security and integrity of the M-Health system is maintained, based on the dimension choices.

The rest of the article is organized as follows. "Background of M-Health" features a brief background of the evolution of M-Health. "Mobile Healthcare Delivery System Networks" provides an overview of existing and new mobile technologies suitable for the healthcare sector. "The M-Health Reference Model" introduces the five dimensions of the M-Health reference model. The penultimate section uses the reference model to decompose a real-life mobile health delivery system in order to illustrate the security concerns. This is followed by the conclusion.

BACKGROUND OF M-HEALTH

The phenomenon of providing care remotely using ICT can be placed into a number of areas, such as M-Health, telemedicine, and e-health, but, as summed up by R. L. Bashur, President Emeritus of the American Telemedicine Association, the terminology is not the important aspect. "It does not really matter what we call it or where we draw boundaries. … collective and collaborative efforts from various fields of science, including what we call now telemedicine is necessary" (Bashshur, 2002, p. 7). Emphasis on the various components and objectives of the various types of systems should be the priority, as this will increase the chances of implementing efficient and effective systems, of generating viable business models, and of creating secure systems that meet the needs of the stakeholders (Olla & Patel, 2003). Over the evolution of telemedicine, new terminologies have been created as new health applications and delivery options became available and as application areas extended to most healthcare domains. This resulted in confusion and identification of what falls under telemedicine, and what falls under telehealth or e-health became more complicated as the field advanced. New concepts such as Pervasive Health and M-Health also are adding to this confusion. Before understanding the scope and components of M-Health, it is important to mention briefly the history of telemedicine and the advancements of mobile networks, which are collectively the foundation of M-Health. The evolution and growth of telemedicine is correlated highly with ICT advancements and software development. Telemedicine advancements can be categorized into three eras (Tulu & Chatterjee, 2005; Bashshur et al., 2000).

The first era of telemedicine focused solely on medical care as the only function of telemedicine. This era can be named as the telecommunications era of the 1970s. The applications in this era were dependent on broadcast and television technologies, in which telemedicine applications were not integrated with any other clinical data. The second era of telemedicine, a result of digitalization in telecommunications, grew during 1990s. The transmission of data was supported by various communication mediums ranging from telephone lines to Integrated Service Digital Network (ISDN) lines. During this period, there were high costs attached to the communication mediums that provided higher bandwidth. The bandwidth issue became a significant bottleneck for telemedicine in this era. Resolving the bandwidth constraints has been a critical research challenge for the past decade, with new approaches and opportunities created by the Internet revolution. Now, more complex and ubiquitous networks are supporting telemedicine. The third era of telemedicine was supported by a networking technology that was cheaper and more accessible to an increasing user population. The improved speed and quality offered by Internet is providing new opportunities in telemedicine. In this new era of telemedicine, the focus shifted from a technology assessment to a deeper appreciation of the functional relationships between telemedicine technology and the outcomes of cost, quality, and access.

This article proposes a fourth era that is characterized by the use of Internet Protocol (IP) technologies, ubiquitous networks, and mobile/wireless networking capabilities, which can be observed by the proliferation of M-Health applications that perform both clinical and non-clinical functions. Since the proliferation of mobile networks, telemedicine has attracted a lot more interest from both academic researchers and industry (Tachakra, Istepanian, Wang, & Song, 2003). This has resulted in many mobile/wireless telemedicine applications being developed and implemented (Budinger, 2003; Istepanian & Laxminaryan, 2000; Jovanov, Lords, Raskovic, Cox, Adhami, & Andrasik, 2003; Pattichis et al., 2002; Istepanian & Lacal, 2003; Webb, 2004). Critical healthcare information regularly travels with patients and clinicians, and therefore, the need for information to become securely and accurately available over mobile telecommunication networks is key to reliable patient care and reliable medical systems.

Health organizations are required by the Health Insurance Portability and Accountability Act of 1996 (HIPAA) to maintain the privacy of protected health information and to provide individuals with notice of its legal duties and privacy policies with respect to this information (HIPPA, 2005). Specifically, the law requires the following:

- Any medical information that identifies a person must be kept private.
- Notice of legal duties and privacy practices with respect to protected health information must be made available.
- The organization must abide by the terms of its notices currently in effect.

HIPAA also requires:

- Computers and data containing protected health information (PHI) must be protected from compromise or loss.
- Audit trails of access to PHI must be kept.
- Electronic transmissions of PHI must be authenticated and protected from observation or change.

Failure to comply with these requirements (for both HIPAA privacy and security) can result in civil and criminal penalties ranging from fines of $100 to $250,000 and up to 10 years in prison. Medical information always has been considered sensitive and never more so than now due to the security issues created by mobility. Many countries have imposed strict regulations with heavy penalties in order to ensure the confidentiality and authorized distribution of personal medical information. Following the lead of European Union Directive 95/46/EC that protects both medical and financial personal information, the United States and Canada passed important legislation (HIPAA and PIPED, respectively) that imposes substantial penalties, both civil and criminal, for negligent or intentional exposure of personal medical information to unauthorized parties.

The telecommunication industry has progressed significantly over the last decade. There has been significant innovation in digital mobile technologies. The mobile telecommunication industry has advances through three generations of systems and is currently on the verge of designing the fourth generation of systems (Olla, 2005b). The recent developments in digital mobile technologies are reflected in the fast-growing commercial domain of mobile telemedical services (Istepanian et al., 2006). Specific examples include mobile ECG transmissions, video images and teleradiology, wireless ambulance services to predict emergency and stroke morbidity, and other integrated mobile telemedical monitoring systems (Istepanian, Jovanov, & Zhang, 2004; Future trends: Convergence is all, 2005; Warren, 2003). There is no doubt that mobile networks can introduce additional security concerns to the healthcare sector.

Implementing a mobile trust model will ensure that a mobile transaction safely navigates multiple technologies and devices without compromising the data or the healthcare systems. M-Health transactions can be made secure by adopting practices that extend beyond the security of the wireless network used and by implementing a trusted model for secure end-to-end mobile transactions. The mobile trust model proposed by Wickramasinghe and Misra (2005) utilizes both technology and adequate operational practices in order to achieve a secure end-to-end mobile transaction. The first level highlights the application of technologies in order to secure elements of a mobile transaction. The next level of the model shows the operational policies and procedures needed in order to complement technologies used. No additional activity is proposed for the mobile network infrastructure, since this element is not within the control of the provider or the hospital.

The next section will discuss the mobile network technologies and infrastructure, which is a key component of any M-Health system. The network infrastructure acts as a channel for data transmission and is subject to the same vulnerabilities, such as sniffing, as in the case of fixed network transaction. The mobile networks discussed in the next section are creating the growth and increased adoption of M-Health applications in the healthcare sector.

MOBILE HEALTHCARE DELIVERY SYSTEM NETWORKS

The implementation of an M-Health application in the healthcare environment leads to the creation of a Mobile Healthcare Delivery System (MHDS), which can be defined as the carrying out of healthcare-related activities using mobile devices such as a wireless tablet computer, Personal Digital Assistant (PDA), or a wireless-enabled computer. An activity occurs when authorized healthcare personnel access the clinical or administrative systems of a healthcare institution using a mobile device (Wickramasinghe & Misra, 2005). The transaction is said to be complete when medical personnel decide to access medical records (patient or administrative) via a mobile network either to browse or to update the record.

Over the past decade, there has been an increase in the use of new mobile technologies in healthcare, such as Bluetooth and Wireless Local Area Networks (WLAN) that use protocols different than the standard digital mobile technologies such as 2G, 2.5, and 3G. A summary of these technologies is presented next, and a précis of the speeds and range is presented in Table 1.

These mobile networks are being deployed in order to allow physicians and nurses easy access to patient records while on rounds, to add observations to the central databases, to check on medications, as well as to perform a growing number of other functions. The ease of access that wireless networks offer is matched by the security and privacy challenges presented by the networks. This serious issue requires further investigation and research in order to identify the real threats for the various types of networks in the healthcare domain.

Second Generation Systems (2G /2.5G)

The second-generation cellular systems (2G) were the first to apply digital transmission technologies such as Time Division Multiple Access (TDMA) for voice and data communication. The data transfer rate was on the order of tens of kbits/s. Other examples of technologies in 2G systems include Frequency Division Multiple Access (FDMA) and Code Division Multiple Access (CDMA).

The 2G networks deliver high quality and secure mobile voice and basic data services, such as fax and text messaging, along with full roaming capabilities around the world. Second-generation technology is in use by more than 10% of the world's population, and it is estimated that 1.3 billion customers in more than 200 countries and territories around the world use this technol-

Table 1. Comparison of mobile networks based on range and speed

Networks	Speed	Range and Coverage	Main Issues for M-Health
2nd Generation GSM	9.6 kilobits per second (KBPS)	World wide coverage, dependent on network operators roaming agreements.	Bandwidth limitation, Interference.
High Speed Circuit Switched Data (HSCSD)	Between 28.8 KBPS and 57.6 KBPS.	Not global, only supported by service providers network.	Not widely available, scarcity of devices.
General Packet Radio Service (GPRS)	171.2 KBPS	Not global, only supported by service providers network.	Not widely available.
EDGE	384 KBPS	Not global, only supported by service providers network.	Not widely available, scarcity of devices
UMTS	144 KBPS-2 MBPS depending on mobility	When fully implemented should offer interoperability between networks, global coverage.	Device battery life, operational costs.
Wireless Local Area	54 MBPS	30–50 m indoors and 100–500 m outdoors. Must be in the vicinity of hot spot.	Privacy, security.
Personal Area Networks-Bluetooth	400 KBPS symmetrically 150-700 KBPS asymmetrically	10-100m	Privacy, security, low bandwidth.
Personal Area Networks-Zigbee	20 kb/s-250 KBPS depending on band	30m	Security, privacy, low bandwidth.
WiMAX	Up to 70MBPS	Approx. 40m from base station.	Currently no devices and networks cards.
RFID	100 KBPS	1 m Non line-of-sight and contact less transfer of data between a tag and Reader.	Security, privacy.
Satellite Networks	400 to 512 KBPS new satellites have potential of 155MBPS.	Global coverage.	Data costs, shortage of devices with roaming capabilities. Bandwidth limitations.

ogy (GSM_Home, 2005). The later advanced technological applications are called 2.5G technologies and include networks such as General Packet Radio Service (GPRS) and Enhanced Data rates for GSM Evolution (EDGE). GPRS-enabled networks provide functionality such as always-on, higher capacity, Internet-based content and packet-based data services enabling services such as color Internet browsing, e-mail on the move, visual communications, multimedia messages and location-based services. Another complimentary 2.5G service is EDGE, which offers similar capabilities to the GPRS network.

Third-Generation Systems (3G)

The most promising period is the advent of 3G networks, which also are referred to as the Universal Mobile Telecommunications Systems (UMTS). A significant feature of 3G technology is its ability to unify existing cellular standards, such as code-division multiple-access (CDMA), global system for mobile communications (GSM_Home, 2005), and time-division multiple-access (TDMA), under one umbrella (Istepanian, Jovanov, & Zhang, 2004). More than 85% of the world's network operators have chosen 3G as the underlying technology platform to deliver their third-generation services (GSM-Information, 2004). Efforts are underway to integrate the many diverse mobile environments and to blur the distinction between the fixed and mobile networks. The continual rollout of advanced wireless communication and mobile network technologies will be the major driving force for future developments in M-Health systems (Istepanian, Jovanov, & Zhang, 2004). Currently, the GSM version of 3G alone saw the addition of more than 13.5 million users, representing an annual growth rate of more than 500% in 2004. As of December 2004, 60 operators in 30 countries were offering 3GSM services. The global 3GSM customer base is approaching 20 million and already has been launched commercially in Africa, the Americas, Asia Pacific, Europe, and the Middle East (GSM_Home, 2005), thus making this technology ideal for developing affordable global M-Health systems.

Fourth Generation (4G)

The benefits of the fourth-generation network technology (Istepanian et al., 2006; Olla, 2005a; Qiu, Zhu, & Zhang, 2002) include voice-data integration, support for mobile and fixed networking, and enhanced services through the use of simple networks with intelligent terminal devices. 4G also incorporates a flexible method of payment for network connectivity that will support a large number of network operators in a highly competitive environment. Over the last decade, the Internet has been dominated by non-real-time, person-to-machine communications (UMTS-Forum-Report14, 2002). The current developments in progress will incorporate real-time, person-to-person communications, including high-quality voice and video telecommunications along with extensive use of machine-to-machine interactions in order to simplify and to enhance the user experience.

Currently, the Internet is used solely to interconnect computer networks. IP compatibility is being added to many types of devices such as set-top boxes to automotive and home electronics. The large-scale deployment of IP-based networks will reduce the acquisition costs of the associated devices. The future vision is to integrate mobile voice communications and Internet technologies, bringing the control and multiplicity of Internet application services to mobile users (Olla, 2005b). 4G advances will provide both mobile patients and citizens with the choices that will fit their lifestyle and make it easier for them to interactively get the medical attention and advice they need when and where it is required and how they want it, regardless of any geographical barriers or mobility constraints.

Worldwide Interoperability for Microwave Access (WiMAX)

WiMAX is considered to be the next generation of Wireless Fidelity (WiFi /Wireless networking technology that will connect you to the Internet at faster speeds and from much longer ranges than current wireless technology allows (http://wimaxxed.com/). WiMax has been undergoing testing and is expected to launch commercially by 2007. Research firm Allied Business Research predicts that by 2009, sales of WiMax accessories will top $1 billion (Kendall & Christopher, 2005), and Strategy Analytics predicts a market of more than 20 million WiMAX subscriber terminals and base stations per year by 2009 (ABI Research, 2005).

The technology holds a lot of potential for M-Health applications and has the capabilities to provide data rates up to 70 mbps over distances up to 50 km. The benefits to both developing and developed nations are immense. There has been a gradual increase in the popularity of this technology. Intel recently announced plans to mass produce and release processors aimed to power WiMax-enabled devices (WiMax, 2005). Other technology organizations that are investing in the further advancement of this technology include Qwest, British Telecom, Siemens, and Texas Instruments. They aim to get the prices of the devices powered by WiMax to affordable levels so that the public can adopt them in large numbers, making it the next global wireless standard. There are already Internet Service Providers in metropolitan areas that are offering pre-WiMAX services to enterprises in a number of cities, including New York, Boston, and Los Angeles (WiMax, 2005).

Wireless Local Area Networks

Wireless Local Area Networks (WLAN) use radio or infrared waves and spread spectrum technology in order to enable communication between devices in a limited area. WLAN allows users to access a data network at high speeds of up to 54 Mb/s, as long as users are located within a relatively short range (typically 30–50m indoors and 100–500m outdoors) of a WLAN base station (or antenna. Devices may roam freely within the coverage areas created by wireless access points, the receivers, and transmitters connected to the enterprise network. WLANs are a good solution for healthcare today; in addition, they are significantly less expensive to operate than wireless WAN solutions such as 3G (Daou-Systems, 2001).

Personal Area Networks

A wireless personal area network (WPAN) (Personal area network, n.d.; Istepanian, Jovanov, & Zhang, 2004) is the interconnection of information technology devices within the range of an individual person, typically within a range of 10 meters. For example, a person traveling with a laptop, a personal digital assistant (PDA), and a portable printer could interconnect wirelessly all the devices using some form of wireless technology. WPANs are defined by IEEE standard 802.15 (IEEE 802.15 Working Group for WPAN, n.d.). The most relevant enabling technologies for M-Health systems are Bluetooth (n.d.) and ZigBee Alliance (n.d.). ZigBee is a set of high-level communication protocols designed to use small, low-power digital radios based on the IEEE 802.15.4 standard for wireless personal area networks. ZigBee is aimed at applications with low data rates and low power consumption. ZigBee's current focus is to define a general-purpose, inexpensive, self-organizing network that can be shared by industrial controls, medical devices, smoke and intruder alarms, building automation, and home automation. The network is designed to use very small amounts of power so that individual devices might run for a year or two with a single alkaline battery, which is ideal for use in small medical devices and sensors. The Bluetooth specification was first developed by Ericsson and later formalized by the Bluetooth Special Interest Group established by Sony, Ericsson, IBM, Intel, Toshiba, and Nokia and later joined by many other companies. A Bluetooth WPAN is also called a *piconet* and is composed of up to eight active devices in a master-slave relationship. A piconet typically has a range of 10 meters, although ranges of up to 100 meters can be reached under ideal circumstances. Implementations with Bluetooth versions 1.1 and 1.2 reach speeds of 723.1 kbit/s. Version 2.0 implementations feature Bluetooth Enhanced Data Rate (EDR) and thus reach 2.1 Mbit/s (Bluetooth, n.d.; Wikipedia, n.d.).

Radio Frequency Identification (RFID)

RFID systems consist of two key elements: a tag and a reader/writer unit capable of transferring data to and from the tag. An antenna linked to each element allows power to be transferred between the reader/writer and a remotely sited tag through inductive coupling. Since this is a bi-directional process, modulation of the tag antenna will be reflected back to the reader's /writer's antenna, allowing data to be transferred in both directions. Some of the advantages of RFID that make this technology appealing to the healthcare sector are as follows:

- No line of sight required between tag and reader.
- Non-contact transfer of data between tag and reader.
- Tags are passive, which means no power source is required for the tag component.
- Data transfer range of up to one meter is possible.
- Rapid data transfer rates of up to 100 Kbits/sec.

The use of RFID in the healthcare environment is set to rise and is currently being used for drug tracking. RFID technology is expected to decrease counterfeit medicines and to make obtaining drugs all the more difficult for addicts (Weil, 2005). There are also applications that allow tagging of patients, beds, and expensive hospital equipment.

Satellite Technologies

Satellite broadband uses a satellite to connect customers to the Internet. Two-way satellite broadband uses a satellite link both to send and to receive data. Typical download speeds are 400 to 512 kbps, while upload speeds on two-way services are typically 64 to 128 kbps. Various organizations (Inmarsat announces availability of the 64 kbit/s mobile office in the sky, 2000) have been investigating the development of an ultra-high-data-rate Internet test satellite for making a high-speed Internet society a reality (JAXA, 2005). Satellite-based telemedicine service will allow real-time transmission of electronic medical records and medical information anywhere on earth. This will make it possible for doctors to diagnose emergency patients even from remote areas and also will increase the chances of saving lives by receiving early information, as ambulance data rates of 155 mbps are expected. One considerable drawback associated with using this technology is cost (Olla, 2004).

This section has summarized the various mobile network technologies that are being used in the healthcare sector. The mobile technologies described have a significant impact on the ability to deploy M-Healthcare application and systems; however, there are combinations of other important factors described in the next section. The next section will present a five-layered model that uses mobile network technologies along with device type, data display, application purpose, and application domain to categorize M-Health systems.

THE M-HEALTH REFERENCE MODEL

The financial cost of delivering quality healthcare is increasing exponentially not only in North America but also around the world. To satisfy increasing healthcare challenges, organizations in the healthcare sector are investing innovative technological solutions in order to meet the higher expectations placed on practice management (Wickramasinghe & Misra, 2005; Wickramasinghe & Silvers, 2002). The evolution of M-Health solutions, such as the use of wireless networks to access patient records and other healthcare services such as billing and prescription, are becoming popular, especially in the U.S. and Canada (Goldberg & Wickramasinghe, 2002) and in many European countries (Sorby, 2002). However, these wireless solutions also bring with them their own challenges, with security being one of the most important issues in M-Health. This is because M-Health systems transmit highly sensitive information such as patient data over cyberspace, and there is a need for high-level, end-to-end security, confidentiality, and privacy.

This section presents a reference model to describe each M-Health system in order to address the security challenges. Specifically, we present a reference model and show how our model can facilitate a higher level of end-to-end security in a wireless or mobile healthcare environment. Despite the advancements of wireless technologies, the use of wireless security in healthcare is in its infancy, and a robust mobile healthcare model needs to be developed in order to allow further innovation of M-Health applications and services that will fulfill the growing needs of the healthcare sector.

This section presents a reference model that can be utilized to classify M-Health systems by identifying the system components for a specific M-Health system. As previously indicated, the approach used to develop the reference model concentrated on identifying different dimensions of a Mobile Healthcare Delivery System (MHDS), which provides the capability to identify user security requirements for different categories in an organized manner. The reference model was formulated by reviewing the literature, and the model reflects an amalgamation of various classification schemes proposed by previous experts to categorize telemedicine and telehealth systems. The following subsections will provide a description of these five domains for the M-Health reference model illustrated in Figure 1. These domains include communication infrastructure, device type, data display, application purpose, and application domain.

Mobile Communication Infrastructure

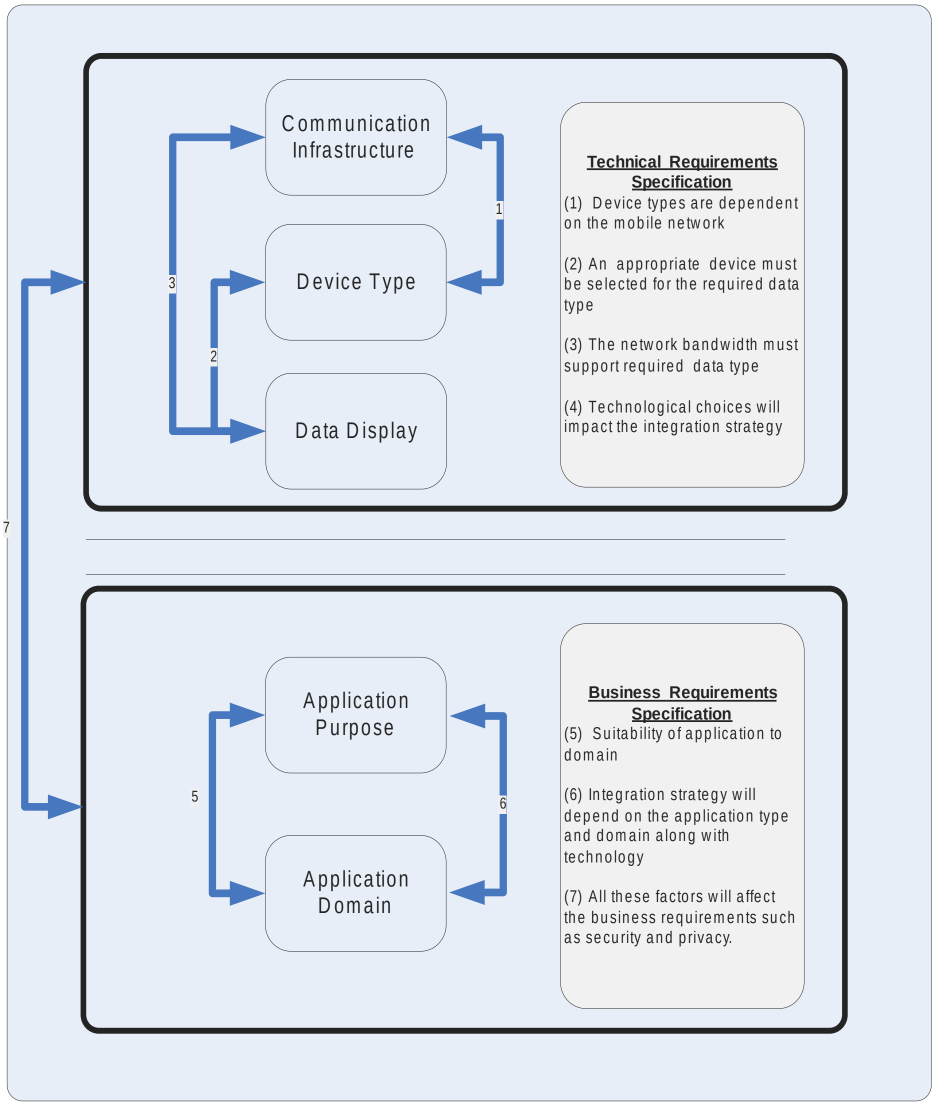
The first dimension defines the mobile infrastructure used to transmit, encode, and receive data. There are numerous wireless infrastructures available from which healthcare providers can choose. Examples of networks that provide personal area networks include Bluetooth and Zigbee. Mobile networks that provide connectivity within buildings include Wireless Local Area Networks (WLAN), which use different protocols from the standard digital mobile technologies such as 2G, 2.5, and 3G, which provide wide area connectivity. A summary of these technologies was provided in "Mobile Healthcare Delivery System Networks."

Device Type

The second section is referred to as Device Type. Due to the advances in medical sensor technologies and handheld computers, M-Health has the potential to deliver services beyond the scope of mobile telemedicine discussed in the previous section. The integration of medical sensors with ICT allows physicians to diagnose, monitor, and treat patients remotely without compromising standards of care. Advances of new materials and signal processing research is allowing the design of smart medical sensors that utilize real-time data recording and processing of multi-physiological signals.

There is a multiplicity of sensors that are available commercially, such as piezo-electrical materials for pressure measurements, infrared sensors for body temperature estimation, and optoelectronic sensors that monitor heart rate and blood pressure. These sensors are being embedded into wearable items and accessories such as sunglasses and rings that can be carried easily. With the continual improvements to the sensors and the miniaturization of computing devices, these wearable devices for monitoring, diagnosing, and treating illnesses are becoming more readily available. Hitachi has developed a wristwatch that can measure the wearer's health conditions such as pulse, moves, and body temperature. The monitoring wristwatch can send data wirelessly via the Internet to remote monitoring services. Hitachi aims to launch the wireless wristwatch commercially within three years (Hitachi's wireless health monitoring wristwatch, 2005), targeting the healthcare service industry for elderly people. The benefit of wearable sensors for physiological data collection is that they can be used to monitor human health constantly without unsettling the patient's day-to-day activities. Examples include rings, vests, watches, wearable intelligent sensors, and systems for e-medicine (Winters, Wang, & Winters, 2003). Wireless communication such as WAN is used for transmitting information

Figure 1. M-health reference model for identifying components of healthcare delivery systems



and accessing healthcare databases wherever appropriate in order to allow free movement of the user (Jovanov et al., 2005).

Personal Digital Assistants (PDAs) have been popular among healthcare practitioners in the last few years. PDA penetration among physicians is at 25%, much higher than the 4% penetration for the general population (Fischer et al., 2003; Holmes, Brown, & Twist, 2001; Harris-Interactive, 2001; Parmanto et al., 2005). Studies have shown that the integration of PDAs into clinical practice has led to decreased medication error rates (Grasso et al., 2002) and the improvement in physicians' adherences to clinical practice guidelines (Shiffman, 1999, 2000).

Although the PDA has great potential to support evidence-based medicine, there are some considerable drawbacks (Parmanto et al., 2005). The small screen is a poor match for information resources designed for full-sized desktop computers. Presenting vast amounts of information in the limited space of the PDA display is a significant technical barrier to the realization of the PDA's potential (Fischer et al., 2003; Larkin, 2001; Peterson, 2002). Other technical PDA drawbacks include low resolution, limited memory, slow processor, and problematic data input. An alternative for the PDA is a tablet PC. These are being adopted in the healthcare sector for information capture at the patient's bedside.

Data Display

This key dimension describes how the information from the M-Health application is to be processed and transmitted. The chosen delivery options can have an important effect on the final quality of the telemedicine event and the outcome. Delivery options in telemedicine can be categorized under two main groups, according to Tulu and Chatterjee (2005): (1) synchronous and (2) asynchronous. Information transactions that occur between two or more participants simultaneously are called synchronous communications. In asynchronous communications, these transactions occur at different points in time. The data displays for both synchronous and asynchronous presentations can be grouped into the following categories: text, data, video, and multimedia (combination).

Application Purpose

The fourth dimension is the Application Purpose. This domain describes the intended use of the application. Field (1996) categorized this field into two main groups: clinical and non-clinical systems.

In 1996, the Committee on Evaluating Clinical Applications of Telemedicine (1996) grouped clinical applications into five categories:

1. Initial urgent evaluation
2. Supervision of primary care
3. Provision of specialty care
4. Consultation
5. Monitoring

Due to rapid advances in the M-Health field, three additional categories were added by Tulu and Chatterjee (2005):

6. Use of remote information and decision analysis resources to support or guide care for specific patients
7. Diagnostic
8. Treatment (surgical and non-surgical)

In addition to these groups, two additional categories have been added by the authors in order to reflect future trends of M-Health systems:

9. Drug delivery
10. Patient identification

The use of M-Health systems for non-clinical purposes includes medical education and administrative duties that do not involve decisions about care for particular patients. Some examples of M-Health non-clinical applications include:

1. Mobile access to the latest drug reference database
2. Bedside access to patient records (increase efficiency by reducing demand for paper records)
3. ePrescribing (mobile prescription writing and verification of drug interactions)
4. Prescription formulary reference (electronically identify most economic pharmaceuticals for a patient)
5. Electronic billing for in-home healthcare workers
6. Patient/drug verification (scan patient and drug bar codes to help to ensure that the appropriate medicine is being administered to the correct patient)
7. Delivery applications (healthcare supply delivery, tracking, and billing)
8. Patient encounter data capture

Application Area

The final dimension is called Application Area. This dimension describes the medical field implementing the M-Health technology. The application area also can be subdivided into clinical and non-clinical use. Clinical use relates to medical departments such as emergency, ophthalmology, pediatrics, and surgery. Non-clinical use relates to maintenance areas such as, charge capture, billing, and administration. The importance of this domain is to highlight the differences in the environments and to identify procedures that are specific to a particular healthcare domain.

In summary, the reference model highlights some of the important elements of an M-Health system. It is important to understand all of the dimensions of the reference model for each mobile healthcare system. Using the previous dimensions to describe an application will allow an M-Health system to be broken down into meaningful components.

The next section presents a case that features the use of an M-health application that uses a PDA to collect private medical data in remote African communities. The case in the next section illustrates how the reference model can be used to derive the components of an M-Health system, which allows issues such as security, privacy, hosting requirements, interference of devices, and integration to be investigated simultaneously.

OVERVIEW OF BLOOD DONOR RECRUITMENT (BDR) PROJECT

As an increased number of healthcare industry professionals adopt mobile-enabled handheld devices to collect, store, and retrieve critical medical information, the need for security has become a top-priority IT challenge. These mobile benefits come with immense corporate and regulatory risk. PDA devices left unsecured while electronic health information is being hotsynced to a PC can become a primary source for the intentional malicious interception of confidential information. Furthermore, the recent adoption of the Patient Privacy and Federal Health Insurance Portability and Accountability Act (HIPPA) is a strong reason not to have protected health information on any unsecured devices, especially a PDA. This section will provide an example of how the reference model described in the previous sections can be used to describe a project initiated by the Red Cross.

The medical informatics data analyst/IT manager of the Uganda Red Cross Society identified a potential weakness in the use of mobile devices in the community health field. One of the authors was consulted to assist in developing a model that addressed data security in order to ensure the main IT security goals: confidentiality, integrity, and availability. Mobile devices were used in Blood Donor Recruitment (BDR) activities to register blood donors' details. The medical history was stored for blood screening purposes. The data are very confidential because some of the blood donors' results are positive for HIV, hepatitis, or syphilis. The system specifications are as follows:

Palm M125 with the following software:

- Palm Desktop software for Windows v. 4.0
- Pendragon form 3.2
- Pendragon Distribution Toolkit 2.0
- Windows XP on the PC
- Microsoft Access 2000
- The volume of data is about 10mb records on the PDA

This information is transferred to the computer server either by wired synchronization (HotSynced) or by wireless network. The hotsync option requires the user to be in the confines of the hospital or office, because it involves the use of the wired cradle. The other alternative that is being considered is the use of a wireless network such as WiFi or GSM to transfer data to the computer. Once the data are stored on the computer, they are manipulated with various software tools.

Using the M-Health reference model described in the previous section, the BDR system was broken down into the various components, as highlighted in Tables 2a and 2b. The following general recommendations were suggested to the client.

1. Upgrade memory on mobile device.
2. Formulate an enforceable end-to-end security policy.
3. Evaluate encryption software that can run both on the PDA and on the PC
4. Implement a solution that ensures that captured data on the PDA can be encrypted before being hotsynced on the PC.
5. Try out software prior to purchase.

There are various software solutions that can be used in order to secure the health system described previously. These solutions vary by encryption method, price, and device compatibility; however, there is no single solution to securing a system. It is important that each scenario be addressed differently, irrespective of how similar they seem.

A software solution that would fulfill the requirements for the scenarios described in Tables 2a and 2b, the PDASecure™by Trust Digital software, provides six different selectable encryption algorithms. Additional protection features include strong password protection to prevent unauthorized synchronizing or beaming of data, unauthorized deletion of files due to viruses or malicious code, and user authentication to devices before data can be decrypted. IT administrators can control who can access data and networks with wireless handheld devices, encrypt 100% of the data, and password-protect devices so that they are useless if stolen or compromised.

Using mobile networks allows data to travel over the open air. When data transmission is wireless, there is a possibility of interception of the radio transmission as well as unauthorized access to the hospital system. Poorly designed WiFi local area networks can be leaky and accessible beyond the intended boundary of use. If an unauthorized access of any kind occurs, it not only could lead to the loss of a patient's privacy but also could lead to other potential consequences.

In Table 2a, the wireless network is not used, and the mobile device must be colocated with the computer system and server in order to transfer and copy data. This option also presents a different set of security problems. Due to the storage capacities of the mobile devices and the growing trend of today's mobile workforce, every port, external disk drive, or JumpDrive can become a security risk. In the corporate environment, IT experts have turned to soldering and gluing USB ports in order to prevent intrusion; they also have installed titanium chastity belts around computers. USB ports and PDA cradles can be used for a variety of functions from input devices as inconspicuous as iPODs, which now are capable of downloading more than 30 gigabits of data.

In summary, using the reference model presented in this article to decompose a Red Cross blood donor system ensured that the security issues were identified and appropriately addressed. The model also was valuable for discovering technical specifications of the system and for understanding what upgrades were required in order to reduce the vulnerabilities of the system. In the future, this model can be improved by the creation of a set of guidelines and standards that are appropriate for a particular type of M-Health system. This will be valuable for governments and private vendors that implement innovative healthcare systems. The model also aids the healthcare stakeholders to identify the technological infrastructure, business requirements, and operational needs for the different types of M-Health systems.

Table 2a. Using the reference model to describe the blood donor system using a cradle solution

System	Purpose	Data	Network	Device	Application Area
Blood Donor Recruitment (BDR)	Blood donor data capture and transfer	Textual data	Fixed cable	Palm PDA	Rural community health
Security Model • Data sent from the PDA and to the PDA must be secured with encryption. • User and server authentication must be in place. • Access to data in the server from a computer or PDA different from where the server runs must be secured with encryption and user / server authentication. • Advances authentication methods such as biometrics will provide added security features.					

Table 2b. Using the reference model to describe the blood donor system using a wireless solution

System	Purpose	Data	Network	Device	Application Area
Blood Donor Recruitment (BDR)	Blood donor data capture and transfer	Textual data	WIFI/ GPRS	Palm PDA	Rural community health
Security Model • Data sent to/from the PDA must be secured with encryption. • User and server authentication must be in place. • No data storage in network operator environment (GPRS/UMTS Operator) • WiFi network must be secured.					

CONCLUSION

In the clinical domain, considerable leaps in the fields of biomedical telemetry, nanotechnology, ICT, and drug delivery techniques will encourage M-Health applications to evolve rapidly over the next decade. These new M-Health applications will take advantage of technological advances, such as device miniaturizations, device convergence, high-speed mobile networks, reduction in power consumption, and improved medical sensors. This will lead to the increased diffusion of clinical M-Health systems, which will impact the reshaping of the healthcare sector.

The non-clinical use of M-Health enterprise systems also will see an increased rate of adoption due to the potential benefits of the mobile solutions. Potential quality improvements may be achieved by implementing platforms that contain functionality that allows organizations to perform clinical results viewing, e-Prescribing, medication administration, specimen collection, charge capture, and physician dictation capture. These systems help to reduce errors by making the patient's medical record and appropriate medical knowledge available at the point of care and at the point of decision.

The reference model presented in this article aids in the breakdown of M-Health systems into the following five key dimensions: (1) Communication Infrastructure — describes mobile telecommunication technologies and networks; (2) Device Type — relates to the type of device being used, such as PDA, sensor, or tablet PC; (3) Data Display — describes how data will be displayed to the user and transmitted, such as images, e-mail, or textual data; (4) Application Purpose — identification of the objective for the M-Health system; and (5) Application Domain — definition of the area in which the system will be implemented.

The reference model described in this article encapsulates the emerging M-Health field and is needed to assist in clarifying the rapid progress in this field. Currently, the M-Health field is disjointed, and it is often unclear what component constitutes an M-Health system. This is addressed by the reference model. Further work is needed in order to identify the appropriate standards and guidelines for implementing the various types of M-Health systems. This would benefit vendors and system implementers by allowing them to understand the various conditions that may apply to a system that uses a specific set of the M-Health reference model variables. The reference model also would benefit from further work that defines how the choice of dimensions can impact the business models and the security policy for the implementation of the system.

REFERENCES

ABI Research. (2005). *WIFI/WIMAX*. Retrieved from http://www.abiresearch.com/category/Wi-Fi_WiMAX

Akyildiz, I., & Rudin, H. (2001). Pervasive computing. *Computer Networks — The International Journal of Computer and Telecommunications Networking, 35*(4), 371-371.

Amy, C., & Richards, G. (2004). A BioMEMS review: MEMS technology for physiologically integrated devices. *Proceedings of the IEEE, 92*(1), 6-21.

Bashshur, R. L. (2002). Telemedicine/telehealth: An international perspective. *Telemedicine and e-Health*, *8*(1), 5-12.

Bashshur, R. L., Reardon, T. G., & Shannon, G. W. (2000). Telemedicine: A new health care delivery system. *Annual Review of Public Health, 21*(1), 613-637.

Bluetooth. (n.d.). Retrieved from http://www.bluetooth.org/

Budinger, T. F. (2003). Biomonitoring with wireless communications. *Annual Review of Biomedical Engineering, 5*, 412.

Committee on Evaluating Clinical Applications of Telemedicine. (1996). *Telemedicine: A guide to assessing telecommunications in health care*.

Daou-Systems. (2001). *Going mobile: From eHealth to mHealth*. Retrieved from http://www.daou.com/emerging/pdf/mHealth_White_Paper_April_2001.PDF

Holmes, B. J., Brown, E. G., & Twist, A. E. (2001, June 25). Doctors connect with handhelds. *The Forrester Report*. Retrieved September 2005, from http://www.forrester.com/ER/Research/Report/Summary/0,1338,11449,00.html

Field, M. J. (1996). *Telemedicine: A guide to assessing telecommunications in health care*. Paper presented at the National Academy Press, Washington, DC.

Fischer, S., Stewart, T. E., Mehta, S., Wax, R., & Lapinsky, S. E. (2003). Handheld computing in medicine. *Journal of the American Medical Informatics Association, 10*(2), 139-149.

Future trends: Convergence is all. (2005, March 12). *NewScientist*, 53. Retrieved January 2006, from http://www.sciencejobs.com/data/pdf/insider/insider129.pdf

Godoe, H. (2000). Innovation regimes, R&D and radical innovations in telecommunications. *Research Policy, 29*(9), 1033-1046.

Goldberg, S., & Wickramasinghe, N. (2002). *Mobilizing healthcare*. Paper presented at the 5th ICECR Conference, Montreal.

Grasso, B. C., Genest, R., Yung, K., & Arnold, C. (2002). Reducing errors in discharge medication lists by using personal digital assistants. *Psychiatric Services, 53*(45), 1325-1326.

Hatcher, M., & Heetebry, I. (2004). Information technology in the future of health care. *Journal of Medical Systems Issue, 28*(6), 673-688.

Health Insurance Portability and Accountability Act of 1996 (HIPPA). (2005). United States Department of Health and Human Services. Retrieved September 30, 2005, from http://www.hhs.gov/ocr/hipaa/

Hitachi's wireless health monitoring wristwatch. (2005). Nikkei-Business-Publications. Retrieved from http://techon.nikkeibp.co.jp/english/NEWS_EN/20050526/105103/?ST=english

IEEE 802.15 Working Group for WPAN. (n.d.). IEEE 802.15. Retrieved from http://www.ieee802.org/15/

Inmarsat announces availability of the 64kbit/s mobile office in the sky. (2000). Inmarsat Swift64. Retrieved from http://www.inmarsat.com/swift64/press_1.htm

Istepanian, R. S. H., & Lacal, J. (2003). *M-Health systems: Future directions*. Paper presented at the 25th Annual International Conference of IEEE Engineering Medicine and Biology, Cancun, Mexico.

Istepanian, R. S. H., & Laxminaryan, S. (2000). UNWIRED, the next generation of wireless and Internetable telemedicine systems [Editorial paper]. *IEEE Transactions on Information Technology in Biomedicine, 4*(3), 189-194.

Istepanian, R. S. H., Laxminarayan, S., & Pattichis, C. (Eds.). (2006). *M-health: Emerging mobile health systems*. London: Springer-Verlag.

Istepanian, R. S. H., & Wang, H. (2003). Telemedicine in UK. In L. Beolchi (Ed.), *European telemedicine glossary of concepts, standards, technologies and users* (5th ed.) (pp. 1159-1165). Brussels, Belgium: European Commission Information Society Directorate-General.

Istepanian, R. S. H., Jovanov, E., & Zhang, Y. T. (2004). M-health: Beyond seamless mobility and global wireless healthcare connectivity [Editorial paper]. *IEEE Transactions on Information Technology in Biomedicine, 8*(4), 405-412.

JAXA. (2005). Aerospace Exploration Agency.

Jovanov, E., Lords, A., Raskovic, D., Cox, P., Adhami, R., & Andrasik, F. (2003). Stress monitoring using a distributed wireless intelligent sensor system. *IEEE Engineering in Medicine and Biology Magazine, 22*(3), 49-55.

Jovanov, E., Milenkovic, A., Otto, C., & Groen, P. C. (2005). A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation. *Journal of NeuroEngineering and Rehabilitation, 2*(6). Retrieved January 17, 2006, from http://www.jneuroengrehab.com/content/2/1/6

Kendall, P., Christopher, T. (2005, May 12). *WiMAX — 3G killer or fixed broadband wireless standard?* Retrieved September 2005, from http://www.strategyanalytics.net/default.aspx?mod=ReportAbstractViewer&a0=2393

Larkin, M. (2001). Can handheld computers improve the quality of care? *Lancet, 358*(9291), 1438.

GSM_Home. (2005). Retrieved from http://www.gsmworld.com/index.shtml

GSM-Information. (2004). Retrieved from http://www.gsmworld.com/index.shtml

Moore, S. K. (2002). Extending healthcare's reach: Telemedicine can help spread medical expertise around the globe. *IEEE Spectrum, 39*(1), 66-71.

Olla, P. (2004). A convergent mobile infrastructure: Competition or co-operation. *Journal of Computing and Information Technology, 12*(4), 309-322.

Olla, P. (2005a). Evolution of GSM network technology. In M. Pagani (Ed.), *Encyclopedia of multimedia technology and networking* (pp. 290-294). Hershey, PA: Idea Group Reference.

Olla, P. (2005b). Incorporating commercial space technology into mobile services: Developing innovative business models. In M. Pagani (Ed.), *Mobile and wireless systems beyond 3G: Managing new business opportunities* (pp. 82-113). Hershey, PA: IRM Press.

Olla, P., & Patel, N. (2003). Framework for delivering secure mobile location information. *International Journal of Mobile Communications, 1*(3), 289-300.

Parmanto, B., Saptono, A., Ferrydiansyah, R., & Sugiantara, W. (2005). *Transcoding biomedical information resources for mobile handhelds*. Paper presented at the 38th Hawaii International Conference on System Sciences, Hawaii.

Pattichis, C. S., Kyriacou, E., Voskarides, S., Pattichis, M. S., Istepanian, R. S. H., & Schizas, C. N. (2002). Wireless telemedicine systems: An overview. *IEEE Transaction of Antennas Propagation Magazine, 44*(2), 143-153.

Personal area network. (n.d.). *IBM mobile computing*. Retrieved from http://www.research.ibm.com/topics/popups/smart/mobile/html/phow.html

Peterson, M. F. T. (2002). *The right information at the point of care library delivery via hand-held computers*. Paper presented at the EAHIL Conference of Health and Medical Libraries, Cologne.

Physicians' use of handheld personal computing devices increases from 15% in 1999 to 26% in 2001. (2001). *Harris Interactive Health Care News, 1*(25). Retrieved September 2005, from http://www.harrisinteractive.com/news/allnewsbydate.asp?NewsID=345

Qiu, R. C., Zhu, W., Zhang, Y. Q. (2002, May 26-29). *Third-generation and beyond (3.5G) wireless networks and its applications*. Paper presented at the IEEE International Symposium on Circuits and Systems (ISCS), Scottsdale, Arizona.

Raskovic, D., Martin, T., & Jovanov, E. (2004). Medical monitoring applications for wearable computing. *Computer Journal, 47*(4), 495-504.

Shiffman, R. N. (1999). *User satisfaction and frustration with a handheld, pen-based guideline implementation system for asthma*. Paper presented at the AMIA Symp.

Shiffman, R. N. (2000). A guideline implementation system using handheld computers for office management of asthma: Effects on adherence and patient outcomes. *Pediatrics, 105*(4), 767-773.

Sorby, I. (2002). Characterising cooperation in the ward — A framework for producing requirements to mobile electronic healthcare records. In *Proceedings of the 2nd Hospital of the Future Conference*, Chicago, IL. Retrieved January 2005, from http://www.hctm.net/events/papers_2002.html

Tachakra, S., Istepanian, R. S. H., Wang, H., & Song, Y. H. (2003). Mobile e-health: The unwired evolution of telemedicine. *Telemedicine and E-Health Journal, 9*(3), 247-257.

Tulu, B., & Chatterjee, S. (2005). *A taxonomy of telemedicine efforts with respect to applications, infrastructure, delivery tools, type of setting and purpose*. Paper presented at the 38th Hawaii International Conference on System Sciences, Hawaii.

UMTS-Forum-Report14. (2002). *Support of third generation services using UMTS in a converging network environment*. UMTS Forum. Retrieved from http://www.umts-forum.org/servlet/dycon/ztumts/umts/Live/en/umts/Resources_Reports_index

Warren, S. (2003). *Beyond telemedicine: Infrastructures for intelligent home care technology*. Paper presented at the Pre-ICADI Workshop

Technology for Aging, Disability, and Independence, London.

Webb, C. E. (2004). Chip shots. *IEEE Spectrum, 41*(10), 48-53.

Weil, N. (2005). Companies announce RFID drug-tracking project: Unisys and SupplyScape plan to track Oxycontin through the supply chain. *Computerworld*. Retrieved January 2006, from http://www.computerworld.com/softwaretopics/erp/story/0,10801,102142,00.html

Wickramasinghe, N., & Misra, S. K. (2005). A wireless trust model for healthcare. *International Journal of Electronic Healthcare, 1*(1), 90-110.

Wickramasinghe, N., & Silvers, J. B. (2003). IS/IT the prescription to enable medical group practices to manage managed care. *Health Care Management Science, 6*(2), 75-86.

Wikipedia. (n.d.). Encyclopedia. Retrieved from http://en.wikipedia.org/wiki

WiMax. (2005). Retrieved September, 2005, from http://wimaxxed.com

Winters, J. M., Wang, Y., & Winters, J. W. (2003). Wearable sensors and telerehabilitation: Integrating intelligent telerehabilitation assistants with a model for optimizing home therapy. *IEEE Engineering in Medicine and Biology Magazine, 1*(22), 56-65.

ZigBee Alliance. (n.d.). Retrieved from http://www.zigbee.org/

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 1, Issue 2, edited by J. Tan, pp. 1-19, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 2.10
Creating a Multimedia Instructional Product for Medical School Students

Mitchell Weisburgh
Academic Business Advisors, LLC, USA

ABSTRACT

Because most medical school textbooks do not adequately address pain management, the American Academy of Pain Medicine wanted to create TOP MED, an online textbook that would address this need for different specialties and which also could be used as a textbook for the Introduction to Pain Management course. This online textbook would cover 11 topics and consist of the latest findings from the most renowned experts in the different disciplines of pain medicine. This case study is a description of the process of designing and producing the online textbook.

INTRODUCTION

The American Academy of Pain Medicine (AAPM) is the medical specialty society representing physicians practicing in the field of pain medicine. Because most medical school textbooks do not adequately address pain management, the academy wanted to create TOP MED, an online textbook that would address this need for different specialties and which also could be used as a textbook for the Introduction to Pain Management course. This online textbook would cover 11 topics and consist of the latest findings from the most renowned experts in the different disciplines of pain medicine.

This case study is a description of the process of designing and producing the online textbook, including how we determined what to cover and how we involved the subject matter experts and translated their content into interactive, entertaining learning segments.

E-LEARNING PROGRAM

The academy had created reference materials online, and they published articles, but this was their first program that was designed for teaching medical professionals. One requirement was that the program have similar production values to television; they did not want something that looked like PowerPoint slides; they did not want talking heads; and they did not want just video. SmartPros, Inc. developed TOP MED for the American Academy of Pain Medicine. The author served as the instructional designer and overall project manager.

ACADEMIC AND ADMINISTRATIVE ISSUES

The academy wanted TOP MED to be an online textbook, not an online class. The typical use would be, say, in a course on pediatrics. When the professor wanted to cover pediatric pain, the students would turn to TOP MED to find out the different ways the children felt pain, how to assess pain in children, how children react differently to drugs, and how pain affects children's, and their family's, lives.

The academy also wanted there to be an assessment at the end of each section so that the professor and the students could determine knowledge acquisition.

The use of video created an interesting problem. The client wanted the course to look as if the video was full screen, but bandwidth considerations prevented the use of full-screen video. One solution could have been to find or build a proprietary solution to serve and access the video. Another could have downloaded the video onto student machines during off-hours. We wanted a more standardized and immediate solution, so we took advantage of a feature in Flash that allowed us to blend video into a Flash animation. We built a virtual "set," which blended in with the video to give the appearance of full screen video without the huge bandwidth requirements.

As a textbook for medical school students, TOP MED has to be authoritative, drawing scientific content from experts. We needed people at the top of their field, individuals who were either conducting or utilizing the latest research. Then we needed to distill and transform their knowledge into lessons for individuals who might become general practitioners, not necessarily pain specialists or researchers.

But if we just wanted to present content, we could have produced a book, audiotape, or video. We wanted to benefit from the unique advantages of a Web-based instructional system, using high quality video, student interaction, assessment and feedback, flexible navigation, tracking, and reporting.

The client wanted the actual lessons to primarily be delivered via video, but they were adamant that the content not be delivered as talking heads with bullet points. They wanted the material to look like full screen video. This is problematic over Internet protocols, because, to be of reasonable production quality, video requires significant bandwidth. We were able to solve this problem by using video embedded in Flash animations. By blending the video in with a digital set, we minimized the size of the video, but the set looked like the video was full screen.

We decided on 12 units, which could later be expanded:

1. Introduction
2. Neurobiology of pain
3. Neuropathic pain
4. Analgesics: NSAIDs and COXIBs
5. Analgesics: Opioids and Adjuvants
6. Patient evaluation
7. Acute and postoperative pain
8. Musculoskeletal pain
9. Cancer pain and palliative care
10. Pediatric pain
11. Misuse and abuse of pain medications
12. Race, culture, and ethnicity in pain management

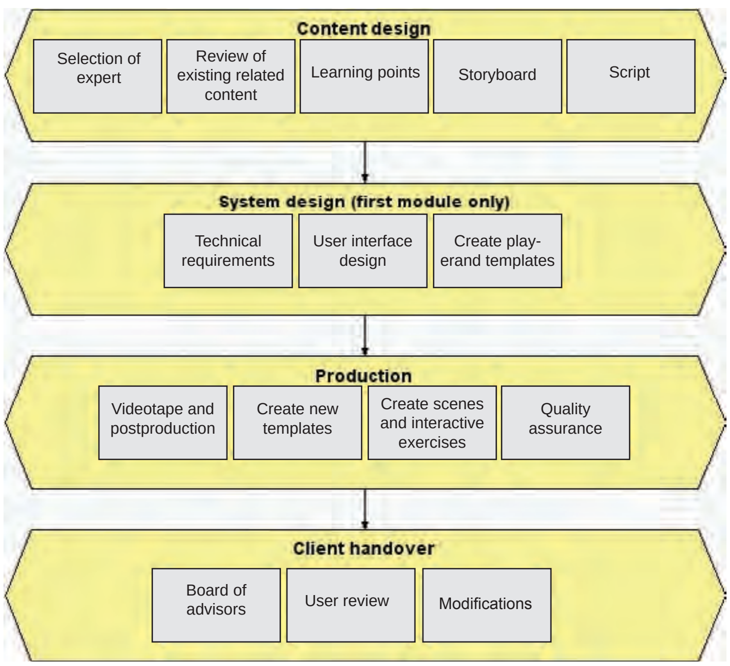
The rationale for sequencing the modules was that we would begin with an introduction to TOP MED and a review of how pain is perceived in the medical community and the population as a whole. The next two units are on how the body reacts to pain and painful stimuli. The two units on analgesics focus on the common medical treatments for pain. Evaluation is necessary for any manifestation of pain. Then the next series of units focus on different common causes and the corresponding treatment for pain. The last two units focus on particular aspects of pain management, which go across all types of pain symptoms and treatments. We started by designing and producing two units: neurobiology of pain and pediatric pain.

Our original goal was that each module should last about a half hour. This was based on initial guidance from the advisory board. We planned the following sections for each module:

1. Introduction: what this topic includes and why it is important.
2. Series of four to eight lessons, each consisting of:
   - 2.1. Introduction to the lesson
   - 2.2. Video lecture with animated, text, and graphic aids
   - 2.3. Summary
   - 2.4. One to three questions, problems, or exercises
   - 2.5. Possible links to supplementary material
3. Conclusion of the topic
   - 3.1. Animated summary of the whole topic with voice
   - 3.2. Quiz of 10 to 20 questions
     - 3.2.1. Each question will have an explanation of how the correct question was arrived at and what learning material was represented
     - 3.2.2. Student responses will be tracked
4. Extra materials
   - 4.1. Printable files of all textual and graphical content
   - 4.2. Glossary with definitions of key terms
   - 4.3. Index of all key topics

Subsequently we decided that the end of module quiz would have a question bank of 30 questions, and that students would get a random

Figure 1. Phases in TOP MED development



sampling of 10 questions each time one took a quiz. Additionally, we combined the index and glossary into a searchable *glindex* that both defines key terms and links to content in the modules.

Our project plan divided our work into four phases. For all modules, there would be a content design phase, a production phase, and a client handover phase. On the first module, we would also perform a system design phase, which would include prototypes that used material from the first module. Figure 1 shows the steps in each phase. The balance of this article will detail how each phase was carried out.

CONTENT DESIGN

Selection of Expert

For this project, the experts were chosen by the client, generally some of the best-known people in the field of pain medicine. In addition to the expert, there was an overall medical editor, who chaired the group that wrote the original federal guidelines on the treatment of pain.

Review of Existing Related Content

AAPM was able to provide videotapes of many of the experts teaching these topics, or at least lecture notes. These videos were generally from 45 minutes to 90 minutes long. The experts also provided copies of materials that they used when they lectured.

Very few of the materials had the production values we needed. There were minimal graphs, grainy unlabeled photographs, and, with one exception, rudimentary animations. But often the content was extremely relevant leading us to conclude that we would have graphic artists recreate and animate graphics and charts.

Learning Points

This phase of development produced an interim and then a final learning points document. The interim document contained a series of learning points with questions for the medical expert, while the final document was used to produce the storyboard.

The instructional designer (ID) first transcribed any video lectures and then outlined the concepts that were covered. There were two key tests for content: (a) would a doctor knowing this information do a better job with a patient (relevance), and (b) if a person did not already know the information, was the way it was presented sufficiently clear to learn from (clarity).

In terms of relevance, if the ID was certain that the information would not help a general practitioner, it was not used. If it was clear from the materials how it was helpful for a doctor, it was included. If there was a question, the information was tagged so that it could be reviewed with the expert. Sometimes the instructional designer could research the topic.

For example, there were concepts that were backed up with intricate descriptions of scientific research. While the concept itself was useful, and the research would have been necessary for people going into pain research, the detailed research explanations were excised from the learning point document, pending approval from one of the content experts.

In terms of clarity, if the explanation was sufficient, it was copied into the learning points. If it was not clear, the subject was researched and/or tagged for discussion with the expert. For example, the following point from the neurobiology of pain module is important to the material and would be understood by those who were already familiar with the topic. But we deemed it too erudite for an introduction to pain management textbook:

There is NMDA receptor-mediated central sensitization that amplifies the input coming both from the injured tissue and also the unharmed tissue surrounding the area of injury.

This statement would be tagged for a more expansive explanation. So that you can see the contrast, refer to Example A for this explanation as it appears in the final script.

And while this text would never make it into a Stephen King best seller, it is something that a second year medical student should be able to understand.

Finally, items were grouped into general topics or sections; all technical terms were defined in footnotes; and a request was made for clinical examples.

The interim learning points document thus had:

- A list of topics that had been covered in lectures or in any additional research.
- Points tagged with, "how is this useful to a doctor?" Or, "why should someone learn this?"
- Other points tagged with, "how else can this be explained?"
- Terms that were tagged with definitions were questioned as, "is an explicit definition necessary in the lecture, or can it just be defined in the glossary?"
- For each topic or section, there was a question, "can you provide a patient history that will illustrate some of the points in this section?"

Once the interim learning points document was assembled, the ID met with the expert face to face. This meeting tended to take about four hours to review all of the questions.

Armed with the answers to the questions and the case studies, the ID would then assemble the final learning point document and submit it to the expert for review. Generally, a one-hour phone meeting was sufficient to review and approve this document.

It also became evident very early on that, for the topics contained in the TOP MED course, modules were going to be 45 to 60 minutes long, not 30 minutes. The learning point documents (Figure 2) tended to run about 20 pages of 12-point font.

Example A.

How does central sensitization occur?

Glutamate appears to play a key role. Glutamate is an amino acid released in excitatory synapses. It affects several types of receptors, both in the spinal cord and in the brain. One of these receptors can be labeled with an artificial reagent, N-methyl D-aspartate (NMDA). That receptor is therefore termed the NMDA receptor.

The NMDA receptor has binding sites for the neurotransmitter glutamate and several other substances that modulate its activity, including glycine, zinc, and various polyamines.

We know that two things have to happen for activation of the NMDA receptor in the spinal cord.

First is the firing of the nociceptor, which causes depolarization and ion flux across its cell membrane, which in turn causes the release of glutamate into the synapse. Prolonged firing by C-fiber nociceptors causes release of glutamate.

Second is the binding of glutamate to the postsynaptic NMDA receptor—keeping in mind that other chemicals also bind and may have additional roles.

Activation of NMDA receptors causes the spinal cord neuron to become more responsive or sensitive to all of its inputs.

Figure 2. Sample learning points page

· *Causes and implications of cancer pain*¶

· Neurology¶

Pain is nociceptive, neuropathic, or both; many cancer patients have a mixture of both¶

Nociceptive¶

- Bone invasion by tumor can be one of the most severe pains associated with metastatic disease.¶
- Infiltration and occlusion of blood vessels can cause ischemia[2] and pain, after which neuropathic pain may be persistent.¶
- Obstruction of a hollow viscous can lead to visceral pain syndromes, which can be quite severe.¶
- Swelling of a structure invested by fascia or periosteum can be very painful.¶
- Necrosis and/or infection of cancerous tissues, with inflammation and ulceration, are tumor-specific pain.¶
- Post-chemotherapy and radiation therapy syndromes are very common, even when a patient is in remission or has been cured.¶
- Any persistent pain syndrome can lead to sensitization and so-called "wind up" of the central nervous, leading to a chronic neuropathic pain state (see TOP MED module on Neurobiology)¶

Neuropathic¶

- Compression or infiltration of nerves, nerve roots or spinal cord is a potent stimulus for acute or persistent neuropathic pain¶
- Concomitant neurological disease, e.g., diabetic neuropathy, post-herpetic neuralgia¶
- Post-surgical, chemotherapy and radiation therapy neuropathies¶

Storyboard

We decided based on the length of the material that we needed to have more than one presenter for each module. We fixed on an intro/review person, an explanatory person, and the medical expert. We also learned that we generally needed to change speakers every minute or so, although a few segments, as long as three minutes, could still maintain interest.

The storyboard document was produced in PowerPoint. The slide showed a stick figure of the actor along with any bullet points or graphics that were going to be displayed, and the speaker notes contained a copy of the learning points that were going to be covered in that scene. If a learning point could be diagrammed, then there would sometimes be a full screen animation along with a voice-over instead of a video.

If there was going to be a graphic, chart, or animation, the ID would create a "wire-frame" from which a graphic artist could work. At the end of each section there were general instructions for an interactive exercise to be completed by the student, generally some type of drag and drop problem. The storyboard was then reviewed with the medical editor. This was also generally a three-hour meeting. The medical editor would typically point out places where the medical expert needed to make his/her sources more explicit, needed more updated references, or, in a few cases, where recent data had modified the assertions.

The comments and storyboard were then reviewed by the medical expert and passed back to the ID. The ID created a final storyboard (Figure 3), which was then reviewed one more time by the editor and medical expert.

Script

The scriptwriting is in three stages. A scriptwriter writes the script. It goes through a review process with the medical editor and instructional designer. Blocking and animation instructions are added. For this project, we used a scriptwriter who had written video scripts for pharmaceutical companies. He provided a fresh perspective to the material. The scriptwriter also wrote any bullets that would appear on the screen and completed

Figure 3. Sample of storyboard page

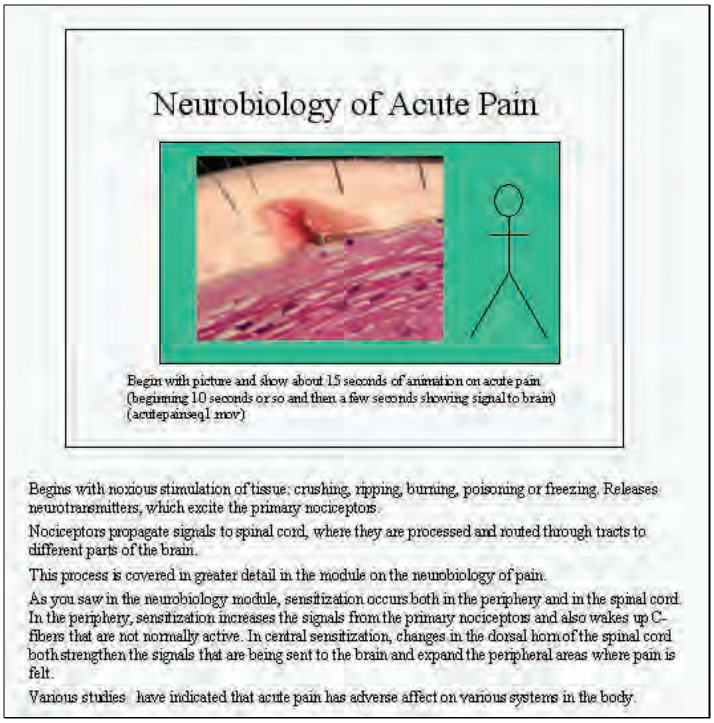
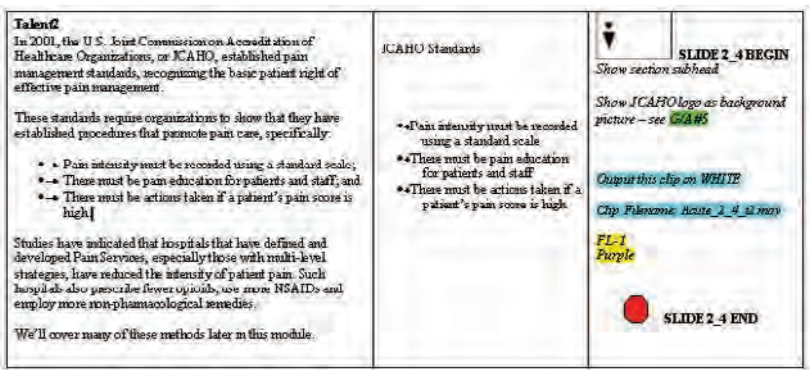


Figure 4. Script sample



Talent2 In 2001, the U.S. Joint Commission on Accreditation of Healthcare Organizations, or JCAHO, established pain management standards, recognizing the basic patient right of effective pain management. These standards require organizations to show that they have established procedures that promote pain care, specifically: • Pain intensity must be recorded using a standard scale; • There must be pain education for patients and staff; and • There must be actions taken if a patient's pain score is high. Studies have indicated that hospitals that have defined and developed Pain Services, especially those with multi-level strategies, have reduced the intensity of patient pain. Such hospitals also prescribe fewer opioids, use more NSAIDs and employ more non-pharmacological remedies. We'll cover many of these methods later in this module.	JCAHO Standards • Pain intensity must be recorded using a standard scale • There must be pain education for patients and staff • There must be actions taken if a patient's pain score is high	**SLIDE 2_4 BEGIN** *Show section subhead* *Show JCAHO logo as background picture – see G/A#5* *Output this clip on WHITE* *Clip Filename: Acute_2_4_t2.mov* *FL-1* *Purple* **SLIDE 2_4 END**

the end of section exercises. We determined that there should be some bullet or some change to the screen every 5 to 10 seconds in order to maintain interest.

The script (Figure 4) was then reviewed by the instructional designer and then the medical editor, and then the production team added animation, bullet, and stage instructions.

SYSTEM DESIGN

Technical Requirements

There were three broad areas that needed to be defined for TOP MED: (1) the player that would be able to display something that looked like nearly full screen video over the Internet; (2) the features of the learning management system (LMS) that would be used for student registration, tracking, and reporting; and (3) the hosting to meet response time and security requirements.

We knew that we would not be able to send a 600x800 pixel, 15 frames per second video over the Internet. It would be too large, requiring too much bandwidth on both the hosting and the user side. On the other hand, we knew that the end product needed a high production value; students and professors would want something comparable to what they see on television. We also knew that the target audience would have access to high-speed Web access, using advanced desktop and laptop computers.

Our conclusion was to use Flash Video. We would tape all segments on a chroma-key background and then blend the video image into the rest of the Flash stage in order to give the appearance of full screen video, cutting the bandwidth requirements by over 75%. Of course, it sounds simpler than it is; but the SmartPros Interactive staff was able to work out all the glitches and make the process relatively straightforward.

The lesson player also needed to be SCORM compliant so that the lessons could be easily ported to different LMSs. We also knew we were going to work with the SmartPros LMS: the Professional Education Center (PEC). The LMS gave us the capabilities to interface with the Flash player, register students, track student progress, provide end of module quizzes, and create custom reports for medical schools and for the AAPM. We needed to define the specific information that needed to be sent back and forth between the LMS and the player and also to specify the key data that the AAPM wanted to maintain.

Hosting presented an interesting conundrum. Between the U.S. and Canada there are over 140 medical schools. What if all the medical schools were using the program and 16,000 medical students needed to access lessons or answer questions at the same time? What would we need for hosting? On the other hand, since we would be starting off with just a few schools for the first semester, why should the client pay for excess capacity that they were not using? On the other hand, the number of students could change rapidly, in as little as a few weeks and with little advance notice.

For us, one of the strong attractions to the SmartPros PEC was that it was already serving hundreds of thousands of users, with bandwidth, processing, and storage to accommodate spikes in demand. If we started with just a few schools and the numbers of students increased by a factor of five three months in a row, it would not be a significant jump in the demands faced by the overall system.

User Interface Design

The user interface design involved both the LMS and the lesson player. For each, we first defined the types of operations a student was likely to perform. From the LMS, a student would need to register, logon, go to a specific module, take a test for a specific module, view results from previous tests, and view descriptions of the different modules. The priority for the LMS was to make these as easy as possible and keep the screen as simple as possible.

A student would be looking at a particular module from the player. From within that module, a student would want to continue, go to a particular section, fast forward or backward, pause, turn off the sound, take a quiz, see results for the previous quiz, look up a word, find where a topic is covered, and go back to the LMS.

We wanted to make these tasks simple; we wanted a unified color scheme and slide layout, but with a few different options to maintain interest. We initially brainstormed the different ways that the features could work and how they would relate to each other. A graphic designer then created screen layouts for five different looks, which were discussed with the client and sample users. Once we had narrowed the look down to two, we created working prototypes using the materials from the first module. These were again discussed with the same groups, and we ended up with our final look and feel.

Player and Template Creation

In order to allow for multiple developers to work on the material, we created the Flash player. All navigation and media controls were built into this player. Inside the player, there was a stage area. The player was designed to run each scene on that stage. The scene developer thus only had to worry about the specific content that was going to be deployed in that scene.

We then created different digital set templates for the different types of scenes: Introduction, Summary, Activity, person on left, person on right. The scene developer would receive a flash video file (flv file), a list of animation elements and/or bullets, and background photographs. The job of the developer was to assemble these source materials and choreograph them with the video. Figure 5 is a sample of the player with a digital set running.

PRODUCTION

Videotape and Post Production

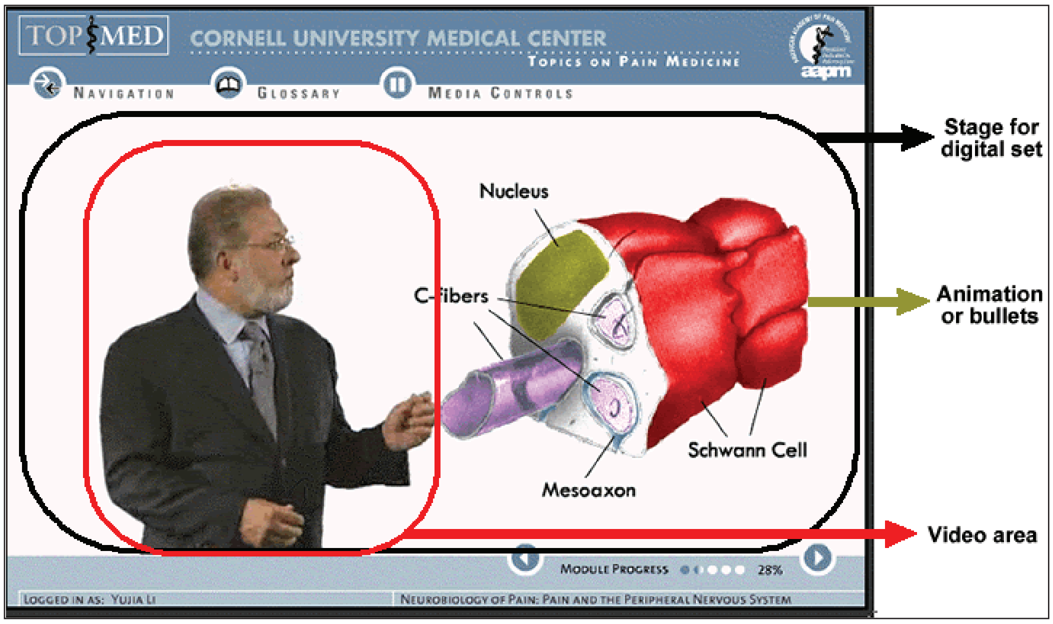
Videotaping was performed in a green room with a teleprompter on the main video camera. Before each scene the actor would be coached on how to enter the frame, when to gesture, and where to stand.

With the medical experts, our challenge was to get them to relax and move freely in front of the camera. With the actors, it was pronunciation of medical terms like parenterally, NSAID, and allodynia. We found that one-hour of video required about one and one half days of shooting. The good takes were then transferred to flash video files and the sound was normalized.

Create New Templates

As we progressed with additional modules, we found we periodically needed to add to the number of templates, for example, person starting on left moving to right, person starting on right moving

Figure 5. Sample player view



to left, and voice over with animation were all added to the original templates.

Create New Scenes and Exercises

We created a styles document and sample scenes that were given to the developers along with the templates and objects that were relevant to them. Before development started we had a kickoff meeting to discuss the purpose and audience of TOP MED, the story for that particular module, the graphics, bullets, and animations for each scene, and how to use the style document and templates.

Quality Assurance

There were two initial rounds of quality assurance for each module, with four individuals going through each scene. For the third round, the medical editor and two representatives of AAPM also participated in the review. The status of all suggestions and findings was maintained in a database that was accessible to all of the QA staff. The project manager was responsible for discussing each finding with the QA individual(s) who found it, verifying it, and then ensuring that the developer understood the changes.

CLIENT HANDOVER

Board of Advisors Review

In February 2005, the board of directors of the American Academy of Pain Medicine viewed the first two modules, and the result was very well received. In fact, one of the directors mentioned that this would be a good prototype for the way medical textbooks should be produced. The general comments were that the length of the modules was appropriate, the content was current, relevant, and focused, and the presentation would maintain the interest of the target student.

User Review

There have been quite a few doctors, residents, and students who have viewed the material. More formal review with medical school classes is also planned.

Modifications

There were three main changes that were requested from our initial feedback. First, since the first impression is going to be from the introduction to each module, we needed to make the intro compelling, and we have added graphics and animations to the background of these sections.

Second, while we initially felt the washed out colors would reduce learner fatigue, the feedback was that we needed a brighter color scheme. We have increased the color contrast for the second and third modules. Third, the initial feedback on the second module was that we needed to add a few animations, which we have done. We also increased our use of animations in the second two modules.

Program Evaluation

A panel of medical experts evaluated all contents for accuracy, currency, and relevance before the program was created. Since the first modules were created, the program has not yet been fully evaluated. We have shown the program to doctors, nurses, and med school students. The feedback has been very positive, but we do have a punch list of changes, most of which are minor. Currently the academy is waiting for the next level of funding to conduct a full evaluation and complete the additional modules.

Networking and Collaboration

TOP MED was designed to be run asynchronously; there is no collaboration built into the program itself. Instead, it is designed as a tool to be used by a professor to provide content to students, with projects to be assigned by the professor as part of the class.

Policy Implications

The advisory board of the AAPM has indicated that they would like to provide more materials in this fashion and that the cost for creating this course was similar to the cost of developing a textbook with similar content. They feel that this style more closely addresses how the current generation of doctors will want to study.

SUSTAINABILITY AND CONCLUSION

TOP MED was designed in such a way that additional modules could be added. Because of the use of video, though, it will be difficult to edit existing modules should methods become obsolete; it is difficult to splice in a new video and have it look continuous.

By having the project sponsored, the AAPM is ensuring that there are no monetary hurdles for students to obtain the content. On the other hand, TOP MED is dependent on the ability of the AAPM to find partners who can afford the approximately $75,000 it takes to create each one-hour module.

Lessons Learned and Best Practices

We thought that the process of defining the content, developing scripts, building in interactivity, and then taping, developing, and editing worked very well. After watching the first two modules, we learned how to make additional transitions to keep the students' attention and the importance of frequent reviews, especially reviews that explain the content in a different manner that when it is first explained.

The three actions that can be identified as best practices are:

1. Determine the needs and content that is needed before doing any of the work. This means pulling the content from a variety of sources and having it checked by content experts.
2. Develop a script with stage and animation instructions before any videotaping or animation development. It is very valuable to have these instructions reviewed by people experienced in animation and video to make sure that what you have envisioned will maintain the interest of the audience and to see if there are alternatives that might be lower cost but just as effective.
3. Flash was a great tool to work with, but you do need experts in Flash, and you need to build in four to five quality assurance rounds. You need to test to see if the product functions and also if it works for the intended audience.

This work was previously published in Cases on Global E-Learning Practices: Successes and Pitfalls, edited by R. Sharma and S. Mishra, pp. 261-275, copyright 2007 by Information Science Publishing (an imprint of IGI Global).

Chapter 2.11
Intelligent Agents Framework for RFID Hospitals

Masoud Mohammadian
University of Canberra, Australia

Ric Jentzsch
Compucat Research Pty Ltd., Canberra, Australia

ABSTRACT

When dealing with human lives, the need to utilize and apply the latest technology to help in saving and maintaining patients' lives is quite important and requires accurate, near-real-time data acquisition and evaluation. At the same time, the delivery of a patient's medical data needs to be as fast and as secure as possible. One possible way to achieve this is to use a wireless framework based on radio-frequency identification (RFID). This framework can integrate wireless networks for fast data acquisition and transmission while maintaining the privacy issue. This chapter discusses the development of an agent framework in which RFID can be used for patient data collection. The chapter presents a framework for the knowledge acquisition of patient and doctor profiling in a hospital. The acquisition of profile data is assisted by a profiling agent that is responsible for processing the raw data obtained through RFID and a database of doctors and patients.

INTRODUCTION

The use and deployment of radio-frequency identification (RFID) is a relatively new area and it has been shown to be a promising technology (Glover & Bhatt, 2006; Lahiri, 2005; Shepard, 2004). This technology has the capability to penetrate and add value to nearly every field, lowering costs while improving service to individuals and businesses. Although many organizations are developing and testing the deployment of RFIDs, the real value of RFID implementation is achieved in conjunction with the use of intelligent systems and intelligent agents. The real issue is how intelligent-agent technologies can be integrated with RFID to be used to achieve the best outcome in business and services areas.

In this research, a new method for integrating intelligent-agent technologies with RFIDs in managing patients' healthcare data in a hospital environment is given. Knowledge acquisition and profiling of patients and doctors in a hospital are assisted by a profiling agent that is responsible for processing the raw data obtained through RFID data that are stored in a hospital database. There are several perspectives for profiling that could be used in a healthcare and hospital environment.

An intelligent agent can assist in profiling patients based on their illness and ongoing diagnostics as reported by the RFIDs. There are certain data and knowledge about each patient in the hospital. This knowledge could be the description of what the patient's symptoms are, monitoring status, and why the patient was admitted to the hospital. Using this information, an evolving profile of each patient can be built.

This data and knowledge can assist in deciding what kind of care he or she requires, the effects of ongoing care, and how to best care for this patient using available resources (doctors, nurses, beds, etc.). The intelligent agent will build a profile of each patient. Along with a profile of each patient, a profile for each doctor can also be developed. Then the patient and doctor profiles can be correlated to find the best doctor to suit the patient.

The patient-doctor profiling can be useful in several situations:

- Providing personalized services to a particular patient, for example, by identifying the services that a patient requires and hence speeding his or her recovery progress in or even out of the hospital.
- Disambiguating a patient's diagnostic based on the patient profile and matching this profile to the available doctor's profile. This may help in matching doctors with the suitable specialization to a patient.
- Providing speedy, reliable reentry of patients into the hospital by having the patients allocated to visit the relevant doctors.
- Presenting information in a way suitable to the doctor's needs, for example, presenting the information about the patients on a continuous basis for the doctors.
- Providing tailored and appropriate care to assist in cost reduction.

Personalization, user modeling, and profiling have been used for many e-commerce applications by IBM, ATG Dynamo, BroadVision, Amazon, and Garden. However, the use of such systems in hospital and personal care and profiling has not been reported. It should be noted that the definitions of personalization, user modeling, and profiling that these companies discuss are not quite the same as our intended meanings.

Many user models try to predict the user's preference in a narrow and specific domain. This works well as long as that domain remains relatively static and as such the results may be limited.

One of the main aims of profiling and user modeling is to provide users with correct and timely responses for their needs. This entails an evolving profile to ensure that as the dynamics of the user and real world change, the profile and user model reflects these changes.

A patient's visit to the hospital can simply be classified as a regular visit, an emergency visit, or an ad hoc appointment (on a need basis). In each of these situations, the needs of the patients are different. During a regular visit, the patient visits the hospital at a regular interval and usually a doctor is assigned to that patient. The patient's profile can assist the patient in a situation where the assigned doctor suddenly becomes unavailable. In this situation, the profile of the patient can be matched with the available doctors with suitable specializations for the needs of the patient. The patient-doctor assignment here is a kind of timetabling problem as we know the profile of the patient and doctors as well as the available doctors. Timetabling of doctors is out of the scope of this research study.

However, in an emergency visit, there is no assigned doctor for such a patient. The doctor in the emergency section of the hospital will provide information about a patient after examination, and a patient profile then can be created. In this case, the intelligent agent can assist the patient by matching the profile of the patient with the doctors suitable for the needs of the patient. Also, the doctors can be contacted in a speedier manner as they are identified and their availability is known.

An appointment visit is very similar to a regular visit, but it may happen only once and therefore the advantages mentioned for regular patients apply here.

We will endeavor to describe several of these, but will expand on one particular potential use of RFIDs in managing patient health data. First some background on RFIDs will be presented in this chapter including some definitions. We will discuss the environment that RFIDs operate in and their relationship to other available wireless technologies such as the IEEE (Institute of Electrical and Electronics Engineers) 802.11b, IEEE 802.11g, and so forth in order to fulfill their requirements effectively and efficiently.

This research is divided into five main sections. The following is based on the patient-doctor profiling and intelligent agents. Then the chapter gives an RFID background that will provide a good description of RFIDs and their components. This section discusses several practical cases of RFID technology in and around hospitals. It lists three possible applicable cases assisting in managing patients' medical data. Next we discuss the important issue of maintaining patients' data security and integrity, and relate that to RFIDs. Finally, the conclusion and further research directions are given.

PATIENT-DOCTOR PROFILING

A profile represents the extent to which something exhibits various characteristics. These characteristics are used to develop a linear model based on the consensus of multiple sets of data, generally over some period of time. A patient or doctor profile is a collection of information about a person based on the characteristics of that person. This information can be used in a decision situation between the doctor, domain environment, and patient. The model can be used to provide meaningful information for useful and strategic actions. The profile can be static or dynamic. The static profile is kept in prefixed data fields where the period between data-field updates is long, such as months or years. The dynamic profile is constantly updated as per evaluation of the situation. The updates may be performed manually or automated. The automated user-profile building is especially important in real-time decision-making systems. Real-time systems are dynamic. These systems often contain data that are critical to the user's decision-making process. Manually updated profiles are at the need and discursions of the relevant decision maker.

The profiling of patients and doctors is based on the patient-doctor information:

- The categories and subcategories of doctor specialization and categorization. These categories will assist in information processing and patient-doctor matching.
- Part of the patient's profile is based on symptoms (past history problems, dietary restrictions, etc.) and can assist in the prediction of the patient's needs specifically.
- The patient's profile can be matched with the available doctor profiles to provide doctors with information about the arrival of patients as well as presentation of the patient's profile to a suitable available doctor.

A value denoting the degree of association can be created from the above evaluation of the doctor-patient profile. The intelligent-agent based on the denoting degrees can identify (and allocate) an appropriate available doctor to the patient.

In the patient-doctor profiling, the intelligent agent will make distinctions in attribute values of the profiles and match the profiles with the highest value. It should be noted that the intelligent agent creates the patient and doctor profiles based on data obtained from the doctors and patient using the following:

- Explicit profiling based on the data entered by hospital staff about a patient.
- Implicit profiling by filling that gap for the missing data by acquiring knowledge about the patient from his or her past visit or other relevant databases if any, and then combining all these data to fill in the missing data. Using legacy data for complementing and updating the user profile seems to be a better choice than implicit profiling. This approach capitalizes on the user's personal history (previous data from previous visits to the doctor or hospital).

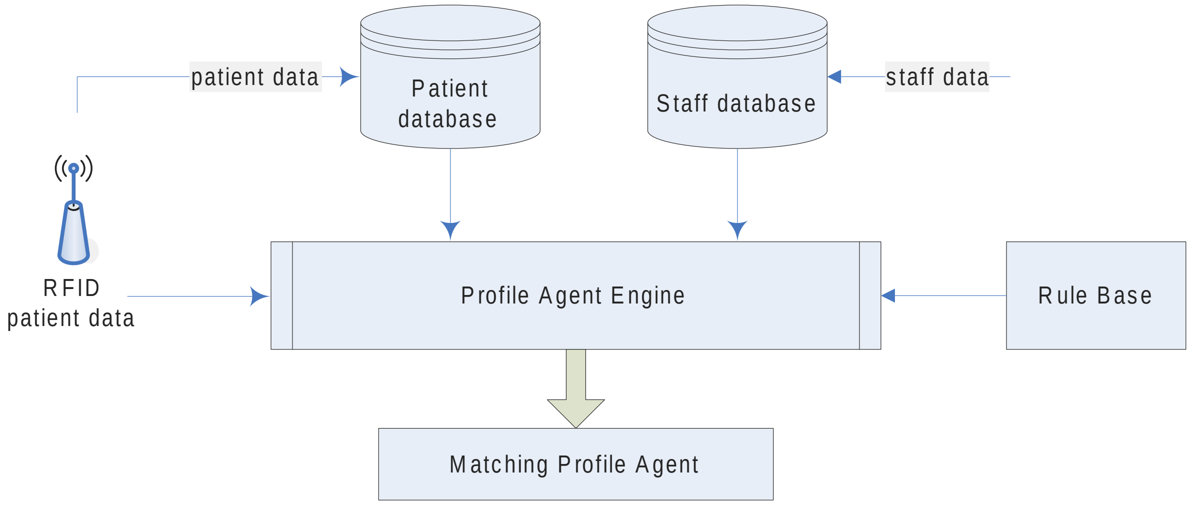
The proposed intelligent-agent system architecture allows user profiling and matching in such a time-intensive and important application. The architecture of the intelligent-agent profiling system using RFID is given in Figure 1.

INTEGRATION OF INTELLIGENT AGENTS, RFID TECHNOLOGIES, AND PROFILE MATCHING

Intelligent agents have been used in order to provide the needed transformation of RIFD passive data collection into an active organizational knowledge assistant. An intelligent agent should be able to act on new data and already stored profile knowledge and thereafter examine its current actions based on certain assumptions. It then inferentially plans its activities. Furthermore, intelligent agents must be able to interact with other agents if required (Bigus & Bigus, 1998; Watson, 1997; Wooldridge & Jennings, 1995) and be able to substitute for a range of human activities in a situated context. In our case, the activities are medical patient assignment and the context is a hospital environment. The integration of RFID capabilities and intelligent-agent techniques provides promising development in the areas of performance improvements in RFID data collection, inference and knowledge acquisition, and profiling operations.

Profile matching is based on a vector of weighted attributes. To get this vector, a rule-based system can be used to match the patient's

Figure 1. Intelligent-agents profiling model using RFID



attributes (stored in the patient's profile) against a doctor's attributes (stored in the doctor's profile). If there is a partial or full match between them, then the doctor will be informed based on his or her availability from the hospital doctor database. The matching is done through the rule-based system by examining the attributes of the patient profile and matching them using the rules already created to the doctor's profile based on the availability of doctors. Figure 1 displays the integration of intelligent agents, RFID technologies, and the profile-matching module.

RFID Description

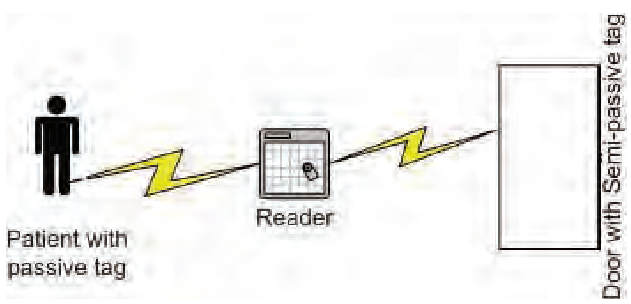
RFID is a progressive technology that has been said to be easy to use and is well suited for collaboration with intelligent agents. Basically, an RFID can be read only, volatile read and write, or write once and read many (WORM) times. RFIDs are non-contact and non-line-of-sight operations. Being non-contact and non-line-of-sight will make RFIDs able to function under a variety of environmental conditions while still providing a high level of data integrity (Glover & Bhatt, 2006; Lahiri, 2005; Shepard, 2004). A basic RFID system consists of four components. These are the RFID tag (sometimes referred to as the transponder), a coiled antenna, a radio-frequency transceiver, and some type of reader for the data collection.

The RFID tag (transponder) emits radio waves in ranges of anywhere from 2.54 centimeters to 33 meters. Depending upon the reader's power output and the radio frequency used, and if a booster is added, that distance can be somewhat increased. When RFID tags pass through a specifically created electromagnetic zone, they detect the reader's activation signal. Transponders can be online or off line and electronically programmed with unique information for a specific application or purpose. A reader decodes the data encoded on the tag's integrated circuit and passes the data to a server for data storage or further processing.

RFID tags can be categorized as active, semi-active, or passive. Each has and is being used in a variety of inventory-management and data-collection applications today. The condition of the application, place, and use determines the required tag type.

Active RFID tags are powered by an internal battery and are typically read-write. Tag data can be rewritten and/or modified as the need dictates. An active tag's memory size varies according to manufacturing specifications and application requirements; some tags operate with up to 5 megabytes of memory. For a typical read-write RFID work-in-process system, a tag might give a machine a set of instructions, and the machine would then report its performance to the tag. This encoded data would then become part of the tagged part's history. The battery-supplied power of an active tag generally gives it a longer read range. The trade-off is greater size, greater cost, and a limited operational life that has been estimated to be a maximum of 10 years, depending upon operating temperatures and battery type.

Figure 2. Semi-passive tag



The semi-active tag comes with a battery. The battery is used to power the tag's circuitry and not to communicate with the reader. This makes the semi-active tag more independent than the passive tag, and it can operate in more adverse conditions. The semi-active tag also has a longer range and more capabilities than a passive tag. Linear bar codes that reference a database to get product specifications and pricing are also data devices that act in a very similar way. Semi-passive tags are preprogrammed, but can allow for slight modifications of their instructions via the reader or interrogator. However, they are bigger, weigh more, and are more complete than passive tags. A reader is still needed for data collection.

Passive RFID tags operate without a separate external power source and obtain operating power generated from the reader. Passive tags have no power source embedded in them and are consequently much lighter than active tags, less expensive, and offer a virtually unlimited operational lifetime. However, the trade-off is that they have shorter read ranges than active tags, and require a higher powered reader. Read-only tags are typically passive and are programmed with a unique set of data (usually 32 to 128 bits) that cannot be modified.

RFID systems can be distinguished by their deployment and frequency range. RFID tags generally operate in two different types of frequencies that make them adaptable for nearly any application. These frequency ranges are as follows.

Low-frequency (30 KHz to 500 KHz) systems have short reading ranges and lower system costs. They are most commonly used in security access, asset tracking, and animal-identification applications.

High-frequency (850 MHz to 950 MHz, or in industry, science, and medical applications, 2.4 GHz to 2.5 GHz) systems offer longer reading ranges (greater than 33 meters) and high reading speeds. These systems are generally used for such applications as railroad-car tracking, container dock and transport management, and automated toll collection. However, the higher performance of high-frequency RFID systems incurs higher system-operating costs.

The coiled antenna is used to emit radio signals to activate the tag and read or write data to it. Antennas are the conduits between the tag and the transceiver that controls the system's data acquisition and communication. RFID antennas are available in many shapes and sizes. They can be built into a door frame, mounted on a tollbooth, or embedded into a manufactured item such as a shaver or software case so that the receiver tags the data from things passing through its zone. Often, the antenna is packaged with the transceiver and decoder to become a reader. The decoder device can be configured either as a handheld or a fixed-mounted device. In large, complex, often chaotic environments, portable or handheld transceivers would prove valuable.

RFID for Hospital Environment

In hospitals, systems need to use rules and domain knowledge that is appropriate to the situation. One of the more promising capabilities of intelligent agents is their ability to coordinate information between the various resources.

In a hospital environment, in order to manage patient medical data, there is a need for both types: fixed and handheld transceivers. Transceivers can be assembled in ceilings, walls, or door frames to collect and disseminate data. Hospitals have become large complex environments.

Nurses and physicians can retrieve the patient's medical data stored in transponders (RFID tags) before they enter a ward or patient's room.

Given the descriptions of the two types of RFID tags and their potential use in hospital patient data management, we suggest the following:

- It would be most useful to embed a passive RFID transponder into a patient's hospital wristband.
- It would be most useful to embed a passive

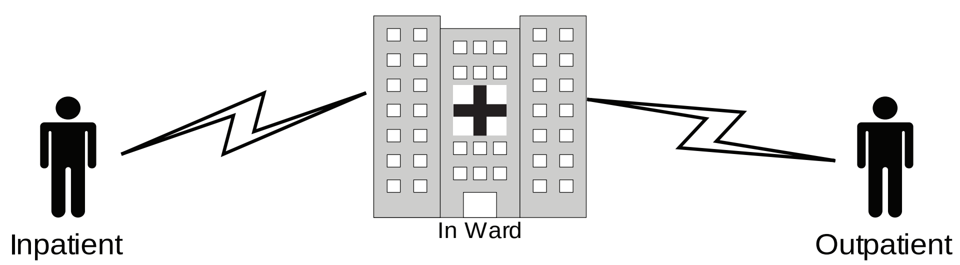
RFID transponder into a patient's medical file.

- Doctors should have PDAs (personal digital assistants) equipped with RFID or some type of personal area network device. Either would enable them to retrieve some patient information whenever they are near the patient instead of waiting until the medical data is pushed to them through the hospital server.
- Active RFID tags are more appropriate for the continuous collection of patient medical data since the patient's medical data need to be continuously recorded to an active RFID tag and an associated reader needs to be employed. Using an active RFID means that the tag will be a bit bulky because of the needed battery for the write process, and there is a concern about radio-frequency emissions. Thus, it is felt that an active tag would not be a good candidate for the patient wristband. However, if the patient's condition is to be continuously monitored, the collection of the data at the source is essential. The inclusion of the tag in the wristband is the only way to record the medical data on a real-time basis using the RFID technology. As more organizations get into the business of manufacturing RFIDs and the life and size of batteries decrease, the tag size will decrease and this may be a real possible use.
- Passive RFID tags can be used as well. These passive tags can be embedded in the doctors PDA, which is needed for determining their locations whenever the medical staff requires them. Also, passive tags can be used in patients' wristbands for storage of limited amount of data on an off-line basis, for example, information such as the date of hospital admission, medical record number (MRN), and so forth.

Low-frequency-range tags are suitable for the patients' wristband RFID tags since it is expected that the patients' bed will not be too far from an RFID reader. The reader might be fixed over the patient's bed, in the bed itself, or over the door frame. The doctor using his or her PDA would be aiming to read the patient's data directly and within a relatively short distance. High-frequency-range tags are suitable for the physician's tag implanted in the PDA. As physicians move from one location to another in the hospital, data on their patients could be continuously being updated.

Finally, the transceivers and interrogators can differ quite considerably in complexity, depending upon the type of tags being supported and the application. The overall function of the application is to provide the means of communicating with the tags and facilitating data transfer. Functions performed by the reader may include quite sophisticated signal conditioning and parity error checking and correction. Once the signal from a transponder has been correctly

Figure 3. Patient and outpatient



received and decoded, algorithms may be applied to decide whether the signal is a repeat transmission, and may then instruct the transponder to cease transmitting or temporarily cease asking for data from the transponder. This is known as the command-response protocol and is used to circumvent the problem of reading multiple tags over a short time frame. Using interrogators in this way is sometimes referred to as hands-down polling. An alternative, more secure but slower tag polling technique is called hands-up polling. This involves the transceiver looking for tags with specific identities, and interrogating them in turn. A further approach may use multiple transceivers multiplexed into one interrogator.

Hospital patient data management deals with sensitive and critical information (patients' medical data). Hands-down polling techniques in conjunction with multiple transceivers that are multiplexed with each other form a wireless network. The reason behind this choice is that there is a need for high-speed transfer of medical data from medical equipment to or from the RFID wristband tag to the nearest RFID reader and then through a wireless network or a network of RFID transceivers or LANs (local area networks) to the hospital server. From there it is a short distance to be transmitted to the doctor's PDA, a laptop, or a desktop through a WLAN (wireless LAN) IEEE 802.11b or 803.11g, or a wired LAN, which operates at the 5.2 GHz band with a maximum data transfer rate of 54 Mbps.

The hand-down polling techniques, as previously described, provide the ability to detect all detectable RFID tags at once (i.e., in parallel), preventing any unwanted delay in transmitting medical data corresponding to each RF tagged patient.

Transponder programmers are the means by which data are delivered to WORM and read-write tags. Programming can be carried out off line or online. For some systems, re-programming may be carried out online, particularly if it is being used as an interactive portable data file within a production environment, for example. Data may need to be recorded during each process. Removing the transponder at the end of each process to read the previous process data and to program the new data would naturally increase process time and would detract substantially from the intended flexibility of the application. By combining the functions of a transceiver and a programmer, data may be appended or altered in the transponder as required without compromising the production line.

Practical Cases using RFID Technology

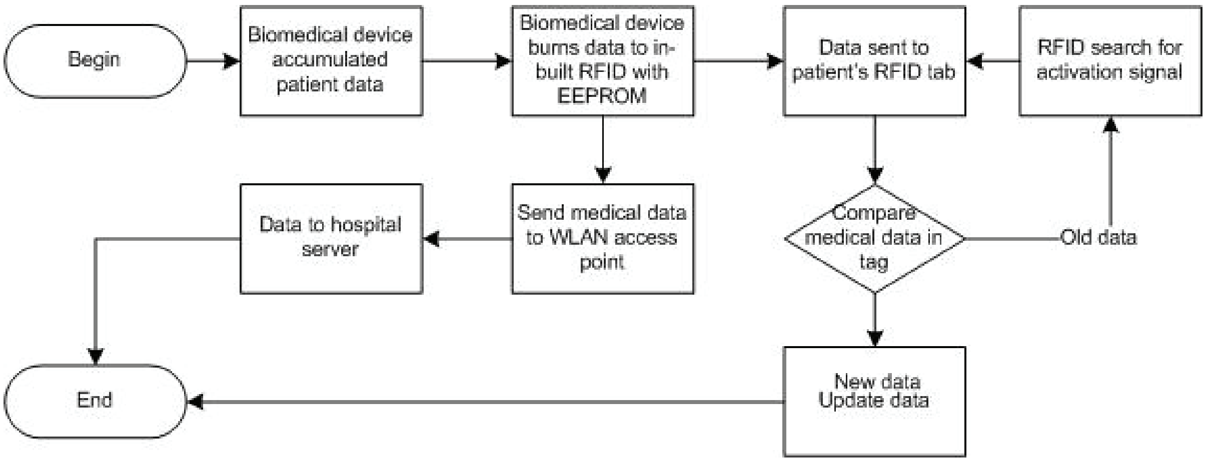
This section explains in detail three possible applications of the RFID technology in three applicable cases. Each case is discussed step by step and then represented by a flowchart. These cases cover issues of the acquisition of a patient's medical data, locating the nearest available doctor to the patient's location, and how doctors stimulate the patient's active RFID tag using their PDAs in order to acquire the medical data stored in it.

Case 1: Acquisition of Patient's Medical Data

Case 1 will represent the method of acquisition and transmission of medical data. This process can be described in the following points as follows:

- A biomedical device equipped with an embedded RFID transceiver and programmer will detect and measure the biological state of a patient. This medical data can be an ECG (electrocardiogram), EEG (electroencephalogram), BP (blood pressure), sugar level, temperature, or any other biomedical reading.
- After the acquisition of the required medical data, the biomedical device will write this data to the RFID transceiver's EEPROM (electronically erasable programmable

Figure 4. Acquisition of patient data



read-only memory) using the built-in RFID programmer. Then the RFID transceiver with its antenna will be used to transmit the stored medical data in the EEPROM to the EEPROM in the patient's transponder (tag) that is around his or her wrist. The data received will be updated periodically once new fresh readings are available by the biomedical device. Hence, the newly sent data by the RFID transceiver will be accumulated with the old data in the tag. The purpose of the data stored in the patient's tag is to make it easy for the doctor to obtain medical information regarding the patient directly via the doctor's PDA, tablet PC (personal computer), or laptop.

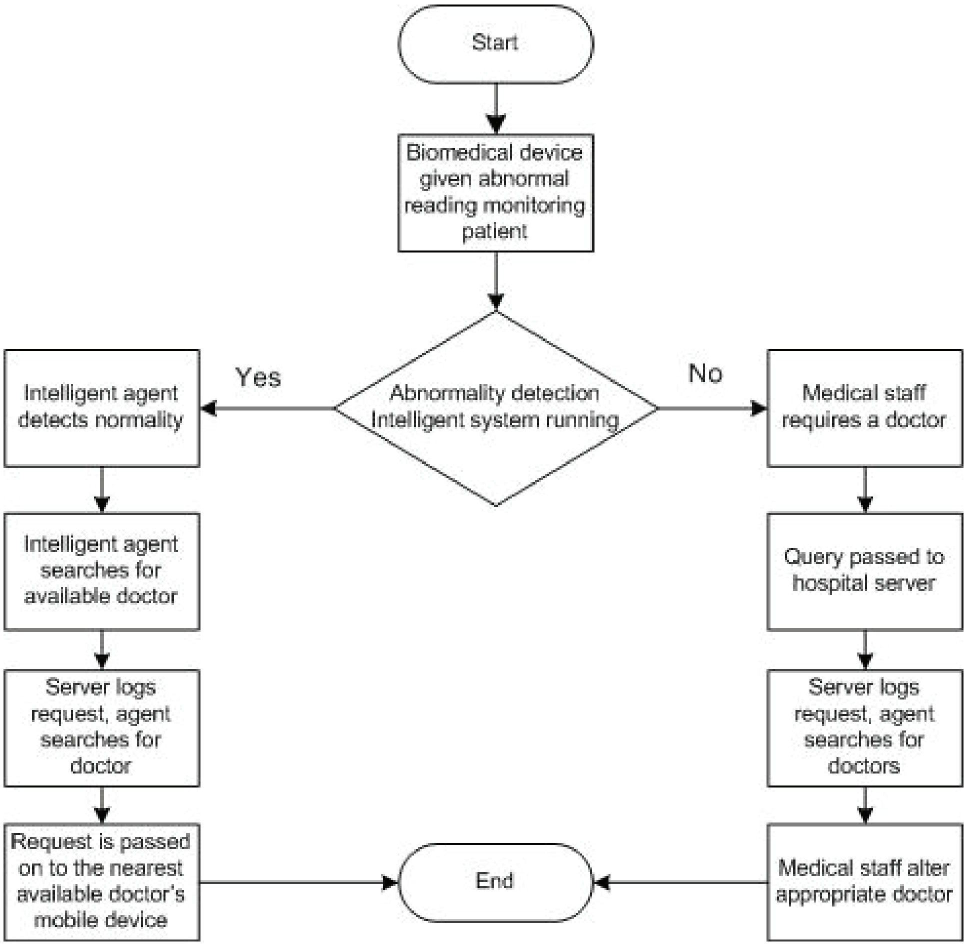
- Similarly, the biomedical device will also transfer the measured medical data wirelessly to the nearest WLAN access point. Since a high data-transfer rate is crucial in transferring medical data, IEEE 802.11b is recommended for the transmission purpose.
- The wirelessly sent data will then be routed to the hospital's main server to be sent (pushed) to the following:
  - Other doctors available throughout the hospital so they can be notified of any newly received medical data.
  - An online patient-monitoring unit or a nurse's workstation within the hospital.
  - An expert (intelligent) software system running on the hospital server to be then compared with other previously stored abnormal patterns of medical data and to raise an alarm if any abnormality is discovered.
- Another option could be using the embedded RFID transceiver in the biomedical device to send the acquired medical data wirelessly to the nearest RFID transceiver in the room. Then the data will travel simultaneously in a network of RFID transceivers until reaching the hospital server.

Case 2: Locating the Nearest Available Doctor to the Patient's Location

This case will explain how to locate the nearest doctor, who is needed urgently, to attend an emergency medical situation. This case can be explained as follows:

- If a specific surgeon or physician is needed in a specific hospital department, the medical staff in the monitoring unit (e.g., nurses) can query the hospital server for the nearest available doctor to the patient's location. In our framework, an intelligent agent can perform this task.

Figure 5. Locating nearest doctor



- The hospital server traces all doctors' locations in the hospital through detecting the presence of their wireless mobile device, for example, PDA, tablet PC, or laptop in the WLAN range.
- Another method that the hospital's server can use to locate the physicians is making use of the RFID transceivers built into the doctor's wireless mobile device. Similar to the access points used in WLAN, RFID transceivers can assist in serving a similar role of locating a doctor's location. This can be described in three steps.
    - The fixed RFID transceivers throughout the hospital will send a stimulation signal to detect other free RFID transceivers, which are in the doctors' PDAs, tablets, laptops, and so forth.
    - All free RFID transceivers will receive the stimulation signal and respond back with an acknowledgement signal to the nearest fixed RFID transceiver.
    - Finally, each free RFID transceiver cell position would be determined by locating to which fixed RFID transceiver range it belongs to or currently is operating with.
- After the hospital server locates the positions of all available doctors, the RFID determines the nearest requested (condition evaluation) physican (pediatries, neuologist, etc...) to the patient's location.

- Once the required physician is located, an alert message will be sent to his or her PDA, tablet PC, or laptop indicating the location to be reached immediately. This alert message would show the following.
  - Case profile over application period
  - The building, floor, and room of the patient (e.g., 3C109).
  - The patient's case (e.g., heart stroke, arrhythmia, etc.).
  - A brief summary description of the patient's case.
- If the hospital is running an intelligent agent on its server as described in our proposed framework, the process of locating and sending an alert message can be automated. This is done through comparing the collected medical data with previously stored abnormal patterns of medical data, then sending an automated message describing the situation. This system could be used in the patient-monitoring unit or the nurse's workstation, who observe and then send an alert message manually.

Case 3: Doctors Stimulate a Patient's Active RFID Tag using Their PDAs in Order to Acquire the Medical Data Stored in It

This method can be used in order to get rid of medical files and records placed in front of the patient's bed. Additionally, it could help in preventing medical errors (reading the wrong file for the wrong patient) and could be considered an important step toward a paperless hospital.

This case can be described in the following steps:

- The doctor enters the patient's room or ward. The doctor wants to check the medical status of a certain patient. Instead of picking up the hard-copy paper medical file, the doctor interrogates the patient's RFID wrist tag with his or her RFID transceiver equipped in a PDA, tablet PC, laptop, or so forth.
- The patient's RFID wrist tag detects the signal of the doctor's RFID transceiver coming from his or her wireless mobile device and replies back with the patient's information and medical data.
- If there was more than one patient in the ward possessing RFID wrist tags, all tags can respond in parallel using hands-down polling techniques back to the doctor's wireless mobile device.
- Another option could be that the doctor retrieves only the patient's number from the passive RFID wrist tag. Then, through the WLAN, the doctor could access the patient's medical record from the hospital's main server.

RFID technology has many potential important applications in hospitals, and the three cases discussed are real practical examples. Two important issues can be concluded from this section. WLAN is preferred for data transfer given that IEEE's wireless networks have much faster speed and greater coverage area as compared to RFID transceiver and transponder technology. Yet, RFID technology is the best for data storage and locating positions of medical staff and patients as well.

Here there is a need for the RFID transceiver and programmer to be embedded in biomedical devices for data acquisition and dissemination. A RFID transceiver embedded in a doctor's wireless mobile device enables the doctor to obtain medical data. With the progress of the RFID technology, it could become a standard as other wireless technologies (Bluetooth, for example) and eventually manufacturers will be building them in electronic devices, or biomedical devices, in our case.

Maintaining Patients' Data Security and Integrity

Once data are transmitted wirelessly, security becomes a crucial issue. Unlike in wired transmission, wirelessly transmitted data can be easily sniffed out, leaving the transmitted data vulnerable to many types of attacks. For example, wireless data could be easily eavesdropped on using any mobile device equipped with a wireless card. In worst cases, wirelessly transmitted data could be intercepted and then possibly tampered with, or in best cases, the patient's security and privacy F be compromised. Hence emerges the need for data to be initially encrypted from the source.

This section of the chapter discusses how we could apply encryption to the designed wireless framework that was explained in the previous section, suggesting exactly where data need to be encrypted and\or decrypted depending on the case that is being examined.

Next, a definition of the type of encryption that would be used in the design of the security (encryption or decryption) framework will be discussed, and this will be followed by a flowchart demonstrating the framework in a step-by-step process. There are two layers of encryption that are recommended to be used.

- **Physical:** (hardware) layer encryption means encrypting all collected medical data at the source or hardware level before transmitting them. This ensures that the patient's medical data would not be compromised once exposed to the outer world on its way to its destination. Even if a person with a malicious intent and also possessing a wireless mobile device steps into the coverage range of the hospitals' WLAN, this intruder will gain actually nothing since all medical data are encrypted, making all intercepted data worthless.
- **Application:** (software) layer encryption means encrypting all collected medical data at the destination or application level once receiving it. Application-level encryption runs on the doctor's wireless mobile device (e.g., PDA, tablet PC, or laptop) and on the hospital server. Once the medical data are received, they will be protected by a secret password (encryption or decryption key) created by the doctor who possesses this device. This type of encryption would prevent any person from accessing the patient's medical data if the doctor's wireless mobile device gets lost, or even if a hacker hacks into the hospital server via the Internet, intranet, or some other means.

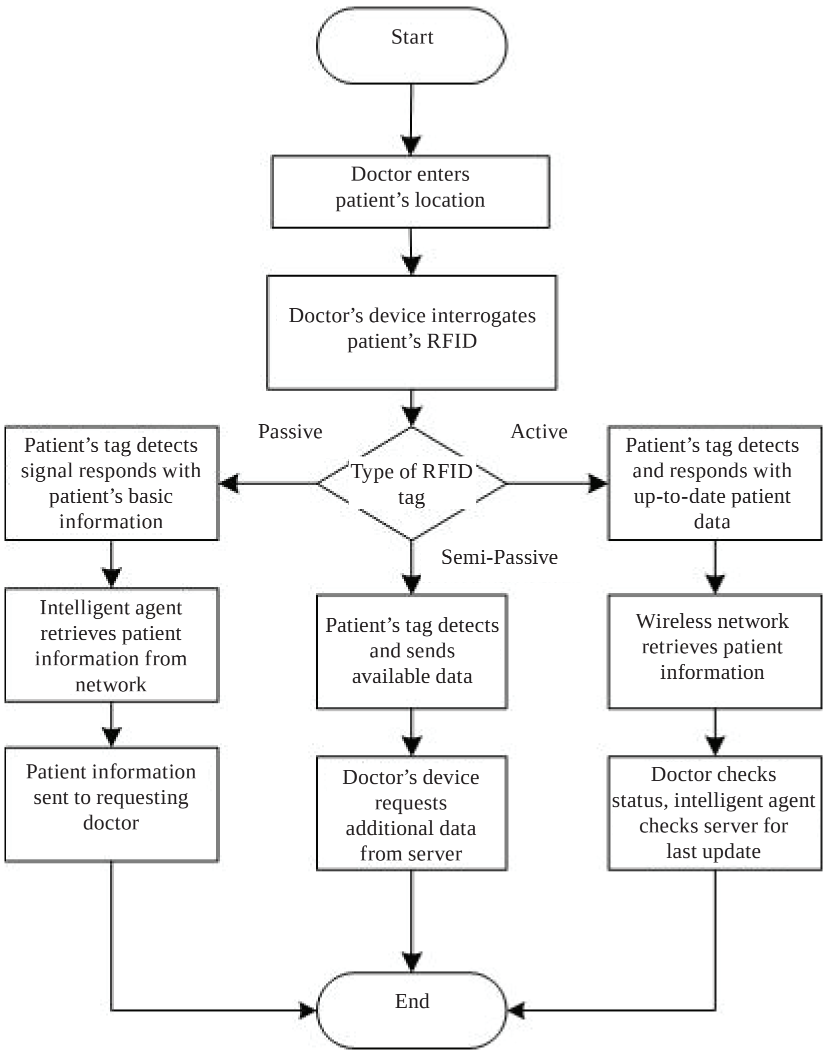
Framework of Encrypting a Patient's Medical Data

The previous section (practical cases using RFID technology) focused on how to design a wireless framework to reflect how a patient's medical data can be managed efficiently and effectively leading to the elimination of errors, delays, and even paperwork. Similarly, this section will focus on the previously discussed framework from a security perspective, attempting to increase security and data integrity during the acquisition of a patient's medical data and when doctors stimulate the patient's active RFID tag using their wireless mobile devices in order to acquire the medical data stored in it. The third case on locating the nearest available doctor to the patient's location is more concerned about locating doctors than transferring a patient's data and it is not discussed here.

The lower part of Figure 6 represents the physical (hardware) encryption layer. This part is divided into two sides. The left side demonstrates the case of a doctor acquiring a patient's medical data via a passive RFID tag located in a band around the patient's wrist. The passive RFID tag contains only a very limited amount of information such as the patient's name, date of admission to the hospital, and above all his or her MRN, which will grant access to the medical record containing the acquired medical data and other information regarding the patient's medical condition. This process is implemented in six steps, and involves two pairs of encryption and decryption. The first encryption occurs after the doctor stimulates the RFID passive tag to acquire the patient's MRN, so the tag will encrypt and reply back with the MRN to the doctors PDA, for example. Then the doctor will decrypt the MRN and use it to access the patient's medical record from the hospital's server. Finally, the hospital server will encrypt and reply back with the medical record, which will be decrypted once received by the doctor's PDA.

The right side of Figure 6 represents a similar case, but this time using an active RFID tag. This process involves only one encryption and decryption. The encryption happens after the doctor stimulates the active RFID tag using a PDA that has an equipped RFID transceiver, so the tag replies with the medical data encrypted. Then the received data is decrypted through the doctor's PDA. The upper part of Figure 6 represents the application encryption layer requiring the doctor to enter a password to decrypt and then access the stored medical data. So, whenever the doctor wants to access a patient's medical data, he or she simply enters a certain password to get

Figure 6. Functional flow



access to either the wireless mobile device or a hospital server, depending on where the medical data actually resides.

Securing medial data seems to be uncomplicated, yet the main danger of compromising such data comes from the people managing it, for example, doctors, nurses, and other medical staff. We have seen that even though the transmitted medical data are initially encrypted from the source, doctors have to run application-level encryption on their wireless mobile devices in order to protect the important data if the device gets lost, left behind, stolen, and so forth. Nevertheless, there is a compromise. Increasing security through using multiple layers and increasing the length of encryption keys decreases the encryption-decryption speed and causes unwanted time delays, whether we are using application- or hardware-level encryption. As a result, this could delay medical data being sent to doctors or online monitoring units. Figure 6 represents the case of high and low levels of security in a flowchart applied to the previously discussed two cases.

We conclude that there are two possible levels of encryption: the software level (application layer) or hardware level (physical layer) depending on the level of security required. Both physical-layer and application-layer encryption are needed in maintaining collected medical data on hospital servers and doctors' wireless mobile devices. Encrypting medical data makes the process of data transmission slower, while sending data unencrypted is faster. Here there is a need to compromise between speed and security. For medical data, it has to be sent as fast as possible to medical staff, yet the security issue has the priority.

CONCLUSION

Managing patients' data wirelessly (paperless) can prevent errors, enforce standards, make staff more efficient, simplify record keeping, and improve patient care. In this chapter, both passive and active RFID tags were used in acquiring and storing medical data, and in linking to the hospital's server via a wireless network. Moreover, three practical applicable RFID cases were discussed and it was explained how the RFID technology can be put in use in hospitals while at the same time maintaining the acquired patient data's security and integrity.

This research in the wireless medical environment introduces some new ideas in conjunction with what is already available in the RFID technology and wireless networks. The aim here is to link both technologies with each other to achieve the research's main goal, which is delivering patient medical data as fast and secure as possible to pave the way for future paperless hospitals.

REFERENCES

Bigus, J. P., & Bigus, J. (1998). *Constructing intelligent agents with Java: A programmers guide to smarter applications.* New York: Wiley. (ISBN:0-471-19135-3)

Glover, B., & Bhatt, H. (2006). *RFID essentials.* O'Railly Media Inc. (ISBN:0-596-0094-4)

Kaptelinin, V. (1992). Human computer interaction in context: The activity theory perspective. In *Proceedings of the East-West Human Computer Interaction Conference*, St. Petersburg, Russia.

Lahiri, H. (2005). *RFID sourcebook.* New York: IBM Press. (ISBN:0-131-8513-73)

Odell, J. (Ed.). (2000). *Agent technology* (OMG Document 00-09-01). OMG Agents interest Group.

Shepard, S. (2004). *RFID.* McGraw-Hill. (ISBN: 0-071-4429-95)

Watson, M. (1997). *Intelligent Java applications for the Internet and intranets.* Morgan Kaufmann. (ISBN: 1-55860-420-0)

Wooldridge, M., & Jennings, N. (1995). Intelligent agents: Theory and practice. *The Knowledge Engineering Review, 10*(2), 115-152. (ISBN: 0-269-888-9)

This work was previously published in Web Mobile-Based Applications for Healthcare Management, edited by L. Al-Hakim, pp. 316-334, copyright 2007 by IRM Press (an imprint of IGI Global).

Chapter 2.12
Design and Development of Standards (HL7 V3) Based Enterprise Architecture for Public Health Programs Integration at the County of Los Angeles

Abdul-Malik Shakir
CAL2CAL Corporation, USA

Debashish Mittra
CAL2CAL Corporation, USA

David Cardenas
L.A. County Department of Public Health, USA

Arindam Basu
CAL2CAL Corporation, USA

Gora Datta
CAL2CAL Corporation, USA

Rini Verma
CAL2CAL Corporation, USA

ABSTRACT

Public Health (PH) applications in County of Los Angeles (LAC), Department of PH have been developed to meet individual PH program's goals. This resulted in lack of county-wide PH data integration, efficiency, and usefulness. LAC encouraged the development of web-based applications utilizing standards-based PH Information Network interoperable service-oriented architecture (SOA). The goal was to stop the evolution of fragmented health data systems which place limitations on the PH mission of safeguarding and improving the health of the community as well as responding to large-scale threats to PH. PH Nursing case management is one example of LAC's initiative for promotion of web-based tools which will be utilized within this SOA. This PH architecture is capable of supporting electronic data exchange from PH partners using a HL7 integration hub. It promotes the development of management tools and applications to assist PH response and recovery activities while providing resources to support departmental integration.

INTRODUCTION

Public health applications in the County of Los Angeles, Department of Public Health have been traditionally developed to assist public health programs in meeting individual program's goals and objectives. This led to the development of systems that collect data only for a specific program area leading to a lack of county-wide public health data integration, efficiency, and usefulness. The existing Program Area Modules (PAMs) do not conform to the National Electronic Disease Surveillance System (NEDSS) nor Public Health Information Network (PHIN) standards recommended by the Centers for Disease Control and Prevention's (CDC, US Govt.) for an interoperable information systems in organizations that participate in public health (Public Health Information Network, 2004; National Electronic Disease Surveillance System, 2004).

The PAMs use a variety of user interfaces, business logic, directory structures, workflow, and data storage systems and are not accessible outside public health firewall. The user and electronic interface for these modules have been implemented differently and do not present a common user or electronic interface for these modules, thereby significantly increasing the cost of data collection and analysis. These modules exist as stand-alone entities each with functionality for disease reporting, incident management, and data storage for specific program areas, and in most cases do not communicate at all with each other or external healthcare partners. A limited number of modules have ad-hoc implementation for electronic data exchange with laboratories as well as some other capabilities needed by the application users.

It was determined that under the standards based architecture, PAMs only handle specialized functionality for each specific area of health care or resource management. All common functionality like user authentication, directory serices, data storage, and so on should be handled by Common Area Modules (CAMs), modules common to all PAMs. An internal message and data broker module will be used for electronic data transfer with partner systems. This initiative will also require that existing investments in legacy systems be leveraged and merged with standards-based Web enabled systems to provide a synchronized view of public health data and resources across all program areas.

The nursing practice management system (NPMS) case management PAM is an example of the Los Angeles County's initiative in promotion of Web-based tools which is to be utilized within the interoperable standards based architecture. This PAM has won several awards since it was implemented including the Los Angeles County Public Health Innovation Award in 2003 and the National Association of Counties Nursing Achievement Award in 2004 (see Figure 1).

The public health nursing unit of Los Angeles County promotes the well being of the community at large and prevents disease, disability, and premature death. Public health nurses (PHNs) make home visits to families with communicable diseases; emphasize nutrition, disease prevention, and health; and improve the quality of neighborhood life by working in partnership with the community. This system tracks multiple interventions given over a period of time by index case, family, and household. It provides a powerful and secure departmental work-flow, assignment/task management and reporting tool with an intuitive interface design that minimizes training and transition costs. The "open" architecture allows the system to communicate with other Public Health programs. This system involves epidemiological data entry during assessments of cases by Public Health Nurses and serves as an efficient tracking system for nurses and supervisors, ensuring that members of the community are assessed and that interventions are made in a timely manner.

The investment into the approach for an enterprise architecture within an existing infrastructure was done with the view that investment of pre-

Figure 1. The Public Health Nursing Unit of the Department of Public Health, Los Angeles County promotes the well being of the community at large and prevents disease, disability, and premature death. This system tracks multiple interventions given over a period of time by index case, family, and household. The "open" architecture allows the system to communicate with other Public Health programs.

existing assets which include people, facilities, and services must now also include information technology as an asset. It was important for County personnel to determine that services must be governed similar to business returns in the corporate environment for maximal return on investment. The decision to implement this architecture was done keeping in mind that there would be large scale reuse of preexisting programs, improved flexibility for change management and improved comprehensibility of software functionality for a clear relationship between services and function (Sprott, 2004)

To help alleviate the evolution of these separate, fragmented public health data systems which over a period of time place limitations on local public health agencies to accomplish the mission of safeguarding and improving the health of the community and to respond effectively to large-scale threats to public health, the County of Los Angeles has been promoting the development of Web-based applications which will utilize a standards-based interoperable service-oriented architecture (SOA).

Service-oriented modeling and architecture was announced by IBM as the first SOA-related methodology in 2004 (IBM, 2005; Arsanjani, 2004). Since 2004, efforts have been made to move towards greater standardization and the involvement of business objectives.

SOA can also be regarded as a style of information systems architecture that enables the creation of applications that are built by combining loosely coupled and interoperable services (Tsai, Chen, & Fan, 2006). These services inter-operate based on a formal definition (or contract, e.g., WSDL) that is independent of the underlying platform and programming language. The interface definition hides the implementation of the language-specific service. SOA based systems can therefore be independent of development technologies and platforms (such as Java, .NET etc). In addition, applications running on either platform can consume services running on the other as Web services (Utschig, Rodriguez, & Buelow, 2006).

The Service Oriented Model is a complex architectural models and it revolves around a few key ideas (Spewak & Hill, 1992). A service is realized by one agent and used by a different agent. Services are mediated by means of the messages exchanged between requester agents and provider agents.

Another important aspect of services is their relationship to the real world: services are mostly deployed to offer functionality in the real world. We model this by elaborating on the concept of a service's owner—which, whether it is a person or an organization, has a real world responsibility for the service. In addition, Service Oriented Model makes use of meta-data, which is a key property of Service Oriented Architectures. This meta-data is used to document many aspects of services: from the details of the interface and transport binding to the semantics of the service and what policy restrictions there may be on the service.

SOA is a mechanism for defining business services and operating models and therefore providing a structure for information technology to deliver the actual business requirements and adapt in a similar way to the business needs (Utschig, Rodriguez, & Buelow, 2006). The purpose of using SOA as a business mapping tool is to ensure that the services created properly represent the business view and are not just what technologists think the business services should be. At the heart of SOA planning is the process of defining architectures for the use of information in support of the business, and the plan for implementing those architectures (Spewak & Hill, 1992).

SOA has also faced some criticisms including concern that Web Services standards and products are still evolving (e.g., transaction, security), and SOA can thus introduce risk unless properly managed and estimated with additional budget and contingency for additional Proof of Concept work (Utschig, Rodriguez, & Buelow, 2006). Web services are designed to support interoperable machine-to-machine interaction over a network. Web Service has an interface which can be processed by a machine in a specific format (specifically WSDL). Other systems interact with the Web service in a manner prescribed by its description using SOAP messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web-related standards

ARCHITECTURE

The Los Angeles County Operational Data Store (LAC-ODS) architecture is based upon recommendations of the Public Health Information Network (PHIN) initiative by Centers for Disease Control (CDC).

In order to maximize the use of pre-existing applications and bring together interoperability the architecture was developed for utilization of the common services while promoting interfaces to new systems with minimal impact on existing systems (LACDHS, 2003).

The architectural design was developed from the premise that change within the county environment is a fact of life and process. The County of Los Angeles continues to provide comprehensive public health programs and services over 17 million people. Therefore the approach to the architecture was to lay out a foundation which could handle change management while stressing on functions and interoperability. The need to reuse preexisting software applications was a major need since this would enable a more effective support of this initiative with greater inputs from the users. The technology capabilities were targeted to be implemented phase wise in order to have optimal feedback and collaborative construction of the architecture (Schuschel & Weske, 2004).

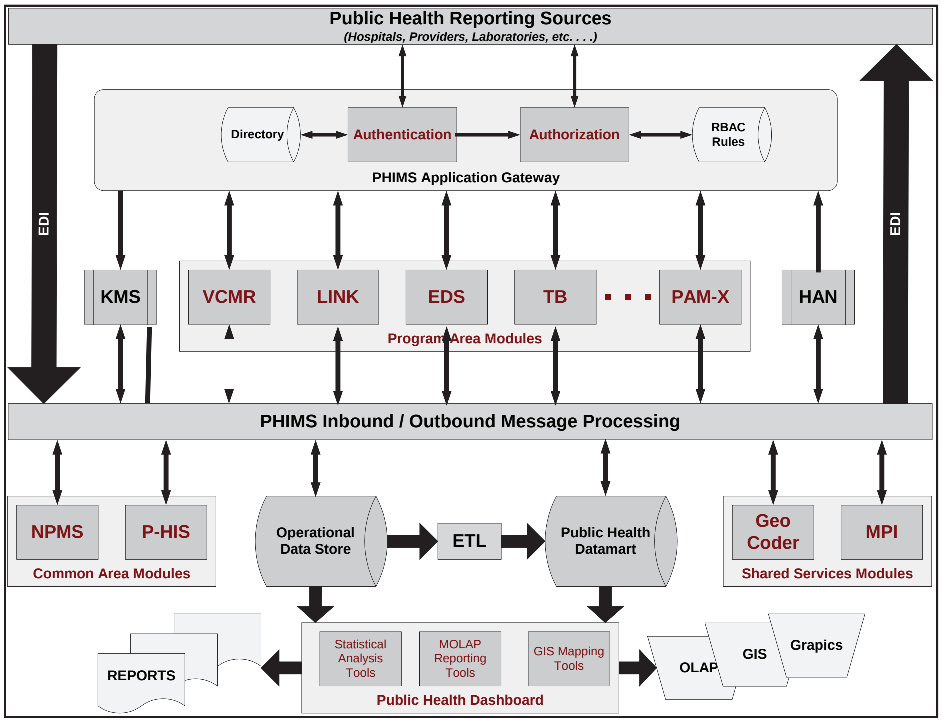
The primary focus of the architecture also kept in mind that the architecture should focus on the services rather than the application. This was done keeping in mind that there was care to adopt a broader scope from building from an existing infrastructure so that an open and flexible architecture on which applications and services can be run. This also eliminated the dependencies on databases supported by vendors. The flexibility of the architecture was ensured by adoption of standard technologies for all layers of the architecture and the messaging itself. There was care taken to design a modularized system to ensure services continuity (LACDHS, 2005).

The decision was to have HL7 RIM V3 as the messaging format for this architecture (Mead, 2005; Regio & Greenfield, 2006). HL7 is a healthcare standard for messaging which is widely being adopted by healthcare and related service providers in their initiative for unified interoperable electronic health standards and implementation of nation wide Electronic Health records. HL7 V3 is currently implemented in UK, Netherlands, Canada, Mexico, Croatia and Germany. Within the US agencies such as Centers for Disease control, Food and Drug Administration and Veterans Administration have adopted the V3 for large scale integration. The messaging included a layer of abstraction in the form of an intermediate XML format. There was a multi version support of the application for database structure and application behaviors for easy change/upgrade and switching functionalities. We created a clearly defined mapping from the physical data model onto the logical data model and from the triggers used by the application to those used in the interfaces. The messaging hub has robust tracking and logging system, which logs everything (including wrappers, etc.) that which is actually send/received is crucial. The design also addressed issues related to failure of real-time interaction and how the message would be handled for receipt and delivery of message. There was also consideration of data synchronization for queries, queues, and so forth (Klein, 2005).

The scope and function of each of the architectural components is as follows (see Figure 2):

1. **Authentication and Authorization**. This portion of the architecture provides technologies and processes used to authenticate users and authorize their access to system applications and resources. Authentication uses a LDAP implementation to authorize personnel and resources to access application system. The directory includes identification, demographic, and authentication credentials for each registered system user. Upon entry to any of the public health applications the user is prompted to provide their identification data and authorization credentials. Upon verification of the authentication credentials for the identified user access is granted to applications for which the authenticated user is authorized.
2. **Health Alert Network (HAN)**. This portion of the architecture provides technologies and processes used to communicate notifications, warnings, and alerts. The alerts may be initiated by an authorized user or by detection of triggering events within application systems. A personnel directory is used to maintain the identification, demographics, and communication addresses for potential alert recipients. A customizable set of rules is used to determine the appropriate communication mode to use for a given type of alert (telephone, email, pager, or fax) and to determine the escalation required for communications that are not acknowledged by the recipient within a predetermined threshold of time.
3. **PHIMS Inbound Message Processing**. This portion of the architecture provides the technologies and processes necessary to receive data electronically from external partners (e.g., Hospitals, Laboratories, and Physician Offices) and internal application systems. The inbound message processing component is configured to accept data in a variety of standard and proprietary formats. Supported standard formats include HL7 v2,

Figure 2. This figure details the existing functional program area modules shown as examples. The inbound and outbound messaging components relationship with the functional program area modules is clearly depicted and shows their role as the communication bus for all data interchange.



v3 and CDA, X12, and NCPDP. Proprietary formats include any form of ASCII file including XML documents, delimited and non-delimited flat files, and encapsulated files such as PDFs and images. Data are extracted from inbound transactions and placed into a staging area for use in importing into the Operational Data Store (ODS) and creation of outbound messages.

4. **PHIMS Outbound Message Processing**. This portion of the architecture provides the technologies and processes necessary to send data electronically to external partners (e.g., Hospitals, Laboratories, and Physician Offices) and other application systems. The outbound message processing component is configured to create outbound transactions in a variety of standard and proprietary formats depending upon the requirements of the receiving application. Supported standard formats include HL7 v2, v3, and CDA, X12, and NCPDP. Proprietary formats include any form of ASCII data file. Data for outbound transactions are retrieved from a staging area populated from the operational data store (ODS) or directly from functional area modules. Outbound message processing transforms the data in the staging areas into the appropriate transaction format for the recipient.

5. **Knowledge Management System (KMS)**. This component of the architecture provides the technologies and processes necessary to maintain coded terminologies and lexicons used in inbound and outbound messages, the ODS, and functional area modules. The KMS component is configured to import clinical terminologies such as SNOMED,

LOINC, and CPT as well as proprietary coding systems. Linkages are maintained between coded terminologies that facilitate the translation of codes from one terminology to another as well as provide a knowledgebase for use in rules processing, inference logic, and workflow management.

6. **Public Health Datamart**. This component of the architecture provides the technologies and processes necessary to extract, transform, and load data from the ODS and inbound message staging area into a star-schema based data structure used for analytical reporting. The data mart is a collection of fact tables and conforming dimension tables designed to provide an integrated multidimensional view of public health data. Common dimensions include factors such as time, location, demographics, and organization. Facts include items such cases, admissions, and observations.
7. **Business Intelligence Environment (Public Health Dashboard).** This component of the architecture provides the technologies and processes necessary to provide access to data in the datamart for use in analysis and visualization. The Business Intelligence Environment included multidimensional analysis tools, statistical tools, and geo-mapping tools to provide a comprehensive view of public health data. The Business Intelligence Environment includes a public health dashboard application that provides a set of measures declared by county management to be of greatest interest, quick links to information of interest, and Web services that expose information from other PAMs and external serices. Users of the dashboard application can drill down into the numbers which influence the measures from there they can slice, dice, and pivot the data as needed. The mapping software enables the data to be plotted on maps that can also be drilled and customized to fulfill a particular analytical need.
8. **Operational Data Store.** This component of the architecture provides the technologies and processes necessary to integrate data from inbound processing and functional area modules into a single data store. The design of the ODS is based upon the Health Level Seven (HL7) Reference Information Model (RIM). The data store is highly abstracted allowing it to collect any data that can be mapped to the entity, role, participation, act paradigm of the HL7 RIM. Data to be imported into the ODS are first placed into a staging area and are then transformed and imported into the ODS. An ODS API is under construction. The API will accept any HL7 v3 styled transactions and generate RIM objects. The RIM objects are then stored in the ODS using ORM software (i.e., Hibernate). The API will simplify the importing of data into the ODS. Inbound messages of any type can be transformed into HL7 v3 styled transactions and automatically mapped for import into the ODS.
9. **Functional Area Modules (PAMS, CAMS and Shared Services).** This component of the architecture is the most essential. The functional area modules include the applications, common routines, and services used to address tactical public health information processing requirements. Applications such as communicable disease reporting, nursing practice management, and laboratory information management make up the functional area modules as well as services such as geo-coding, identity managements, and record locator.

IMPLEMENTATION

An Enterprise Service Bus (ESB) for used for our deployment of SOA at the County of Los

Angeles. The deployment of the SOA required the conversion of existing systems into services. The task involved in achieving this SOA is ongoing for each system and a common set of components (HAN, CAMS, and Shared Services, etc.) are providing additional functionality (such as security and auditing). The bus is a collection of servers and components that will provide functions and set of tools to assist in the conversion of existing systems to functional services.

The SOA implemented in County of Los Angeles used the following major steps and technologies: Existing systems were used core services and business services were derived from these core services. In turn, process modeling was derived from the business services. Microsoft Technologies such as Microsoft.NET and Web Services, Microsoft BizTalk Server and Web Services Enhancements for .NET were utilized for this implementation.

Microsoft BizTalk was used for PHIMS inbound and outbound message processing and transformation. Microsoft SQL Server is used as the relational database management system for the integrated data repository designed after the Public Health Logical Data Model (PHLDM) and the Health Level Seven Reference Information Model (HL7 RIM). COGNOS and ESRI solutions are used to provide multi-dimensional online analytical processing (MOLAP) in a business intelligence environment that supports GIS, statistical analysis, and reporting. Information standards used within the architecture include HL7 Version 2 ADT and ORU messages; HL7 Version 3 RIM, Vocabulary, and data type specifications; and the Common Alerting Protocol (CAP) data interchange format.

We have used the following technology stack in Web services:

- WSDL provides description of service behavior and the interface contract.
- The HTTP protocol for network service. SOAP provides a platform independent protocol for sending messages across the wire.
- UDDI provides the means to discover the service both at design time and at run time.
- XML is the technology that underpins it all. Since XML is text based, it can be read by virtually any machine on any platform giving a high degree of interoperability. WSDL, UDDI, and SOAP are all XML based.

The implementation of this architecture consisted of specific focus on the following functional areas based for creation of a service centric approach.

- **Authentication and Authorization.** Security services provide a single point of entry for user authentication and system access authorization. User authentication service used multifactor authentication methods including passwords, RDS Secure ID tokens and X509 certificates which work with Microsoft single sign on architecture implemented with the help of Microsoft SharePoint Portal Server.

An LDAP based directory and RBAC defined system access privileges and rules control which public health application authenticated user are allowed to access. This is implemented using Microsoft Windows 2003 Active Directory.

Alerting services provided the necessary technical infrastructure to facilitate the distribution of notifications, warnings, reminders, and alerts to public health personnel and partner organizations. This is implemented using combination of MIR3 Enterprise server and custom developed components in Microsoft.NET technologies.

- **PHIMS Inbound/Outbound Message Processing.** The PHIMS Inbound/Outbound Message Processing area of the architecture is a collection of technologies and processes that enable the electronic exchange of information between the county public health application systems and external information systems. PHIMS Inbound/Outbound Message Processing implements standard and proprietary message formats for data exchange. Electronic trading partners include entities external to County of Los Angeles such as healthcare providers, state health department, laboratories, and public health partners. PHIMS

Inbound/Outbound Message Processing can receive and send messages in various standards and legacy formats like HL7 v2.x, HL7 v3, HL7 CDA, CSV, ASCII Files etc. This is implemented using combination of custom developed Microsoft. NET applications and Microsoft BizTalk server.

- **Functional Area (PAMS, CAMS and Shared Services).** The Functional area of the architecture is a collection of program area modules, common area modules, and information routing technologies and processes used to manage conditions and events of public health interest. PAMs support the information and processing needs of a particular public health program such as communicable disease surveillance, environmental health, and bio-terrorism response management. CAM supports information and processing needs that transcend program areas such as laboratory information management, event detection, and controlled medical terminology services. PHIMS information routing technologies and processes include message parsing, transformation, and routing technologies and an integrated operational data store. The centralized operational data sore is developed based on HL7 V3 Reference Information Model (RIM) and implemented in Microsoft SQL Server 2000.
- **Public Health Datamart and Public Health Dashboard.** This is a collection of technologies and processes that facilitate analysis, visualization, and reporting of public health information. The data mart integrates data from multiple county public health, clinical, and administrative systems. The data are organized into facts and dimensions that allow multidimensional online analytical processing. The Public Health Datamart functions as a single source of truth for detail and aggregate public health data for use by a variety of business intelligence tools and applications. Business Intelligence services include both data analysis components and GIS components.

Data integration in healthcare has become a significant factor for any organization. In particular, County of Los Angeles goals are to have public health data be utilized for monitoring its population health dynamics and better plan for large scale health crises.

The decision to have data integration synchronized by use of HL7 healthcare standard messaging was in part to take into account the disparate applications and vendors already in existence in county.

The implementation of this integrated standards-based system in LAC provides a synchronized view of public health data and resources across all *program areas* (see Figure 2). High level implementation and testing strategies were put into place prior the development process. All processes were documented in depth for ongoing integration of PAMS and for future planning.

EVALUATION

The key components of the architecture include (see Figure 2) (1) Infrastructure & housekeeping services to provide authentication, directory and security services for the system and enable single sign-on. (2) A knowledge management module that manages translation and mapping of coded data from other systems to the Operational Data Store (ODS) in a standardized nomenclature. (3) Provision for single sign-on portal for authentication & entry into the system. It also incorporates an event alerting mechanism. (4) An electronic interface to systems that need to send data and receive data from the ODS, program area modules, and external data sources. (5) Analysis and Visualization services and data mart provide data analysis, reporting and GIS services through a dashboard. (6) Operational data store to save integrated data from functional area modules and external data sources.

The architecture provides the necessary infrastructure for a service oriented approach to development of functional area modules. Functional area modules support the business functions of public health. Some functional area modules are program specific (PAM), some are common to multiple program areas (CAM), and others are services invoked by PAMs and CAM. The services included in the architecture not only implement business functions but are essential infrastructure functions; security, alerting, messaging, terminology, and business intelligence. The development of functional area modules is streamlined due to the existence of these foundational features. The inbound and outbound messaging features enable functional modules integration and interoperability.

Web-enabled data management, tracking, decision support, and business intelligence systems, (including NPMS) have been developed for diverse agencies and departments within the Public Health system of Los Angeles County. All iterations from the design and planning phases to development, implementation and expansion were carefully evaluated with program users, administrative staff, and technology experts. The involvement of the user departments at all stages helped to develop systems that were easy and convenient to use.

The use of a model driven SOA has allowed the County of Los Angeles to take advantage of state-of-the-art information technologies while at the same time leveraging the investment made in its legacy application system. The use of industry standards positions the system to be interoperable with similar efforts conducted in other jurisdictions such as in neighboring Counties, the state of California, and the CDC.

It is proving to be a useful tool in assisting the County in coordinating the work of its multiple vendors, ensuring the coherent application of technologies to daily operations and preparedness activities, as well as, fostering collaboration between Los Angeles County Public Health and its partners.

In addition to design and development of a SOA architecture several other issues were brought up which helped in the implementation and planning of this architecture for public health in Los Angeles County. The vision by key county personnel expressing the necessity of data sharing and leveraging data as a department wide asset was clearly communicated. The CDC defined standards for use in enabling a public health information network has reached a level of maturity that could be utilized for the design elements of this architecture. The national recognition of the vulnerability of our public health system has motivated the United States Congress to allocate the funds necessary to underwrite the cost of development. Issues that continue to be part of the long term planning of this implementation include the inertia and tradition of autonomy enjoyed by program area leaders. The existence of silos in the program areas continue to yield challenges in navigation through the silo boundaries. The large scope of the implementation of this system and the presence of multiple vendors brings up coordination issues for business and technical needs and understandings for common grounds. The process of funding and resource limitations also becomes significant factors in the development process periodically.

CONCLUSION

Towards the end of 2003, a project team was assembled to study the requirements of Public Health Information Management System (PHIMS) at Los Angeles County. The objective was to design a system architecture that met CDC's Public Health Information Network (PHIN) standards. The design of the PHIMS architecture was completed in March 2004. The remainder of 2004 was spent on evaluations and analysis of the design with internal and external stakeholders. In addition the specifications were subjected to further refinement in conjunction with vendors. The vendors and County personnel were committed to assume accountability for implementing portions of the design, and developing plans for phased implementation of the architecture. By August 2005, core features of the architecture were implemented in a lab environment including prototype implementations of key functional components of the architecture. The entire countywide architecture is in the process of transitioning into production. There is ongoing emphasis in Los Angeles County Public Health to identify and prioritize PAMs which should be integrated into this architecture. The functional requirements of program area modules and the definition of an action plan for incorporating well-architected functional solutions into the technical infrastructure is ongoing.

It is expected that the Public Health Information Network System effort will expedite the consolidation of critical clinical and public health data across diverse individual IT systems (Verma et al., 2005). It will leverage existing investments in legacy systems and merge them with standards-based Web-enabled systems to provide a synchronized view of public health data and resources across all program areas. The LAC-ODS can be queried for a person-centric view of health-related data across all program areas. The visualization services provide a dashboard view and drill-down report capabilities for decision support and alignment of critical public health resources.

The PHIN compliant systems will specifically address the following (Health Alert Network, 2002; Los Angeles County Department of Health Services, 2004): (a) The development of a Web-based system architecture for Public Health programs and health districts that is capable of supporting electronic data exchange from public health partners using a HL7-based integration hub, (b) the development of management tools and applications to assist public health response, and (c) recovery activities while providing resources to support departmental integration. The PHIN/NEDSS systems will leverage individual system components for the overall improvement of public health information technology infrastructure while contributing to the development of a common enterprise data warehouse that will unify public health and clinical data under a unique person identifier.

There are multiple systems in place that support communications for public health labs, the clinical community, and state, and local health departments. However, most of these systems operate in isolation. Numerous benefits will start accruing as parts of the system are built and integrated into the business processes of the local health services. The implementation of a unifying system will further improve access to laboratory data and response protocols, advanced capabilities for rapid notification of public health partners, response agencies, the media, and the general public. There will be an enhanced capability to train public health staff and a uniform data exchange standard for exchanging data between the public health partners. Real-time collection of data from heterogeneous healthcare systems, program area modules, consolidation, and cross-indexing of data, integrated directory infrastructure for public health personnel and identity management, integration of related healthcare, and patient data from heterogeneous systems into a common interface, provide access through a ubiquitous Web-based portal that will obviate the necessity of client-side implementations of application systems, provide a mechanism to disseminate critical and public-interest information to the community in general are additional benefits.

ACKNOWLEDGMENT

This work was performed by CAL2CAL Corporation at the LAC-Department of Public Health. The ODS and ODS related applications are developed under a grant from the CDC to state and local jurisdictions to support public health preparedness and response to bioterrorism (Los Angeles County Department of Health Services, 2003). This work was done in collaboration with Public Health Information Systems—Los Angeles County Department of Public Health.

REFERENCES

Arsanjani, A. (2004). Service-Oriented modeling and architecture: How to identify, specify, and realize services for your SOA, *IBM*.

Health Alert Network (2002). *Public health information technology, functions and specifications (for emergency preparedness and bioterrorism)* (February 2002). Retrieved from http://www.cdc.gov/cic/functionsspecs/IT_Functions Specifications_final_21402.pdf

IBM (2005). Web sphere integration reference architecture: A service-based foundation for enterprise reference architecture.

Klein, J. (2005). Integrating (2005) electronic health records Using HL7 clinical document architecture. *Business Integration Journal.*

Los Angeles County Department of Health Services, *Public Health Information System*s (October, 2003). [a] SARS-EDS Software Requirement Specifications v1.0.2, [b]SARS-EDS Functional Document v1.0.2.

Los Angeles County Department of Health Services, Public Health Information Systems, (March, 2003). *PHIS Action Plan for Bio-terrorism Preparedness Initiative* – The application of NEDSS architecture to Public Health Information systems in Los Angeles County.

Los Angeles County Department of Health Services, Public Health, *Bio-terrorism Preparedness & Response Unit* (April, 2005). ODS and ODS-API design and architecture documents.

Los Angeles County Department of Health Services, Public Health, Bio-terrorism Preparedness & Response Unit (March 2004). *Bio-terrorism Preparedness and Response Initiative* – The application of PHIN Standards & NEDSS Architecture towards integrating Pubic Health Information Systems in Los Angeles County – Program Plan.

Mead, C., (2005) Data interchange standards in healthcare IT—Computable semantic interoperability: Now possible but still difficult, do we really need a better mousetrap? *Journal of Healthcare Information Management, 20*(1), 71-78.

National Electronic Disease Surveillance System (2004). *The surveillance and monitoring component of the public health information network*. Retrieved from http://www.cdc.gov/nedss.

Public Health Information Network (2004). *An overview*. Retrieved from http://www.cdc.gov/phin/index.html.

Regio, M., & Greenfield, J. (2006). Design and implement a software factory. *The Architecture Journal,* 6.

Schuschel, H.. & Weske, M. (2004). Infrastructure for Collaborative Enterprises. In *Proceedings of the 13th IEEE International Workshops on Enabling Technologies (WETICE'04)* (pp. 75-80).

Spewak, S., & Hill, S. (1992). *Enterprise architecture planning publishers*. John Wiley and Sons.

Sprott, D. (2004). The business case for service oriented architecture. *CBDI Journal*, *3-11.*

Tsai, W.T., Chen, Y., & Fan, C. (2006). PESOI: Process Embedded Service-Oriented Architecture. *Journal of Software, 17*(6), 1470-1484.

Utschig, C., Rodriguez, J., & Buelow, H. (2006). Practical interoperability approaches, WS- Security and WS-Addressing Explained. *SOA Webservices Journal.*

Verma, R., Shakir, A.M., Cardenas D., Mittra D., Marwaha A., & Datta G (2005). Integrated disease surveillance application in a standards (HL7 V3) based public health information network. *The Journal on Information Technology in Healthcare, 3*(6).

This work was previously published in the International Journal of Healthcare Information Systems and Informatics, Vol. 2, Issue 2, edited by J. Tan, pp. 53-66, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 2.13
The Development of a Health Data Quality Programme

Karolyn Kerr
Simpl, New Zealand

Tony Norris
Massey University, New Zealand

ABSTRACT

Data quality requirements are increasing as a wider range of data becomes available and the technology to mine data shows the value of data that is "fit for use." This chapter describes a data quality programme for the New Zealand Ministry of Health that first isolates the criteria that define "fitness" and then develops a framework as the basis of a health sector-wide data quality strategy that aligns with the sector's existing strategies and policies for the use of health information. The framework development builds on existing work by the Canadian Institute for Health Information, and takes into account current data quality literature and recognised total data quality management (TDQM) principles. Strategy development builds upon existing policy and strategy within the New Zealand health sector, a review of customer requirements, current sector maturity and adaptability, and current literature to provide a practical strategy that offers clear guidelines for action. The chapter ends with a summary of key issues that can be employed by health care organisations to develop their own successful data quality improvement programmes.

INTRODUCTION

The New Zealand health sector data quality improvement programme attempts to provide a structure for managing data to prevent data quality errors at the source of collection and to maintain the meaning of the data as they move throughout the health sector. This approach requires viewing data quality from a holistic perspective — going beyond a one-dimensional assessment of quality based only on accuracy — and assessing other dimensions (Ballou & Tayi, 1999) such as relevance, timeliness, comparability, usability, security, and privacy of data.

As data quality affects everyone in the health sector, the whole sector is responsible for maintaining and improving data quality. The role of the New Zealand Ministry of Health is one of leadership and support, whilst data collectors need to employ all possible processes to ensure only high quality data are collected, using agreed national and international standards, where available. Data quality needs to be the responsibility for high-level managers in an organisation to ensure the entire organisation makes the required changes for improvement. "All too often data quality is seen as something that is the responsibility of informatics staff alone and is often seen with disinterest by clinicians and managers, despite being so critical to the quality of the decisions they make" (Data Remember, UK National Health Service, 2001; UK Audit Commission, 2002).

This chapter describes the development of a data quality evaluation framework (DQEF) and underpinning strategy for the Ministry of Health in New Zealand and outlines the process to "institutionalise" total data quality management throughout the whole of the health sector.

THE IMPORTANCE AND ELEMENTS OF DATA QUALITY PROGRAMMES

Bill Gates (1999) states:

The most meaningful way to differentiate your company from your competition, the best way to put distance between you and the crowd, is to do an outstanding job with information. How you gather, manage and use information will determine whether you win or lose.

Organisations are becoming more and more dependent on information. Virtually everything the modern organisation does both creates and depends upon enormous quantities of data. A comprehensive data management programme is therefore essential to meet the needs of the organisation (Pautke & Redman, 2001). Many authors, for example, Levitin and Redman (1993), also draw attention to the importance of data quality in managing information as a resource of the organisation.

The first step in setting up a data quality (improvement) programme is therefore to decide the determinants that define quality. A framework is then required to apply these determinants and their associated metrics that can assess the level of data quality and establish processes such as collection, storage, access, and maintenance that lead to quality improvements where they are necessary. Finally, whilst a data quality framework models the data environment, it must be underpinned and supported by a strategy that is broader in scope. This strategy establishes the business purpose and context of data and aims to make the framework a routine tool and part of day-to-day operations.

These three elements, quality determinants, assessment framework, and implementation strategy are the core components of a data quality programme and we now look at these stages in more detail.

Data Quality Determinants

Strong, Lee, and Wang (1997) take a consumer (people or groups who have experience in using organisational data to make business decisions) focused view that quality data are "data that are fit for use," and this view is widely adopted in the literature (Wang, Strong, & Guarascio, 1996). Redman (2001) comes to the following definition based on Juran and Godfrey (1999):

Data are of high quality if they are fit for their intended uses in operations, decision-making, and planning. Data are fit for use if they are free of defects and possess desired features.

Clearly, however, fitness for purpose depends upon the purpose and so the set of data quality determinants will vary according to the application. In addition, modern views of data quality have a wider frame of reference and many more features than the simple choice of attributes such as accuracy and currency. There are therefore multiple approaches to designing and applying data quality systems as well as competing terminologies to trap the unwary (Canadian Institute for Health Information, 2005; Eppler & Wittig, 2000). The approach we have adopted here is based on a hierarchical system (see next section) developed by the Canadian Institute for Health Information (CIHI) (2003a, 2005). At the uppermost level are the familiar attributes such as accuracy, relevance, and so forth. These attributes are referred to in the scheme as "dimensions." Each dimension is defined in context by appropriate determinants known as "characteristics." For example, characteristics of the accuracy dimension might include the tolerated level of error and the population to which the data accuracy must apply. Characteristics require answers to "what is/are" questions. Underpinning these characteristics are "criteria" that define processes and metrics used to assess the presence of potential data quality issues. Thus, the level of error characteristic might be assessed by asking if the error falls into a predefined range and if the level of bias is significant. Criteria typically demand "yes" or "no" answers. In this chapter we are concerned mainly with data quality dimensions and criteria although we describe the hierarchical process for their selection.

Data Quality Evaluation Frameworks

At its most basic, a data quality evaluation framework (DQEF) is defined by Wang et al. (1996) as "a vehicle that an organisation can use to define a model of its data environment, identify relevant data quality attributes, analyse data quality attributes in their current or future context, and provide guidance for data quality improvement."

In our terminology, a DQEF enshrines and enacts the processes and metrics (criteria) that assess whether the level of a dimension (e.g., accuracy) is acceptable or not. However, Eppler and Wittig (2000) describe a data quality framework as:

A point in time assessment and measurement tool, integrated into organisational processes, providing a benchmark for the effectiveness of any future data quality improvement initiatives and a standardised template for information on data quality both for internal and external users.

The same authors also add that a framework should not only evaluate but also provide a scheme to analyse and solve data quality problems by proactive management.

Ideally, therefore, a DQEF goes beyond straightforward quality assessment and becomes an integral component of the processes that an organisation puts in place to deliver its business goals (Willshire & Meyen, 1997). The framework then uses data quality to target poor quality processes or inefficiencies that may reduce profitability or lead to poor service (Wang et al., 1996). The implication of this extended framework is that the organisation will need to engage in a certain amount of business process reengineering. That is why the framework needs to be part of an organisational improvement programme and strongly tied to a data quality strategy.

Data Quality Strategies

Thus far, little has been published on what constitutes a data quality strategy, let alone an evaluation of a structured and tested improvement programme. Recently, however, this type of strategy has become increasingly important as a core requirement for many businesses. It is likely that some large organisations do have such strategies, improvement programmes, or components of them, but these are not currently documented and available in the literature. Davis (2003), publishing on the FirstLogic Web site, wrote several articles on his vision of a data quality strategy. According to Davis, a data quality strategy should include:

- A statement of the goals (what is driving the project)
- A description of the primary organisational processes impacted by the goals
- A high-level list of the major data groups and types that support the operations
- A description of the data systems where the data groups are stored
- A statement of the type of data and how they are used
- Discussion of cleansing solutions matching them to the types of data
- Inventory of the existing data touch points
- A plan for how, where, and when the data can be accessed for cleansing
- A plan for how often the cleansing activity will occur and on what systems
- A detailed list of the individual data elements

Whilst Davis' list is a useful starting point, it is based on an information providers' perspective. Other components should be added to incorporate the needs of consumers and to define and document these. This is still not easy to do given that customers often do not know what their needs are (Redman, 2001). A first step would be the identification of the organisation's customers, or important customer groups, where there are too many individual customers for initial improvement programmes.

With this survey of the elements of data quality improvement programmes we can now look briefly at international attempts to apply the principles to health care.

INTERNATIONAL HEALTH DATA QUALITY PROGRAMMES

Health care delivery and planning rely heavily on data and information from clinical, administrative, and management sources, and quality data lead to quality and cost-effective care, improving patient outcomes and customer satisfaction. Data for health care delivery range from the clinical records of individual patients, detailing their interactions with medical services, to the administrative data required to manage the complex business of health care. When abstracted and aggregated in warehouses, and informed by other management and policy information, these "unit level data" produce knowledge bases for heath care planning and decision support. The totality of data from these various sources can then be used for further policy development (Al-Shorbaji, 2001). A prime example is the planning at government level to provide services that address the prevalence and distribution of diseases such as diabetes in a population. Clearly, any programme that improves the quality of data at all levels will improve the quality and cost-effectiveness of care and patient outcomes.

A review of international data quality improvement programmes in health care, including the National Health Service (NHS) (Department of Health, 2004) in the United Kingdom, the Canadian Institute for Health Information (CIHI) (2003b), HealthConnect Australia (Department of Health and Aging, 2003), and the United States Department of Health and Human Services (2002) identified similarities between the various programmes. All of the reviewed programmes note the multilevel, multidimensional complexity of data quality improvement initiatives. They seek to manage data proactively and ensure integrity by preventing data quality problems using a systematic total data quality management (TDQM) approach (Wang, 1998). There is also commonality of role expectations — the data suppliers are responsible for the quality of the data they provide to the central government, whilst central government is required to provide leadership and assistance to data suppliers by developing sectorwide standards and best practice guidelines.

The NHS in particular outlines clearly, and in detail, the substantial work that is required by any health care provider to ensure good data quality. The NHS developed an accreditation scheme that was initially thought to be all that would be needed to ensure the supply of good quality data. The scheme is extensive and was found to be very successful but did not sufficiently identify the responsibilities of the data supplier; central government was still monitoring more than it was leading. This discovery led to the more extensive guidelines developed around principles of data quality supported within the NHS. Several NHS Trusts have published data quality strategies on their Web sites that align to the NHS core strategy requirements. These efforts have been recognised in a recent review of the programme by the UK Audit Commission (2004), which found significant improvements to levels of data quality. However, similar issues to those initially found are still apparent after five years of targeted improvements and the report recommended:

- The development of a more coordinated and strategic approach to data quality
- Development of an NHS-wide strategy for specifying, obtaining, and using both national and local information
- Making more and better use of patient-based information
- Involving Trust board members
- Training and developing staff
- Keeping systems up to date

In Canada, the CIHI also briefly discusses accreditation for enhancing collaboration with data suppliers. They have undertaken extensive work on data quality through collaborative work with experienced statisticians from Statistics Canada and base their theories on research by the Massachusetts Institute of Technology in the U.S.

We now look at the relevance of this international experience in health data quality programmes to New Zealand health.

TOWARDS A HEALTH DATA QUALITY PROGRAMME IN NEW ZEALAND

The New Zealand Health Information Service (NZHIS) is a specialised group within the Ministry of Health responsible for the collection and dissemination of health-related data. NZHIS has as its foundation the goal of making "fit-for-purpose" information readily available and accessible in a timely manner throughout the health sector to support the sector's ongoing effort to improve the health status of New Zealanders. The vision of NZHIS is to be a leader in the provision of health information services in New Zealand, and to be recognised and respected as a leading organisation internationally. Effective and timely use of information is crucial to achieving this vision. NZHIS has responsibility for:

- The overall collection, processing, maintenance, and dissemination of health data, health statistics and health information
- The ongoing quality improvement of data entering the national data collections1
- The continuing maintenance and development of the national health and disability information systems
- The provision of appropriate databases, systems, and information products
- The development and provision of health and disability information standards and quality-audit programmes for data
- Coordination of ongoing national health and disability information collections and proposals for their development
- Analysis of health information, performance monitoring, benchmarking, and advice on the use of information obtained from NZHIS

Thus, while the NZHIS is responsible for the lead on data quality, this does not mean that it is solely accountable for solving data quality problems. The role of the NZHIS is rather to define data quality criteria and establish a framework in which health care organisations can assess the quality of their own data. This framework must also ensure that data quality does not degrade when data are moved between organisations. These data move with the patients they refer to creating reciprocal dependence between the organisations so that poor data management in one organisation can adversely and incrementally affect other organisations and the care a patient receives. A national "systems" framework is therefore needed to certify that data used for decision making meet the same quality criteria and standards both within and between organisations.

Assessment of the national health collections at NZHIS (2003) showed that the required framework was unlikely to lead to sustainable improvements unless it was placed in the context of a national strategy that would reengineer data processes and embed the quality-centric revisions in normal, everyday practice. The resources to improve quality were already available, but they were being applied to the wrong processes and without consistency or direction. In addition, short-term priorities needed to focus on areas where benefits could be realised easily with long-term projects concentrating on implementing change.

Thus, as suggested by the previous discussion, the approach to the New Zealand Health Data Quality Improvement Programme is seen to consist of three stages.

1. Determination of the criteria needed to judge health data quality in a New Zealand context
2. Development of a practical framework to apply these criteria to existing and new database collection
3. Development of a strategy to implement the framework and embed it in normal practice

The next sections describe the research methodology and the design and application of these three stages.

RESEARCH METHODOLOGY FOR THE PROGRAMME DEVELOPMENT

The research utilised several qualitative methodologies — action research, semi-structured interviews, focus groups, and a questionnaire to develop and formally assess a health data quality framework that could be tied to an implementation strategy and promulgated as a data quality programme. The two focus groups were derived from a Ministry Data Quality Team (MDQT) formed specifically to look at ways of improving quality in a consistent way across the organisation. Membership of the MDQT was selected from across the ministry to bring together business units that appeared to have similar issues with data quality but at that time had no formal infrastructure to coordinate quality initiatives. Members were mostly "information users" such as information analysts and business intelligence staff, but some were also members of the already existing operational Clinical Analysis Team. All regularly used data for a wide range of different purposes.

A grounded theory (Strauss & Corbin, 1998) approach was used to analyse the data for content themes that could reveal new concepts. The research concentrated on eliciting the opinions of the participants on areas such as:

- The applicability of the criteria, characteristics, and dimensions for the assessed collection
- Proposal of other dimensions that may be applicable
- The language used in the framework
- The language and examples provided in the user manual
- The length of time required to complete the assessment using the framework

- The value to users of the information provided from using the framework
- The table of contents for the data quality documentation folder

The DQEF then went through a pilot evaluation process using two health data collections. Initial assessment was made on the mortality data collection which is a national health collection considered to have good data quality in relation to other collections. The mortality collection has been established to provide data for public health research, policy formulation, development and monitoring, and cancer survival studies. A complete dataset of each year's mortality data is sent to the World Health Organization to be used in international comparisons of mortality statistics.

The second data collection consisted of clinical data held in a local hospital setting. These data are used to determine best health outcomes for clinical care pathways and they are consequently stored at a more granular level than the national health data.

Further details of methodology are given with the discussion of the research findings.

RESEARCH FINDINGS

Selection of Health Data Quality Dimensions

The development of suitable dimensions for assessing health data quality in a New Zealand context was based as indicated on the Canadian Institute of Health Information (CIHI) data quality framework (Long, Richards, & Seko, 2002). This framework is itself based on Statistics Canada guidelines and methods, information quality literature, and the principle of continuous quality improvement (Deming, 1982). The CIHI is comparable in function to NZHIS and the health care systems of the two countries are also similar in many respects.

The CIHI data quality framework operationalises data quality as a four-level conceptual model (Long & Seko, 2002). At the foundation of the model are 86 criteria and these are aggregated using the framework algorithm into the second level of 24 characteristics that in turn define 5 dimensions of data quality: accuracy, timeliness, comparability, usability, and relevance. Finally, the five dimensions can be reduced using the algorithm into one overall database evaluation. Figure 1 provides a summary of the four-level conceptual CIHI model.[2]

The CIHI framework was first assessed for completeness, applicability, and ease of adaptation in New Zealand against current ministry information strategy documents. These include regional health care providers strategic plans and the WAVE report (working to add value through e-information) (WAVE Advisory Board, 2001), which is New Zealand's national information management strategy for health. Compliance with New Zealand legislation was also considered.

The MDQT focus groups made minimal changes to the basic content of the CIHI framework retaining the hierarchical approach but removing some criteria that were thought inappropriate and adding others that were considered important in the New Zealand context, yielding a total of 69 from the original 86. The most significant change was the inclusion of an additional dimension, privacy and security, to satisfy New Zealand concerns. The CIHI states that privacy and security are implicit requirements that are embedded in all their data management processes. Whilst the same could also be said of the Ministry of Health, the pervading culture in New Zealand requires that privacy and security of information, in particular health information, are paramount. Therefore, the MDQT felt there was a requirement for explicit and transparent consideration of these quality dimensions. The underpinning characteristics for these new dimensions were developed by the senior advisors on health sector privacy and security to ensure alignment with the ongoing development of new privacy and security policies.

The six data quality dimensions chosen for the DQEF by analysis of the feedback from the focus groups are as follows with accuracy as the most important:

1. Accuracy is defined within the framework as how well data reflect the reality they are supposed to represent.
2. Relevancy reflects the degree to which a database meets the current and potential future needs of users.
3. Timeliness refers primarily to how current or up-to-date the data are at the time of use.
4. Comparability is defined as the extent to which databases are consistent over time and use standard conventions (such as data elements or reporting periods), making them similar to other databases.
5. Usability reflects the ease with which data may be understood and accessed. If data are difficult to use, they can be rendered worthless no matter how accurate, timely, comparable, or relevant they may be.
6. Security and privacy reflect the degree to which a database meets the current legislation, standards, policies, and processes.

Within the research participant group the understanding of the meaning of these dimensions varied. No one definition could be found for even the most commonly used data quality dimensions (Wand & Wang, 1996). For this reason it is important for the development of the DQEF to explicitly define each dimension. It was decided to utilise the definitions provided by the CIHI Framework where possible, as these aligned with the characteristics and criteria adopted throughout the framework. Importantly, the dimensions were found to be mutually exclusive and collectively exhaustive.

Definition of the Health Data Quality Evaluation Framework

The development of the health DQEF began with a current state analysis (New Zealand Health Information Service, 2003) through a preliminary survey of managers and users from across the ministry. This survey consisted of open questions requiring free-text answers to elicit information on a set of factors including historical and contextual information about the collection, the data collection processes, any changes made to data from within the ministry, what the data are used for, where they reside, and the nature and perceived effectiveness of existing data quality initiatives.

The gathering of this information proved difficult. The survey results showed there were currently no compiled and complete records of data quality for any of the national data collections administered or managed by the Ministry of Health and neither was there clear accountability for data quality in the health sector. The information is spread over a range of business units, people, and documents so that the Ministry cannot easily assess the scope or effectiveness of its data quality measures. Furthermore, the varying requirements for quality between centrally held and managed collections and those "at the coalface" led to considerable uncertainty as to what "quality" entailed. This situation and extensive discussions with data users involved in the development of the Ministry of Health Information Systems Strategic Plan served only to confirm the need for the DQEF and the availability of assessment tools that could provide information on the nature and levels of data quality and identify the sources of problems so as to allocate responsibilities.

For the New Zealand Ministry of Health, therefore, the DQEF takes the form of a tool that allows a consistent and accurate assessment of data quality in all national health data collections, enabling improved decision making and policy development in the health sector through better information. The DQEF standardises information on data quality for users, provides a common objective approach to assessing the data quality of all health information databases and registries, and enables the identification and measurement of major data quality issues.

The draft DQEF, consisting of the aligned set of quality criteria, characteristics, and dimensions, was sent to all group participants. A presentation to the MDQT was made prior to the focus groups to ensure all participants had a common understanding of the purpose of the DQEF and the desired outcome goals. The group participated in two focus groups of two hours each. A member of the Ministry's Health Information Strategy and Policy Group (the researcher) led the focus groups and an administrator was present to make audio recordings and to later transcribe the recordings, noting also the interaction between group members on discussion points.

A template was developed to assist data managers and users to assess and document the effectiveness of the DQEF, its user manual, and the proposed data quality documentation folder for each data collection. The documentation folder houses all information pertaining to the quality of each data collection and makes it available in both paper and online format for access by all staff at the Ministry of Health. Following the focus group sessions, a second review of the DQEF, assessed using criteria developed by Eppler and Wittig (2000) ensured that the DQEF remained robust after localised changes.

The information provided by the DQEF evaluation highlighted data quality problems that were already known to the data custodians who had already initiated work to make improvements. However, several other deficiencies were highlighted, indicating where improvements in the current data quality work could be made, or in the case of timeliness, highlighting known issues that perhaps should be made a priority for improvement.

The issue of timeliness of the suicide in the mortality data collection data highlights a trade-off debate between dimensions. For example, data users are aware that some datasets are likely to be incomplete; then the data can be released without delay making them become usable at an earlier date. Eppler and Wittig (2000) note that trade-offs are not commonly addressed in frameworks. Research participants did not raise this issue and the DQEF does not explicitly address trade-offs between dimensions. However, an analysis of possible trade-offs and their implications could be included in the user manual. Different trade-offs would apply to different collections, and again may be different for specific uses of the data (Ballou & Pazer, 2003).

An issues register was kept to supply feedback to the researcher on the usability of the DQEF and user manual. Overall, the DQEF was found to provide useful data quality information by collection users and custodians and to provide sufficient information to make at least preliminary prioritised lists of essential data quality improvement projects. A detailed analysis of the feedback is provided.

Training

Further training was required to ensure that evaluators use the DQEF consistently and that it is a practical and easy tool to use. During training sessions it was found that many users of the DQEF initially made assumptions about the meaning of criteria and needed to refer to the metrics and definitions contained in the user manual. Consistency of meaning of criteria is important since many different users complete the evaluation both within an evaluation cycle and also from one evaluation to the next. For this reason a section was included under each criterion that asked the evaluator to describe in free text how they came to their decision, including references to any data quality information used to make their decision. This requirement also aids any subsequent evaluation of a collection.

Time Taken to Complete Evaluations

As noted in the findings of the Current State Analysis of Data Quality in the Ministry of Health, data quality information is held in many different locations and finding this information takes considerable time. The time taken to complete an evaluation of a data collection by already busy staff was estimated to be a minimum of four hours when all relevant documentation was available. In practice, the information was held in disparate locations by different staff and the initial evaluations took far longer. Repeated evaluations of the same collections would be completed much more efficiently, as much of the information could remain the same or merely need updating.

Intended Unit of Analysis

The researcher and participants also discussed the granularity of evaluation or intended unit of analysis. Participants asked, "was the DQEF able to evaluate a collection as a whole and also a column within a collection?" Price and Shanks (2005) found similar issues when implementing a data quality framework and noted the limitations of a framework on data that are not uniquely identifiable, such as those found in non-key columns. It was therefore decided to measure each collection as a whole. However, there are specific data elements that can be assessed individually and are extensively used for analysis and decision making. An example would be a registry, such as the National Health Index (NHI, a unique patient identifier), where each field of demographic information is the reference for other collections and provides information to prevent duplicate allocation. In this case, the DQEF can be used as a two-dimensional tool, assessing the registry as a whole and then each element.

User Manual

Extensive changes to the CIHI user manual were required to make it useful to the New Zealand health care environment. The findings show that the target audience for the accompanying DQEF user manual needed careful definition to ensure that the language was appropriate and the documents easy to use. In particular, the language used in the original CIHI version was found to be too simplistic for the intended audience. DQEF users are likely to be systems administrators, data quality advisors, and members of the Business Intelligence Team, but the language assumed little underlying knowledge of data and systems. The user manual could also be shortened by moving some of the underlying theory used to develop the Data Quality Improvement Programme to other documents intended for potential users taking part in education programmes.

Extent of Data Quality Information

Initial feedback from the data quality team at NZHIS showed that the DQEF did not provide sufficient information to make decisions on the quality of data. There appeared to be insufficient detail to manage data quality effectively. However, once the team became more familiar with the DQEF and began to understand the context of its use, the feedback became more positive. Users realised that the purpose of the DQEF was to provide guidance on the regular processes and measures needed to answer criteria questions. By presenting a more complete view of the data quality of a collection, these "answers" suggested new measures, not previously thought of by the data quality team. In effect, the DQEF is an invaluable guide or checklist that facilitates the consistent application of appropriate and wide ranging measures across all collections. It has the ability to raise the standard of work expected in data quality by bringing about awareness of areas of deficiency in the current work programme.

Interestingly, findings from the mortality data collection evaluation show that the Business Intelligence Team (as data consumers) required the most detailed information on how the assessment was made for each criterion, whereas managers required only summary information. This variation reflects the different decisions made by the different groups and hence the distinctive uses and required granularity of data quality information. Clearly, reports on the outcomes of assessments should be tailored to provide this information specific to audience requirements. In support of this observation, Chengalur-Smith, Ballou, and Pazer (1999) noted the importance of information format in their research on the impact of data quality information on decision making. They found that users required complex data quality information to make simple decisions but that they did not make use of this information for complex decisions. We consider this reciprocity may be due to "information overload" where too much information can be counterproductive to decision making in more complex decision environments.

Although decisions made in the health care environment are often "complex," what a complex decision is to one user may not be to another. This implies that users need to have input into decisions on the granularity of the supplied data quality information. Chengalur-Smith et al. (1999) note that it may be most effective to provide users with data quality information that focuses on one or two key criteria that exhibit the greatest impact on the data.

A later study by the same group (Fisher, Chengalur-Smith, & Ballou, 2003) observes the impact of experience on the decision maker, whereby increasing use of data quality information is found as experience levels progress from novice to professional and suggests that data quality information should be incorporated into data warehouses used on an ad-hoc basis.

Metrics

The DQEF uses subjective metrics for data criteria that have been developed by the collections' data custodians. Whilst this is a valid form of measurement, the robustness of the DQEF requires additional objective metrics and these are derived from a structured system based on statistical process control (Carey & Lloyd, 2001). The metrics include measures that assess customer requirements for levels of data quality, trends in historical data within the national health collections, current key performance indicators for contracted data suppliers, and legislative requirements for the provision of data by health care providers and the Ministry of Health to international bodies such as the World Health Organisation. Further work is still required to develop applicable metrics.

In summary, the pilot of the DQEF in the NZHIS environment elicited these issues:

- Training was required before using the DQEF.
- Users felt considerable time was needed to complete the evaluations.
- The extent and detail of information (particularly about how evaluation decisions are made) provided by the DQEF assessment process must meet the needs of data users, such as the Business Intelligence Team.
- Granularity/units of analysis need to fit the type of collection; registries such as the NHI require more granular, element level analysis.
- Language in the user manual and DQEF is an important consideration.
- Further work is required on metrics development.

Overall, collection users and managers found the DQEF to offer sufficient information on data quality to make preliminary, prioritised lists of essential data quality improvement projects. Further training has been necessary to ensure assessors use the DQEF consistently and that it is a practical and easy tool to use. The CIHI also found considerable training and change management were necessary to implement the CIHI Framework due in part to already heavy workloads (Long & Seko, 2002).

Pilot of the Framework in a Hospital Environment

The purpose of this pilot was to assess the application of the DQEF to the clinical integrated system (CIS) model that is currently used in clinical practice at Auckland District Health Board (A+). The assessment would indicate the applicability of the DQEF and its chosen data quality dimensions in the wider health sector. The CIS Model is an interdisciplinary computerised model of patient care that replaces paper notes and requires all staff members to document care via the computer. The software programme was developed by the A+ Network Centre for Best Patient Outcomes. This model has been in clinical practice since 2000 and has been used for 6,000 patients (Fogarty, 2004). The main objectives were to review the DQEF criteria against the CIS model database to:

- Determine the appropriateness of the criteria for a clinical database
- Assess the documented data quality of the CIS model against the DQEF criteria
- Assess the clarity of DQEF documentation

Positive feedback was given on the usefulness of the information provided by the assessment process and the applicability of the DQEF. The majority of the criteria could be applied to an external clinical database as shown in Table 1. This table highlights that 52 out of the possible 69 criteria used in the DQEF conform to the data quality requirements of the clinical database held at the local hospital level. The language used in the DQEF was further improved to ensure consistent understanding using the detailed feedback on each criterion provided by this assessment.

The DQEF evaluation process also proved valuable to the hospital submitting the clinical data set. It was suggested by the hospital data analyst who completed the evaluation that some formal, sectorwide criteria based on the DQEF, together with a certification process such as accreditation, would help to ensure that clinical databases are valid and reliable.

Evaluation of the Framework Against Eppler and Wittig Criteria

The resulting DQEF, with changes as recommended by focus group participants and the hospital environment, was assessed using Eppler and Wittig's (2000) criteria. The feedback can be found in Table 2. The recommendations centre on the development of business processes that support the effective use of the DQEF in the Ministry of Health environment and the content of the manual that instructs users on the implementation methodology. This supports Eppler and Wittig's (2000) theory that a framework requires tools, such as guidelines and manuals, to be effectively implemented. CIHI also found that a detailed step by step guiding process is required to implement a framework (Long & Seko, 2002).

The summary information gained from evaluations of all national health data collections was collated to form a prioritised list of data quality improvement initiatives across the ministry. Ongoing assessment using the DQEF will provide information on the success of such initiatives. The DQEF remains an iterative tool, whereby those that use the tool will improve its usefulness with growing knowledge of data quality theory, the level of data quality of the national health collections, and the priorities for improvement. In particular, the data quality improvement strategy described in the remainder of the chapter will support the ongoing refinements to the DQEF.

Development of the Health Data Quality Strategy

As mentioned earlier, little has been published on what constitutes a data quality strategy, let alone an evaluation of a structured and tested improvement programme. It is particularly important to note that a data quality improvement strategy is not an "information technology strategy," nor an "information systems strategy." Although such strategies may provide insight and tools to underpin a data quality improvement strategy, data quality improvements cannot be attained merely through information technology; the problem is one of processes and people. As noted in Ward and Peppard (2002), "clearly, technology on its own, no matter how leading edge is not enough."

Technology can support the achievement of an organisation's business goals at an operational level by increasing efficiency and at a strategic level by reengineering the business processes to be more effective. At both levels, however, the technology is merely an enabler and it is information and its management that makes the difference. Since the quality of the information rests on the quality of the raw data used to derive it, data quality is a critical issue. Technology can be used to improve data quality in some processes and also to exploit the superior quality for better business outcome in others.

Unfortunately, whilst the technical aspects of data quality assurance are usually well-documented, documentation of the business processes that support good data quality is often lacking. Information processes must be carefully documented to make the meaning of data transparent to all users. For example, data sources should always be identified to provide the data user with context around the data collection process. Imposing standards and maintaining consistency across business units with data definitions, business rules, and even systems architecture can lead to greater data integration and utility and hence perceived quality. This is where a "whole of health sector" strategy begins to have a significant effect on the quality of data and where centralised leadership is required to ensure that the views of all those who manage and use data are taken into account. Thus, good documentation carries the two-fold importance of establishing standards and managing change.

Change management is the main driver of the health data quality strategy (DQS) that, together with the data quality dimensions and framework, comprise the overall Data Quality Improvement Programme at the New Zealand Ministry of Health.

The DQS uses the dimensions derived in the DQEF to set data quality standards for designing, developing, and maintaining the national health data collections and data management throughout New Zealand. A practical strategy has to consider the complexity of the health sector since the data, their structure and use, and the products, are potentially much more varied than are found in financial or manufacturing organisations.

The development of the strategy was informed throughout by several stakeholders. Ongoing communication with health sector groups, including the Ministry of Health staff was essential to ensure sector buy-in and to maintain input and interest in the strategy development and implementation. Full consultation with a wide range of data suppliers and users was also necessary. Finally, surveys and discussions with organisations outside of health, both within New Zealand and overseas, on the management of data quality and implementation of strategy help to inform the ongoing and iterative development of the strategy.

The standards established by the DQS aim to achieve the improved health outcomes by:

- Better decision making to target areas of improvement in national health data collections with a transparent prioritisation tool for data quality improvement projects
- Improved relationships with data suppliers, developing a whole-of-sector responsibility for data quality

- Improved awareness of a data quality culture throughout the sector
- Improved understanding of the processes involved in developing an information product
- The development of best practice guidelines leading to accreditation of data suppliers
- Education and support to data suppliers
- Minimum data quality requirements for existing and new national collections
- Minimum requirements for regular operational data quality initiatives across all national health collections and an approach to rooting out persistent quality errors

In particular, the last bullet point demands a systematic approach to data quality management such as total data quality management (TDQM) mentioned earlier in the chapter.

Some approaches to data quality management target specific errors within a collection or an entire collection but often do not devise solutions to prevent systemic problems. In contrast, TDQM focuses on two key aspects of data management: the data flow processes that constitute the organisation's business and the recognition that information is a product, rather than a by-product, of these processes (Wang, 1998). Regarding the processes themselves, TDQM seeks to ensure that none of them changes the initial meaning of the data leading to systematic errors and repeated data quality problems. Systematic process errors can be prevented by several means, some of which will depend upon the nature of the business unit and its data. However, we can identify generic prevention mechanisms across business units to include:

- The systematic and ongoing education of data suppliers
- Education within the Ministry of Health
- Regular reviews of recurrent data quality problems from suppliers and information feedback to suppliers on issues with support provided for improvement
- Internally developed data quality applications to reduce time spent on assessment of data quality problems (limited in use for complex health data)
- A continuous cycle of define, measure, analyse, and improve using the framework to inform the assessment process

Concerning the role of information as a product of process, TDQM draws attention to the need to manage information so that organisations:

- Know their customers/consumers of the information and their information needs
- Understand the relationship between technology and organisational culture
- Manage the entire life cycle of their information products
- Make managers accountable for managing their information processes and resulting products

TDQM at the Ministry of Health

The foundation of the New Zealand data quality strategy is the institutionalisation of TDQM principles by which data quality management processes are embedded in normal business practice and accepted at all levels in the sector. This installation is achieved by the "define, measure, analyse, improve" sequence, as described in the next four sections.

Define

The definition of what data quality means to health sector professionals is clearly a fundamental requirement and this was done through the development of the data quality dimensions and the data quality framework. The dimensions define what health care data collectors, custodians, and consumers consider important to measure.

A cross-organisational data quality group is used to facilitate the discussion of data quality issues from an organisation-wide perspective. The group includes representatives of business units that collect, manage, or use data, as well as clinicians and clinical coders. The strategy then provides for whole of sector input into the data quality dimensions considered important, where data quality improvement should be targeted, and how and where accountability should lie. Continual assessment of the needs of data customers is required and will be managed through a yearly postal survey of customers and discussions at forums and training sessions.

Measure

Active and regular measurement of data quality avoids a passive reliance on untested assumptions as to the perceived quality of the data. Reporting on data quality levels becomes transparent and justifiable. The regular measurement programme involves:

- Regular use of the data quality dimensions and evaluation framework on national collections
- Developing appropriate statistical process control (SPC) measures
- Developing data production maps for all major information products outlining flow of data around the organisation and the health sector and possible areas of data quality issues
- Checking the availability of data quality assessment software or the potential for in-house development

Measuring the quality of data supplied by organisations to the national collections involves performance measurement based on key performance indicators (KPIs) developed through SPC (Carey & Lloyd, 2001). SPC uses historical information to define the ranges of acceptable data so that outliers or variations identify areas for improvement that can be negotiated with data suppliers.

Data production maps for all major information products are another active management tool for improving data quality. These maps model the flow of data around the organisation and highlight potential quality issues by detecting points where data quality may be compromised. When properly developed and supported by senior management, they can be effectively applied to solve complex, cross-functional area problems (Wang, Lee, Pipino, & Strong, 1998). Where data cross from one organisation to another, significant data quality issues can arise, and this is evident within the New Zealand health care system. Maps can also be used to summarise issues for managers and provide them with a tool to understand the implications of data quality.

Analyse

The complexity of health care requires an extensive range of decisions, both administrative and health-related, and a single data element can substantiate many decisions and carry impacts across the sector. But what level of data quality do we need to make good decisions? The importance of data quality analysis is that it allows us to determine the appropriate quality levels and identify areas for improvement where current processes do not support good data quality work. Analysis can draw upon:

- **Clinical analysis and data modelling:** what cannot be done due to data quality
- Surveys of customer needs
- Consideration of current and future uses of data
- Realistic expectations vs. expectations of stakeholders
- **Corporate knowledge:** what the organisation already knows is important and/or relatively easy to fix
- **International standards:** does the organisation compare well with appropriate standards and are our expectations realistic?
- Resources available to make improvements

Improve

Actively finding data quality issues before they cause problems is made easier by the regular assessment of collections using the data quality measurements outlined previously. Prevention, rather than cure, is now a large part of the ministry's work through the proactive approach outlined here:

- Regular minimum data quality initiatives for all collections with specific initiatives for single collections only
- Preventing poor data quality in new collections through a "data quality plan" developed in the early phases of projects with full involvement of the data quality team. Data quality is embedded in the system and processes of the new collection prior to going live
- Continuing "what works" as the data quality team have learnt a considerable amount about the operational requirements of maintaining data quality levels
- The endorsement of standards for the collection, management, storage, and use of specific data elements by the Health Information Standards Organisation; this way all stakeholders know and agree to all steps in the data flow process
- Stewardship and data access policies that provide clear ownership and access guidelines for the data
- The proposed development of a National Data Dictionary to address the standardisation of the major data elements found in the National Health Data Collections. This standardisation is likely to provide for considerable improvement to data quality through a nationally consistent understanding of data definitions

Through these initiatives it is expected that expensive, one-off projects can be avoided and money can be allocated for regular operational requirements that will enable an ongoing prevention programme. Projects may only improve processes and data in one collection, whereby regular prevention mechanisms help to ensure all data are of high quality.

TDQM in the Wider Health Sector

Some of the wider benefits of institutionalising TDQM through the strategy include:

- Getting everyone talking about data quality through the agreement of a strategic direction
- Raising the profile of data quality within health care organisations through organisation-wide data quality groups
- Getting the sector moving in the same direction through the development of organisation-wide data quality strategies that align with the national strategic direction
- Drawing on current practice/knowledge through best practice guidelines developed by the sector and widely disseminated
- Clear expectations of data suppliers through accreditation/KPIs/contracts
- Actively reviewing strategic direction regularly

A key issue raised by the data quality proposals described in this chapter is how to identify and disseminate best practice and embed it in normal day-to-day operation. The approach favoured in New Zealand is accreditation rather than audit. Accreditation suggests devolvement, ownership, and a supporting role from the ministry. For example, the ministry requires District Health Boards to produce annual plans but the boards themselves will be responsible for addressing data quality issues within a national framework and getting them approved at the highest level of regional health care funding and provision.

The ministry, through the Quality Health New Zealand and/or the Royal Colleges health care provider's accreditation process, will provide sector organisations with clear guidelines on how to achieve accreditation as good data suppliers. The accreditation process will be developed in consultation with the sector following the development and testing of best practice guidelines. Those organisations that have been able to implement extensive data quality programmes, as outlined previously, will be accredited as good data suppliers. This should lead to a reduction in the need for peer review and audit.

The proposed sector work programme requires that health care organisations can achieve accreditation if they:

- Take part in a sector peer review/audit process
- Meet KPIs
- Implement an in-house data quality education programme
- Develop and implement a local data quality improvement strategy
- Organise regular meetings with a cross-organisational data quality group
- Align with best practice guidelines (when developed)

CONCLUSION

This chapter has described the development of a health data quality programme for the Ministry of Health in New Zealand. The ministry's purpose is to realise the full value and potential of the data that it collects, stores, and manages. Building "trust" in the data throughout the health sector will ensure that data are used frequently and to their greatest possible benefit. With the right framework and strategy, data that are highly utilised for a range of reasons will incrementally improve in quality. Extensive data mining, combining currently disparate collections, will also provide far more granular information and knowledge to improve these collections. A data quality strategy will provide coherent direction towards total data quality management through a continuous cycle of work. The improved data quality will then ensure that the health sector is better able to make informed and accurate decisions on health care policy and strategy.

These aims are relevant to national and regional health authorities and providers in many countries and the key principles, described here and developed from pioneering work by the Canadian Institute for Health information, can be deployed successfully in many health care and related scenarios.

In summary, for the development and application of the data quality evaluation framework, it is important to:

- Define the underpinning data quality criteria carefully involving all stakeholders to ensure common understanding and direction
- Consider the critical quality dimensions that reflect how the organisation uses data and how data flow throughout the business processes
- Document business processes identifying data sources and their reliability
- Appreciate that the framework requires practical supporting tools (e.g., documentation) to make it effective
- Customise the language of the data quality user manual with regard to the level and experience of the intended users
- Be aware of the importance of both education and training at all necessary stages and levels; training is essential to affect the culture change that must accompany the realisation of the importance of data quality
- Be aware that application of the framework is an iterative, ongoing process; the required outcomes cannot be achieved in a single pass

For the data quality improvement strategy, it is important to:

- Derive and impose standards that facilitate data and information transfer whilst preserving quality
- Reengineer the business processes to deliver the quality data needed for efficient service planning and the effective practice of integrated patient care
- Identify and disseminate best practice to reduce the development time needed to improve data quality
- Ensure data quality levels are not unnecessarily rigorous to maintain user ownership and workloads at reasonable levels
- Define user accountabilities for data quality and the mechanisms to enforce them
- Seek to embed the search for data quality in normal working practices and recognise its achievement in appropriate ways such as with accreditation

Work on extending and improving the health data quality programme continues, particularly with regard to the formulation of objective metrics for the underlying data quality criteria.

REFERENCES

Al-Shorbaji, N. (2001). *Health and medical informatics.* Cairo: World Health Organisation.

Ballou, D. P., & Pazer, H. (2003). Modeling completeness versus consistency tradeoffs in information decision contexts. *Transactions on Knowledge and Data Engineering, 15*(1), 240-243.

Ballou, D. P., & Tayi, G. K. (1999). Enhancing data quality in data warehouse environments. *Communications of the AMC*. Retrieved April 9, 2006, from http://delivery.acm.org/10.1145/300000/291471/p73-ballou.pdf?key1=291471&key2=3297921511&coll=porta&dl=ACM&CFID=11111111&CFTOKEN=2222222

Canadian Institute for Health Information. (2003a). *Data quality framework.* Ottawa, Ontario: Author.

Canadian Institute for Health Information. (2003b). *Earning trust. Key findings and proposed action plan from the Data Quality Strategies Study*. Ottawa, Ontario: Author.

Canadian Institute for Health Information. (2005). *Data quality framework June 2005 revision*. Ottawa, Ontario: Author.

Carey, R. G., & Lloyd, R. C. (2001). *Measuring quality improvement in healthcare. A guide to statistical process control applications*. Milwaukee, WI: Quality Press.

Chengalur-Smith, I., Ballou, D. P., & Pazer, H. (1999). The impact of data quality information on decision making: An exploratory analysis. *IEEE Transactions on Knowledge and Data Engineering, 11*(6), 853-864.

Davis, F. (2003). *Data quality strategy.* Retrieved April 9, 2006, from http://www.firstlogic.com/_gics10643/DQ/article.asp?articleID=32

Deming, W.E. (1982). *Out of the crisis.* Cambridge: Massachusetts Institute of Technology.

Department of Health. (2004). *A strategy for NHS information quality assurance*. Author.

Department of Health and Aging. (2003). *Health Connect*. Canberra: Author.

Department of Health and Human Services. (2002). *HHS guidelines for ensuring and maximising the quality, objectivity, utility, and integrity of information dissemination to the public.* Retrieved April 9, 2006, from http://www.hhs.gov/infoquality/part1.html

Eppler, M. J., & Wittig, D. (2000). *A review of information quality frameworks from the last ten years.* Retrieved April 9, 2006, from http://www.knowledgemedia.org/modules/pub/view.php?id=knowledgemedia-2&page=1

Fisher, C., Chengalur-Smith, I., & Ballou, D. P. (2003). The impact of experience and time on the use of data quality information in decision making. *Information Systems Research, 14*(2), 170-188.

Fogarty, A. (2004). *Review of the MOH data quality evaluation framework on a clinical database*. Auckland: Auckland District Health Board.

Gates, B. (1999). *Business @ the speed of thought: Using a digital nervous system.* London: Penguin Books.

Juran, J. M., & Godfrey, A. B. (1999). *Juran's quality handbook* (5th ed.). New York: McGraw-Hill.

Levitin, A. V., & Redman, T. C. (1993). A model of the data (life) cycles with application to quality. *Information and Software Technology, 35*(4), 217-223.

Long, J. A., Richards, J. A., & Seko, C. E. (2002). *The Canadian Institute for Health Information Data Quality Framework, Version 1: A meta-evaluation and future directions.* Retrieved April 9, 2006, from http://secure.cihi.ca/cihiweb/dispPage.jsp?cw_page=quality_e

Long, J. A., & Seko, C. E. (2002). *A new method for database data quality evaluation at the Canadian Institute for Health Information.* Paper presented at the 7th International Conference on Information Quality, Boston.

New Zealand Health Information Service. (2003). *Current state analysis of data in the national health collections.* Wellington: Ministry of Health.

Pautke, R. W., & Redman, T. C. (2002). The organisation's most important data issues. In G. M. Piattini C. Calero, & M. F. Genero, (Eds.), *Information and database quality.* Dordrecht: Kluwer.

Price, R. J., & Shanks, G. (2005). *Empirical refinements of a semiotic information quality framework.* Paper presented at the 38th International Conference on Systems Sciences, Hawaii.

Redman, T. C. (2001). *Data quality. The field guide.* Boston: Digital Press.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research*. Thousand Oaks, CA: Sage Publications.

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM, 40*(5), 103-110.

UK Audit Commission. (2002). Data remember. Improving the quality of patient-based information in the NHS. In *Management Paper* (p. 42). London: Author.

UK Audit Commission. (2004). *Information and data quality in the NHS*. London: Author.

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM, 39*(11), 86-95.

Wang, R. Y. (1998). A product perspective on total data quality management. *Communications of the ACM, 41*(2), 58-63.

Wang, R. Y., Lee, Y. W., Pipino, L. L., & Strong, D. M. (1998, Summer). Manage your information as a product. *Sloan Management Review*, 95-105.

Wang, R. Y., Strong, D. M., & Guarascio, L. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5-33.

Ward, J., & Peppard, J. (2002). *Strategic planning for information systems* (3rd ed.). Chichester, UK: John Wiley & Sons.

WAVE Advisory Board. (2001). *From strategy to reality. The WAVE (Working to Add Value through E-information) project. Health Information Management and Technology Plan* (Project Report). Wellington: Ministry of Health.

Willshire, M. J., & Meyen, D. (1997). A process for improving data quality. *Data Quality, 3*(1). Retrieved from http://www.dataquality.com/997meyen.htm

ENDNOTES

[1] A national collection is either a long-term collection of nationwide data or a reference dataset, of which NZHIS is the custodian on behalf of the sector, and which is used for analysis and information provision to improve the services and capabilities of the publicly funded health and disability sector.

[2] Version 1 of the CIHI Framework was superseded by a revised version 2 in June 2005 after this work was completed. The two versions (Canadian Institute for Health Information, 2003a, 2005) are essentially the same although the newer version has fewer criteria (58) but has added "documentation" as a sixth dimension.

Chapter 2.14
QoS Provisioning in Sensor Enabled Telemedicine Networks

Chunxiao Chigan
Michigan Tech, USA

Vikram Oberoi
Highmark Inc., USA

ABSTRACT

Enabled by the advances of the wireless sensor network technologies, wireless LANs will play a critical role in providing ubiquitous connectivity for future telemedicine applications. This paper focuses on how to provide QoS over the wireless channel between the Body Sensor Network (BSN) Gateway and the wireless Access Points (AP). Telemedicine applications require the periodic report data and the emergency messages transmitted to the remote health care center in a timely manner. However, unlike the voice and multimedia applications which can be supported by traditional QoS techniques, the sporadic nature of the emergency data in telemedicine systems makes it nontrivial to provide sufficient QoS. This article investigates several alternative schemes for QoS support in the telemedicine systems, and an express dual channel (EDC) based QoS mechanism is proposed. Not only is the proposed mechanism simple and resource efficient, but also it provides the bounded maximum delay guarantee for the unpredictable emergency data transmission for telemedicine applications.

INTRODUCTION

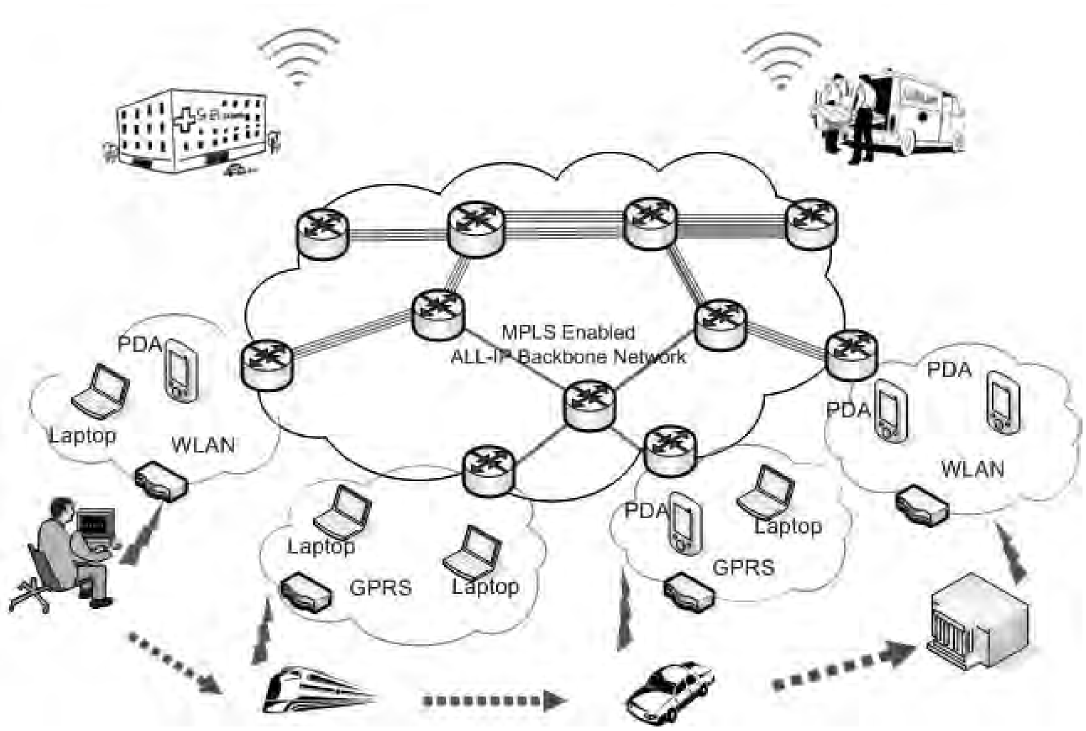
With the advances in wireless technology, the realization of seamless connectivity to support "anywhere" and "anytime" communications becomes more and more plausible. One of the most important advancements is the development of low cost wireless sensors networks to collect data in various environments. In this paper, we explore one such environment, namely the telemedicine systems. Chronic diseases such as diabetes and heart attacks are the most common and costly health problems in the United States. Fortunately, with the help of adequate support infrastructure and seamless connectivity, many fatal situations may be prevented. If the data related to the periodic monitoring reports and potential emergency situations of the patients can get across to medical facilities in a timely manner, even though the patient is free to roam about, many life-death critical situations can be taken care of.

A telemedicine system which can support such application, is depicted in Figure 1, where the data regarding various physiological parameters such as heart rate, blood pressure, and so forth, collected by the in-body sensors, is fed to and aggregated by the Body Sensor Network (BSN) gateway which can be some handheld device (e.g., a modified palmtop or PDA). This BSN gateway then sends this data to a wireless Access Point (AP, supported by IEEE 802.11) connected to the Internet, which further relays the data to the remote medical facilities. In such a telemedicine system, the BSN gateway fuses information from wearable, ingestible, and implantable sensors to provide continuous diagnostic monitoring and generate early warnings. It may also provide control over critical implantable devices such as drug delivery pumps. The emergency or the periodic data fused by a BSN gateway is then relayed via a WLAN access point to remote health care centers. The communication between the WLAN access point and remote health care center takes place via an all IP-based network.

To deploy such a telemedicine application which is life-death critical, certain QoS guarantees have to be provisioned so the data can reach the remote health care center in a reliable and timely manner. This paper focuses on the QoS provisioning between the BSN gateway and the wireless access points (e.g., IEEE 802.11 based Hot Spots) in the telemedicine applications.

We assume that the underlying media access control (MAC) protocol of the WLAN access point network is the IEEE 802.11e standard which supports the traditional QoS service to various classes of applications. Compared to the baseline IEEE 802.11 which does not support any QoS service, IEEE802.11e provides a basic platform to support different priority of the wireless media access to different data traffics if the traffic

Figure 1. Conceptual system architecture of the sensors-enabled telemedicine system



pattern is of certain predictability. However, the occurrence of emergency events in telemedicine application is extremely sporadic and its traffic pattern is totally unpredictable. Therefore, the QoS provisioning for such applications is not trivial. Rather, it can be an extremely difficult task if the traditional scheduling-based QoS approaches or the traditional contention based on-demand QoS approach patterns are applied for the QoS provisioning in telemedicine systems.

In this article, we propose a completely new method of providing QoS support for telemedicine applications with very little complexity and cost-effective QoS guarantees. The rest of the article is organized as follows. In the second section, we present the related work and discuss the uniqueness of the QoS requirements in telemedicine applications. In addition, four alternative potential approaches to support QoS in telemedicine applications are also investigated in this section. The third section presents the details of our proposed Express Dual Channel (EDC) based QoS solution for telemedicine applications. In the fourth section, we present our simulation results, and the fifth section concludes the paper by presenting conclusions and future work.

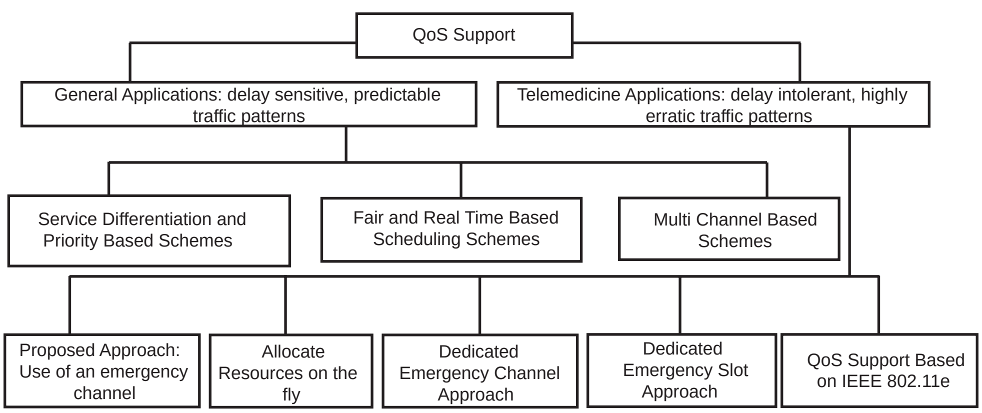
RELATED WORK

In this section, we present the existing approaches providing traditional QoS services and identify the unique challenges of QoS provisioning in telemedicine systems. We then explore the potential approaches to meet these unique QoS requirements in telemedicine applications.

Current QoS Approaches (Traditional QoS)

There are many solutions proposed to support the traditional wireless QoS applications where different classes of data demands are known in advance. For example, it is known that some applications (e.g., video, voice) are more delay sensitive than the others (e.g., data). Because the traffic patterns of these delay sensitive data (video and voice) have been extensively studied and well understood, we refer to the solutions providing QoS to applications of predictable traffic patterns as *traditional QoS* approaches. The current approaches to support traditional QoS requirements can be grouped into various categories. We have classified (refer to Figure 2) some of these approaches which are closely related to our work into three categories. The following is a brief overview of these approaches.

Figure 2. Diagrammatic representation of classification of QoS various approaches



Service Differentiation and Priority (SDP) Based QoS Schemes

The schemes under this category basically use different values of various contention parameters such as contention window size and so forth, to assign different priorities to different services. The work by Ada and Castelluccia (2003), Xiao (2003), and Veres, Campbell, Barry, and Sun (2001) belong to this category. One such scheme has been proposed in Ada and Castelluccia (2003), wherein the authors present four service differentiation schemes for IEEE 802.11. The first scheme is based on scaling the contention window according to the priority of each *flow* or *user*. For different users with different priorities, the second, the third, and the fourth mechanisms assign different minimum contention widow values, different interframe spacing and different maximum frame lengths, respectively. The SDP-based QoS approach can also be used to provide priorities to real-time applications (Xiao, 2003). By differentiating the initial window size, the window-increasing factor and the maximum backoff stage, an analytical model for assigning priorities to real-time applications was proposed in Xiao (2003).

In the SDP-based QoS approach, service differentiation can also be introduced using a fully distributed approach. In Veres et al. (2001), based on the IEEE 802.11 Distributed Coordination Function (DCF) mode, two distributed estimation algorithms are proposed. A Virtual MAC (VMAC) algorithm passively monitors the radio channel and estimates locally achievable service levels. VMAC estimates key MAC level statistics related to service quality such as delay, delay variation, packet collision, and packet loss.

In all of the above schemes, priority is assigned to different services by assigning favorable contention parameters. Therefore, data demanding high priority can only achieve best-effort QoS with unbounded maximum delay. This is due to the fact that the high priority data still has to contend for channel access with the lower priority and equal priority services.

Fair and Real-Time Scheduling (FRS) Based QoS Schemes

The protocols under this category use fair queue scheduling principles to attain QoS. Vaidya, Bahl, and Gupta (2000), Adamou, Khanna, Lee, Shin, and Zhou (2001), and Ranasinghe, Andrew, Hayes, and Everitt (2001) fall into this category, wherein the scheduling can be done based on the traffic analysis for different applications.

Using one or more parameters, such as finish time for a packet, some scheduling policies are developed. One such approach presented in Vaidya et al. (2000) is based on the Self Clocked Fair-Queuing approach of transmitting the packet whose finish tag is the smallest, as well as its mechanism for updating the virtual time. A distributed approach for finding the smallest finish tag is employed, using the back off interval mechanism of IEEE 802.11 MAC. The essential idea is to choose a backoff interval that is proportional to the finish tag of the packet to be transmitted.

The approaches under this category try to attain a specific scheduling goal depending upon the nature of the application. For instance in Adamou et al. (2001), the authors present four different algorithms designed to attain certain specific scheduling goals. The first two are EDF (Earliest Deadline First) and GDF (Greatest Degradation First) that consider only one aspect of the scheduling goal, respectively. EDF is naturally suited for maximizing throughput, while GDF seeks to minimize the maximum degradation. The next two are algorithms, called EOG (EDF or GDF) and LFF (Lagging Flows First), which consider the two aspects of the scheduling goal. EOG simply combines EDF and GDF, whereas LFF tries to favor lagging flows in a nontrivial manner.

Similar to different scheduling goals, the QoS objectives can also be differentiated according to the nature of the application (Ranasinghe et al., 2001). One QoS objective could be to maximize the number of customers receiving good service for real time data services such as multimedia data, MPEG, video streaming, IP telephony, and so forth. In Ranasinghe et al. (2001), a new scheduling scheme called the Dual Queue discipline is proposed which has the flexibility to satisfy a variety of QoS objectives, ranging from existing notions of fairness to maximizing the number of customers receiving good service.

All the schemes under this category, however, require the prior knowledge of the different traffic patterns, which is unavailable in telemedicine applications wherein the traffic pattern of emergency messages is highly unpredictable.

Multi Channel (MC) Based QoS Schemes

In MC-based QoS schemes, usually separate control messages are used to set up a separate dedicated channel or reserve the primary channel for data transmission. Dedicated Channel for Streaming Services (DCSS) (Tsai & Wang, 2005) uses some new control frames to assign a new dedicated channel for heavy traffic or streaming services. On the other hand, Dynamic Channel Assignment (DCA) (Wu, Lin, Tseng, & Sheu, 2000) divides the overall bandwidth into one control channel and “n” data channels. Whenever a node wants to communicate with another node, it sends control packets over the control channel to obtain rights to access one of the data channels. The data channels carry only the Data and the Acknowledgements packets.

Another QoS mechanism is to use different reserved channels for communicating with different neighbors. Sequenced Neighbor Double Reservation (SNDR) (Cai & Lu, 2000) is one such MAC protocol for Multi-Channel Ad Hoc networks. The basic idea is that, for a given node A, if one of its neighbors B wants to send data to A, B must use the same code sequence preassigned to A to send data. In other words, A can receive data only on the channel it has been preassigned. If more than one neighbor wants to send data to A, then they use TDM over that channel.

One of the drawbacks of the schemes mentioned in this category is the delay caused in setting up the separate channel or gaining access to the main data channel. In addition, valuable resources may be underutilized because of reservation of separate channels for separate nodes.

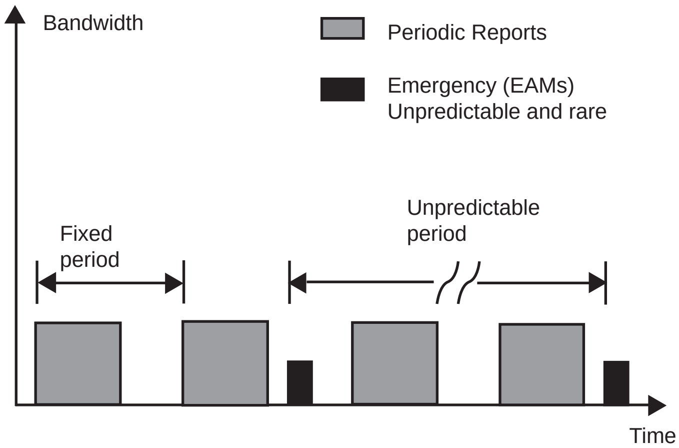
Alternative Approaches (QoS in Telemedicine Applications)

In this section, we first discuss the unique QoS requirements and challenges for telemedicine applications. Then we discuss why the previously surveyed existing approaches may not work well to provide QoS support for such applications. Certain approaches that may be employed for providing QoS guarantee for telemedicine applications are then explored. We use the sensor enabled telemedicine systems illustrated in Figure 1 as an example in the rest of paper for discussion purpose. The QoS issues and principles discussed in this paper, however, are applicable to a variety of sensor enabled telemedicine systems.

Unique QoS Requirements of Telemedicine Applications

As described above, there are many traditional solutions proposed to support the traditional wireless QoS applications where different classes of data demands are known in advance. However, as shown in Figure 3, telemedicine application usually comprises of two types of data: the regular periodic report data which are sent at fixed time intervals and are of predictable nature, and the emergency messages that have highly erratic natures. Therefore, providing QoS support for emergency situations can be very tricky because of this erratic nature.

Figure 3. Traffic pattern for a telemedicine application with emergency situations



To highlight the uniqueness of QoS requirements for emergency data, let us compare its properties with that of traditional QoS applications such as voice streaming and multimedia traffic.

First, the emergency events in telemedicine application are extremely sporadic and unpredictable, therefore may occur just once in several months/years. However, although emergencies occur sporadically, emergency telemedicine data is extremely delay-intolerant and demands bounded maximum delay guarantee. Hence, traditional resource reservation and the scheduling-based QoS approach cannot be applied to this kind of telemedicine applications efficiently. On the other hand, the traditional on-demand QoS approaches are contention-based and may lead to unbounded maximum delay, and therefore cannot ensure the QoS requirements of the delay intolerant telemedicine applications either. Thus, all traditional QoS approaches (resource reservation and scheduling based, or contention based on-demand approach) can hardly suit the telemedicine applications.

Indeed, none of the traditional QoS approaches discussed previously can work well in telemedicine applications. The *SDP Based QoS Schemes* may experience unbounded maximum delay for gaining the channel access due to the contention among different and same priority services. The *FRS Based QoS Schemes* require traffic analysis and thus demand the prior knowledge of traffic patterns that is not available for telemedicine applications. And the *MC Based QoS Schemes* may require unbounded maximum delay in establishing a separate channel or gaining access to one of the data channels. This is also due to the contention and collisions that may occur over the control channel.

Alternative Approaches for Telemedicine QoS

Next, we explore potential approaches that might be applicable for QoS in telemedicine applications.

1. **Allocate Resources On-the-Fly:** Anytime a device with emergency message to be delivered via the access point (AP), it probes the availability of an AP in its vicinity (or passively listens for a beacon from an AP). Then it associates with the AP. After the association phase, it ***contends*** for the channel and can only send the emergency data after obtaining channel access. The obvious limitation of this approach is the nondeterministic delay introduced. In the worst case (with unbounded maximum delay) the

device with emergency event may not get the channel access in the acceptable delay period.

2. **Dedicated Emergency Channel Approach:** Another approach could be to reserve a whole separate channel for transmitting the emergency data. However, as emergencies occur rarely and follow the bursty pattern, this often leads to significant wastage of the resources.
3. **Dedicated Emergency Slot Approach:** A similar approach is that the AP reserves a slot of channel access for the telemedicine application by the periodic beacon signal it sends out. If the application indicates that there is an emergency event in this slot, the AP grants the channel to the telemedicine application. However, the beacon is sent out following the scheduled (determined) intervals, whereas emergencies are of highly unpredictable nature. Thus, if the emergency event occurs after the beacon was sent out, the emergency data cannot be transmitted until the next beacon period. In addition, because the emergency data has to be transmitted within the reserved slot, the periodic share of wireless channel access needs to be large enough to support at least one telemedicine application at a time. With the unpredictable nature of the emergency message, this *slot reservation based approach* is therefore subject to the same wastage of resources as the *dedicated channel based approach*. In general, the larger the reserved slot, the less access delay the telemedicine applications (one or more) experience, while the more significant the resource wastage is.
4. **QoS Support Based on IEEE 802.11e:** This approach uses the Hybrid Coordination Function (HCF) in the IEEE 802.11e standard (refer to appendix for details) to provide higher priority for the telemedicine applications. Under the HCF mode, because the AP is given the highest priority for the wireless media access, it can grant access to nodes with different priorities based on certain criteria. During the Contention Free Period (CFP), the AP can issue a QoS Contention Free (CF) Poll to a particular station to grant it a channel access. The AP schedules these polls based on the traffic analysis sent to it by each station regarding their queue length, and so forth. During the Contention Period (CP), all the stations contend for channel access via Enhanced Distributed Coordination Function (EDCF). In this case, if the AP is triggered properly by the emergency demands, it can grant channel access to the telemedicine stations by sending out a beacon signal. However, for telemedicine applications, the traffic is of a highly unpredictable nature which could render the traffic analysis/predication a very complex issue. Therefore, it would be very difficult to design a scheduling algorithm that can capture the unpredictable nature of the emergency data.

In this paper, we propose a novel mechanism to provide QoS for telemedicine applications. Not only is this mechanism simple and resource efficient, but also it provides bounded maximum delay for the unpredictable emergency data transmission.

EXPRESS DUAL CHANNEL (EDC) BASED QOS PROVISIONING FOR TELEMEDICINE APPLICATIONS

In this section, we present an entirely new and radical approach for providing QoS support in telemedicine applications. We use the telemedicine system depicted in Figure 1 as an example for explanation, but the proposed approach is applicable to most of the telemedicine applications where wireless media access controls are involved.

As described in the first section, for a sensor enabled telemedicine system, the patients are equipped with in-body or wearable sensors which can send the data from different parts of the body to the Body Sensor Network (BSN) Gateway (e.g., modified handheld devices such as PDAs or Palmtops). The BSN gateway sends the aggregated data (periodic reporting data, and unpredictable emergency data) to the remote health care centers via the public Wireless Access Points (APs, e.g., WiFi) which are connected to the Internet. Because APs often support many other applications in the public domain, there are contentions among different applications to gain the wireless channel access from different nodes. Furthermore, multiple telemedicine applications within one AP also contend among each other to gain channel access. In this context, our paper focuses on the QoS support for telemedicine applications by guaranteeing maximum (worst) delay of the public APs access, and at the same time minimizing the resource wastage due to the unpredictability of the emergency data.

Assumptions and Notations

The assumptions that were made are listed below:

- The AP is able to transmit and receive data simultaneously.
- The level of interference experienced by two nodes is the same in both directions (transmission and receiving). The interference level does not change during the transmission period.
- Nodes send association and disassociation messages to the AP when they enter and leave the network, respectively.

The following is the summary of the notations (IEEE 802.11e compatible) used in this paper:

- SIFS—Short Interframe Space
- PIFS—Point Interframe Space
- DIFS—Distributed Interframe Space
- HC—Hybrid Coordinator
- AP—Access Point
- TXOP—Transmission Opportunity
- TBTT—Target Beacon Transmission Time

Network Dynamics in Telemedicine Applications

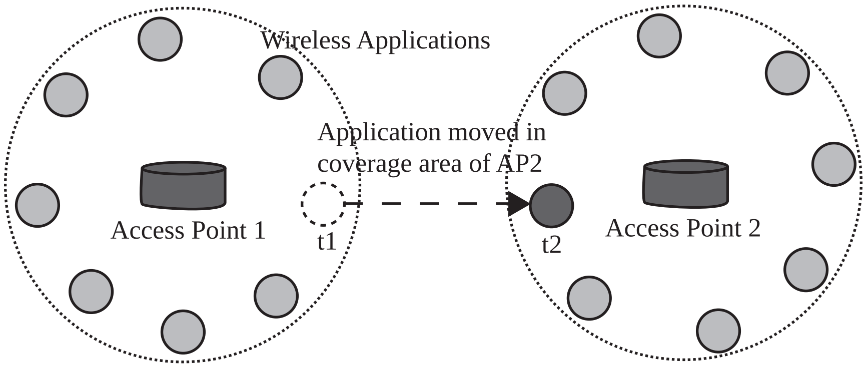
To enable the BSN gateway sending data to remote health center seamlessly, APs at different locations (different spots on campus, shops, airports, etc.) can be used to provide pervasive network connection. In this case, the individual wireless local area network is in infrastructure mode. However, the telemedicine applications supported by enormous numbers of APs are still of ad hoc nature because the BSN gateway can be one time within the coverage of one AP, and another time move to the vicinity of a different AP. This demands a perfect hand-off technology for the WLAN networks, which is beyond the scope of this article.

The above situation is depicted in Figure 4. The network configuration (in which a telemedicine device connected to) at time t1 is different from that at time t2 when some telemedicine applications have moved from the coverage area of AP1 to the coverage area of AP2.

Revisit: Traffic Patterns and Unique QoS Requirements in Telemedicine Applications

As discussed previously, the QoS support required by the telemedicine applications is different to that of conventional applications due to the extreme unpredictability of the emergency data pattern. Conventional techniques for providing QoS, which are based on analyzing the traffic patterns of different flows, do not suit such kinds of telemedicine

Figure 4. Network configuration at time t1, t2



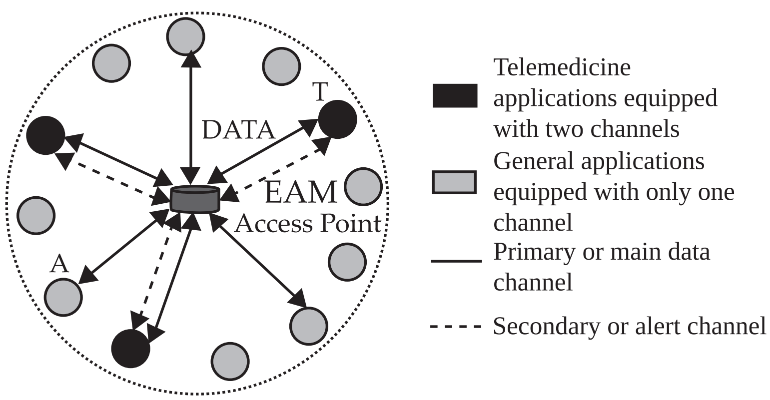
applications due to the highly unpredictable nature of the traffic. In general, for such applications, the emergency events occur very infrequently, and therefore often it is desirable to have the periodic report messages (probably with a slow repetition cycle) sent for event monitoring/predication purpose and to maintain the presence of the telemedicine devices and their association with the network access point.

Supporting Emergency QoS in Telemedicine System: Express Dual Channel-Based Mechanism

To provide QoS guarantee for telemedicine applications supported by wireless LAN, we propose the use of an unsymmetrical dual channel mechanism. With such a dual channel mechanism, the primary channel is dominated for all data transmission, and the secondary channel is very slim, which is dedicated to the short emergency alert message (EAM) transmission.

Figure 5 depicts a scenario where such a dual channel mechanism is used to provide QoS support for a telemedicine application in case of an emergency. Before an emergency occurred, a general application device A was sending data to an AP. Now, as soon as the emergency occurred, the telemedicine application device T sends out a short Emergency Alert Message (EAM) through the reserved secondary, or alert channel. Upon receiving the EAM alarm, the AP schedules (with the highest priority) to send out a beacon to all the applications in the network, indicating the reservation of resources for device T.

Figure 5. Diagrammatic representation of channel access via the use of dual channel



The AP can gain access to the channel anytime because of its highest priority to access the channel under the Hybrid Coordination Function (HCF) mode supported in IEEE 802.11e. Upon receiving the beacon frame, telemedicine device T can send out the emergency data over the primary data channel. At the same time, the regular application device A losses its access to data transmission channel. Only devices with telemedicine requirements are equipped with this dual channel. A separate secondary channel is reserved for sending the short EAM to the AP. Whenever an emergency occurs, the BSN gateway sends a short EAM over this reserved channel to the AP to request the channel access for its emergency data. With this reserved secondary channel, the short EAM can arrive in a timely manner at the AP. Upon receiving the EAM, the AP can schedule to send out a beacon frame indicating an instant/highest priority reservation of resources for this station over the primary (normal) channel for its emergency data transmission. Once the requesting station is granted to access the medium, it can then send the complete emergency data over the primary data transmission channel.

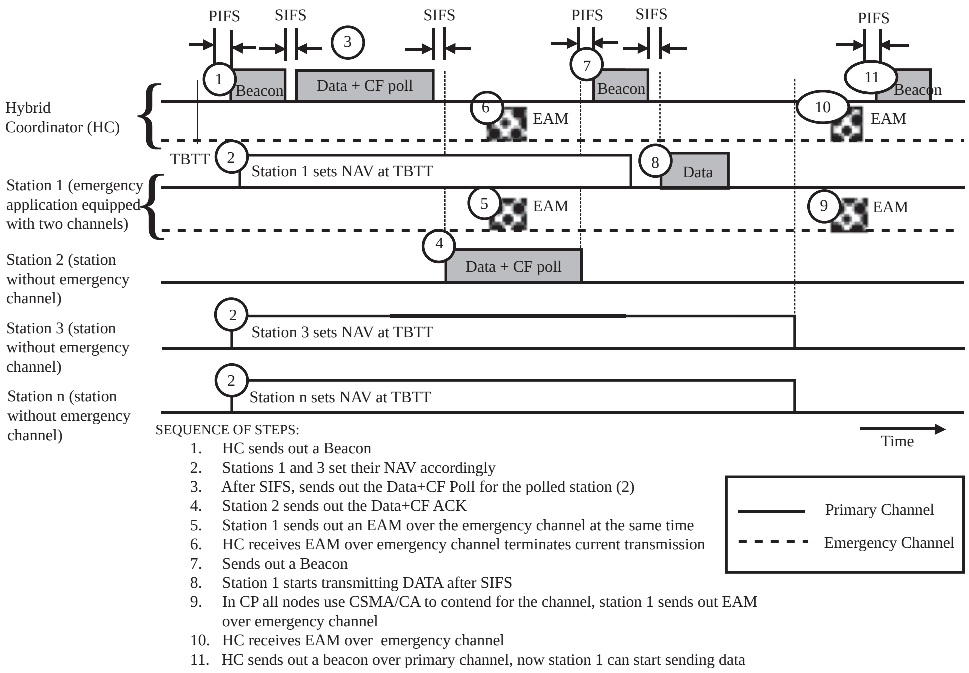
The reserved secondary channel therefore requires a very small amount of bandwidth as the EAMs are very short messages to report the occurrence of the emergency event. Consequently, such a static secondary channel reservation at the APs for the telemedicine applications will not be costly even when there are no devices of telemedicine applications associated with the APs.

Because there could be more than one device of telemedicine applications within one AP's coverage, there can be a contention over this reserved secondary channel too. To resolve such a complication, the AP can apply a Time Division Multiplexing (TDM) scheme over the reserved secondary channel to allow a fixed number of such applications that it can accommodate to associate with itself. This number will depend on the maximum delay these telemedicine applications can tolerate and the statistics of the specific telemedicine applications. Once the EAM reaches the AP through the reserved secondary channel, the primary data channel access will then be scheduled for that application, and thus emergency requests can be processed within a certain time, and hence the delay in processing multiple such requests will be deterministic. To allocate the available bandwidth between the dominant primary data channel and the slim secondary emergency alert channel, Orthogonal Frequency Division Multiplexing (OFDM) (Tutorial: Multi Carrier Modulation and OFDM. http://zone.ni.com/devzone/cda/ph/p/id/150), (Lawrey, 1997) or Frequency Division Multiplexing (FDM) (Tutorial: Multi Carrier Modulation and OFDM. http://zone.ni.com/devzone/cda/ph/p/id/150) or Code Division Multiple Access (CDMA) can be used. Details of these methods can be found in Lawrey (1997). Because these are all mature commercial technologies, the cost of the hardware introduced to support our proposed express dual channel (EDC)-based QoS solution will be rather low.

Figure 6 gives an in-depth description of the wireless channel acquisition for the proposed dual channel-based QoS support for telemedicine applications. Station 1 is a telemedicine application device and is equipped with two channels. Stations 2 and 3 are regular application devices and can send messages over the primary data channel only. The whole process is divided into two periods, the Contention Free Period (CFP) and the Contention Period (CP). The CFP is fixed and bounded by two TBTT (Target Beacon Transmission Time). During the Contention Free Period (CFP), the AP or the Hybrid Coordinator (HC) polls the stations for data transmission by sending a beacon, whereas during the CP, the nodes contend for channel access (refer to appendix for details).

First, let us have a look at channel access during CFP. Let us assume at time 0, the AP has polled and given the Transmission Opportunity (TXOP) to station 2. Station 1 and 3 set the Network Allocation Vector (NAV) to the next Target

Figure 6. Channel access during contention free period (CFP) and contention period (CP) in express dual channel based QoS Mechanism



Beacon Transmission Time TBTT, assuming that all the polling slots are full and it will not have the TXOP during this period. Thus, station 2 can send its data to the AP during its TXOP over the primary data channel.

Now, when an emergency happens at telemedicine station 1, it sends out an EAM over the secondary emergency alert channel to the AP. Once the AP receives this EAM, it terminates the TXOP of station 2 and sends out a beacon frame over the primary channel after waiting for the Point (coordination function) Interframe Space (PIFS) duration. This beacon indicates the reservation of the primary channel for telemedicine station 1 and grants the TXOP to station 1 for the required duration. Stations 2 and 3 can set their NAVs accordingly.

Now let us consider the Contention Period (CP). During the Contention Period (CP), all the stations contend for channel access via Enhanced Distributed Coordination Function (EDCF) (refer to appendix for details). But the AP still can grant access to the channel by broadcasting a beacon. Thus, when station 1 detects an emergency, it sends an EAM over the secondary emergency channel. Upon receiving this EAM, AP broadcasts a beacon over the primary channel. Now, station 1 can send its emergency data over the primary data channel.

To provide QoS support for regular report data for telemedicine applications, we apply the IEEE 802.11e standard. For details of IEEE 802.11e MAC, please refer to the appendix. Such periodic data reports of telemedicine applications are grouped into a separate category with the second highest access priority at the primary data channel. The number of retransmission of such periodic report data is limited by the time after which the report will be outdated and not by a certain threshold number as in IEEE 802.11e standard. When operating in the EDCF mode, the telemedicine applications get the second highest priority to access the primary data channel and when operating in the Hybrid Coordinator (HC) mode, the Hybrid Coordinator (HC) located at the Access Point (AP) polls the telemedicine applications first. As the regular report data will be generated in a periodic manner, the AP can schedule its polling for these applications after it analyses their traffic patterns.

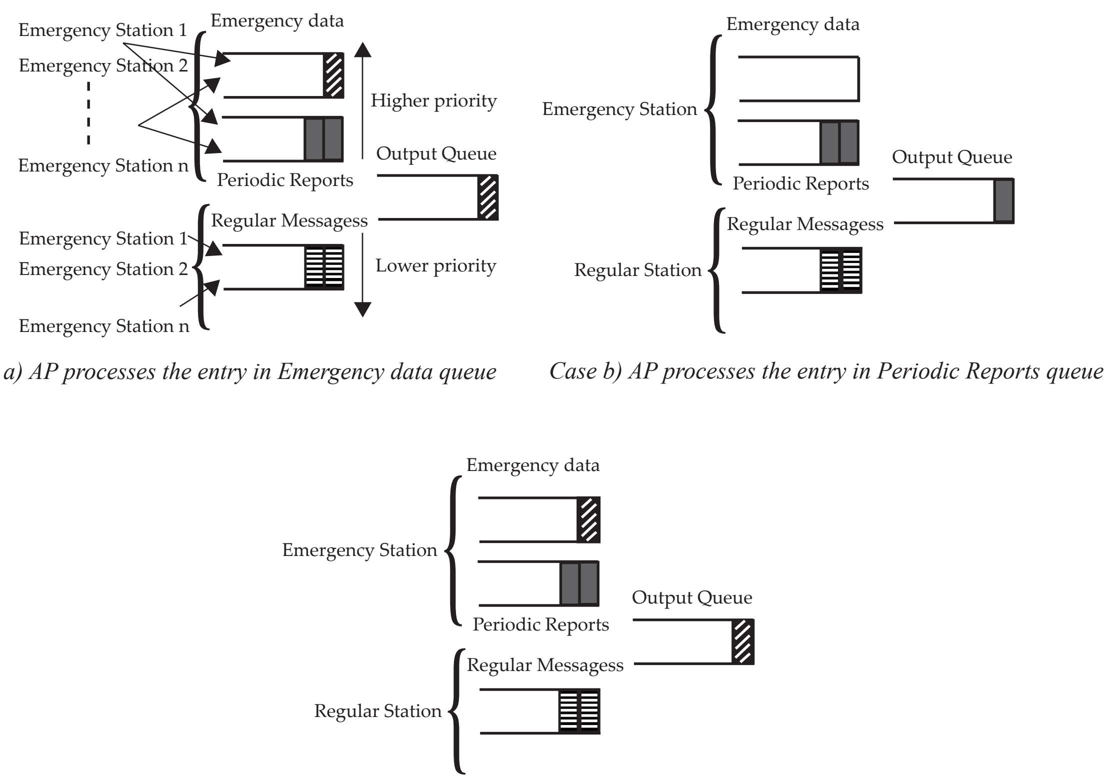
Queue Formation and Processing of Queue Entries at the AP

There are 3 classes of queues maintained at the AP in the proposed telemedicine QoS mechanism. The emergency data queue (EAM queue) has the highest priority. The queue with the second highest priority is dedicated to the periodic report data of the telemedicine applications. The queue dedicated to the other messages (regular applications) has the lowest priority. As shown in Figures 7 and 9, the entries in the EAM queue at the AP are given the highest priority. As long as there is an entry in the EAM queue it will be processed first, thus resulting in very small processing delay for such applications. Thus, as soon as an emergency occurs, the EAM can be sent to the AP without further access contention. When a station leaves the network, it sends a disassociation message to the AP and the AP can release the resources that were allocated to this station.

The proposed Hybrid "Emergency First" and "Fair Queue" algorithm works as follows (refer to Figure 8): the AP processor always checks the EAM queue first. If there is a message in this queue, it processes it; otherwise it goes on to check the queue of the periodic reports. Upon the completion of any message processing, the AP proceed to check the EAM queue first and then the other queues.

In Figure 7 of case "a", at the starting point, there was an entry in the EAM queue, so the processor at AP processes it. In case "b", as there is no entry in the EAM queue, the processor moves on to periodic data reports queue, and processes one entry and then again proceeds to check the EAM queue. If there is an entry in EAM queue,

Figure 7. Hybrid "Emergency First" and "Fair Queue" scheduling at the AP



Case a) AP processes the entry in Emergency data queue

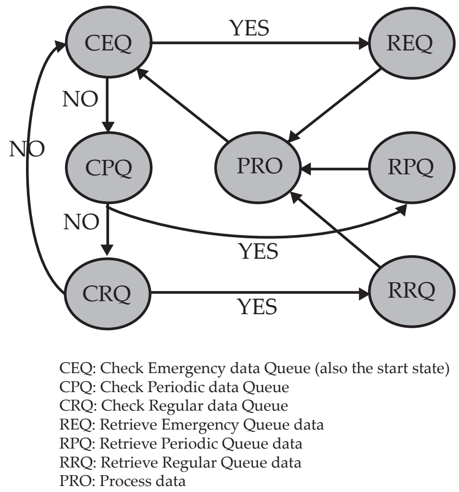
Case b) AP processes the entry in Periodic Reports queue

Case c) AP processes the entry in Emergency data queue

Figure 8. Hybrid "Emergency First" and "Fair Queue" algorithm

1. Check Emergency data Queue
   a. If entry exists,
      • retrieve data, go to Process,
      • go to step 1
   b. Else, go to step 2.
2. Check Periodic Reports Queue
   a. If entry exists,
      • retrieve data, go to Process,
      • go to step 3
3. Check Regular messages Queue
   a. If entry exists,
      • retrieve data, go to Process,
      • go to step 1
   b. Else, go to step 1

Figure 9. Finite state machine (FSM) of queue processing at AP



the AP will process it first and then move on to other queues. This is depicted in case "c". Notice that there can never be more than one message in the output queue at a time. This is due to the fact that the processor only processes one message at a time and then checks the different queues first before processing the next message. Thus, the maximum delay experienced by the emergency data after it reaches the AP is the time required to process one single data message request.

Figure 9 depicts the Finite State Machine (FSM) of our Hybrid "Emergency First" and "Fair Queue" Algorithm, which is described in detail in Figure 8. The FSM diagram and the algorithm (refer to Figure 8) depict the processing flow of queue entries at the Access Point. An EAM basically triggers the AP to reserve resources for its telemedicine application. As soon as an EAM reaches the AP, it sends out a beacon indicating the reservation of resources for the telemedicine application associated with the EAM.

This beacon also acts as an acknowledgment for the telemedicine station that its EAM has reached the AP. For periodic reports, the standard ACK packets can be used as the confirmation of their reception. Once the emergency data reaches the AP, it is placed in Emergency data queue. This queue is given the highest priority. All the queues are scanned by the AP in a round robin fashion, as described in Figure 8.

SIMULATIONS

As discussed in the second and third sections, our proposed Express Dual Channel Based QoS Mechanism can guarantee bounded *maximum delay* experienced by the emergency message transmission, and it can therefore ensure the QoS requirement of the telemedicine applications wherein the occurrence of the emergency event is rare and highly unpredictable. Indeed, without *a prior* knowledge of the traffic pattern of the emergency messages available, it is impossible for the traditional schedule-based QoS approaches to prereserve channel resources for telemedicine applications efficiently. On the other hand, the traditional on-demand QoS approaches are contention-based and may lead to unbounded *maximum delay*, and therefore cannot ensure the QoS requirements of the delay intolerant telemedicine applications either. In this section, we will evaluate the performance of our proposed QoS scheme in terms of the *average end-to-end transmission delay*, and the *percentage of the dropped data bits* via simulation.

We used OPNET, version 8.1A, for simulating the proposed QoS solution in WLAN networks. Various parameters used and their values are shown in Table 1. A total of 10 nodes were placed randomly, but all within the transmission range of the Access Point (AP). The underlying MAC protocol is the IEEE 802.11. IEEE 802.11e has not been simulated in OPNET. For the time being, we used IEEE 802.11 standard as the underlying MAC. The network is operating in infrastructure mode where all the nodes have PCF functionality enabled. For telemedicine stations, only the emergency data is considered, and the periodic report data is not included.

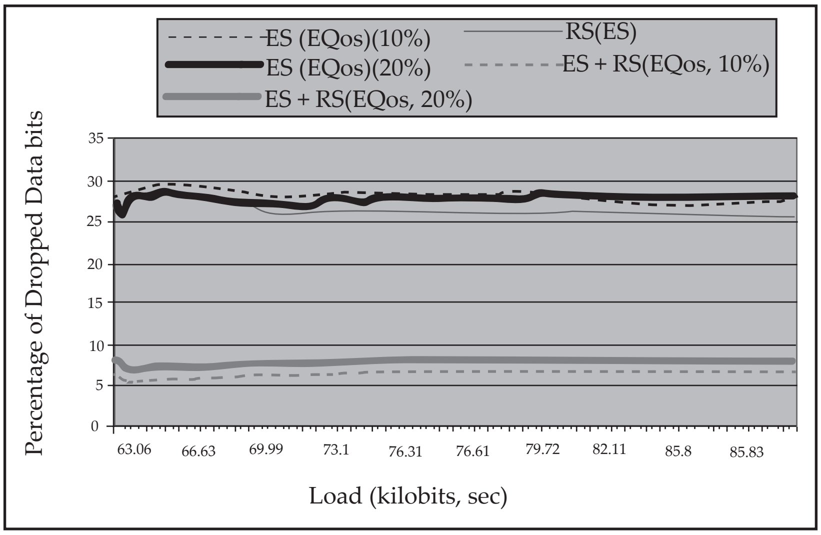
Figure 10 shows the average dropped data bits (kilobits) for the telemedicine stations and the regular stations vs. the average load for a station. Our proposed scheme is compared with the baseline scheme, which is the IEEE 802.11 standard in the PCF mode without priority given to any station. We used this scheme to form a benchmark for the performance of a regular station. Therefore, in the baseline scheme, all the stations are considered as regular stations as opposed to our proposed scheme, in which stations are divided among telemedicine stations and regular stations. *ES (EQoS) (10%)* denotes Enhanced Telemedicine Stations in our proposed telemedicine QoS scheme, and they are 10% of the total stations in the simulated WLAN network. Similarly, RS (BS) denotes Regular Stations in the Baseline Scheme. ES+RS (EQoS, 10%) denotes the overall performance of all telemedicine and regular stations in our proposed EQoS scheme when there are 10% telemedicine stations in the simulated WLAN networks. Conversely, when there are 10% telemedicine stations, then there will be 90% regular stations, and when there are 20% telemedicine stations, there will be 80% regular stations. We can see that the percentage of the dropped data bits are considerably less for the telemedicine station in our proposed scheme, as expected.

It is also shown in Figure 10 that although the *percentage of the dropped data bits* for a regular station in our scheme is greater than that in the baseline scheme, this difference is not much larger as compared to that in the baseline scheme. Hence, even though in our proposed scheme we concentrate on the telemedicine stations, the performance degradation for a regular station is not that large. The graph also shows that there is a slight increase in the percentage of the dropped data bits for the telemedicine stations when their number is increased from 10% to 20%. However, for overall performance of all regular and telemedicine station, this increase is negligible. This may be due to less contention experienced by regular stations as the number of regular stations contending for the same channel has decreased.

Table 1. Simulation parameters and their values

Parameter	Values used
Grid size	100*100 meters
Transmission range	30 meters
Traffic	Bursty
Network Scale	Office
Simulation Time	900 seconds
No. of stations	10
No. of stations equipped w/ emergency channel	2

Figure 10. Percentage of dropped data vs. average load



We can also see that there is a steady increase in the *percentage of the dropped data bits* with the increased load for the telemedicine stations. This is due to the fact that EAMs reach the AP through a time division multiplexed channel. Hence, when one telemedicine station gains the channel access, the other telemedicine station has to wait until the first one completes its transmission. Thus, with an increasing load, more data bits will be dropped as the waiting time will increase.

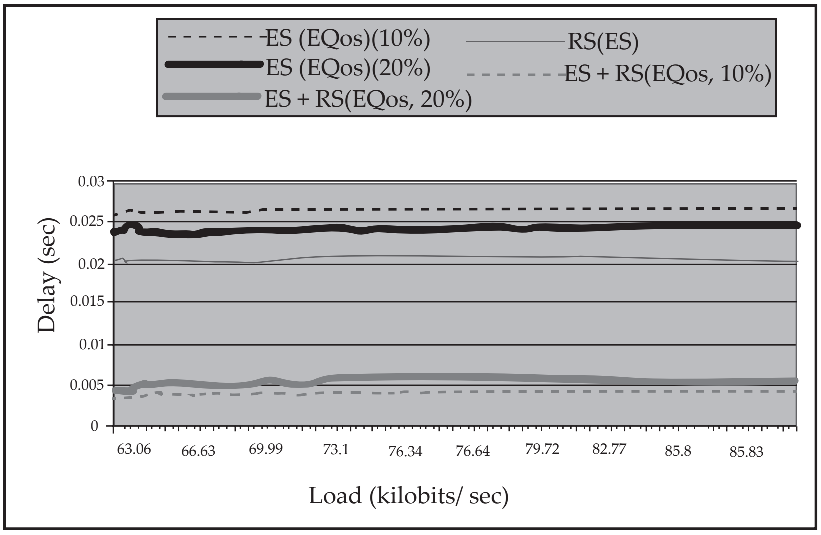
Figure 11 shows the *average delay of the end-to-end transmission* for the telemedicine stations and the regular stations vs. the average load for the station. As the load for the station increases, the average delay experienced by both the telemedicine and the regular station increases. However, the delay experienced by the telemedicine station is much less than the delay experienced by the regular station in our scheme. This is due to the fact that whenever a telemedicine station needs to send data to the AP, it can send an EAM over the emergency channel and get the channel access scheduled in a deterministic and negligible time period. However, regular stations have to contend with other regular applications (which results in nondeterministic delay) or wait to be polled by the AP before they can start sending their data.

Again, even though in our scheme we concentrate on improving the performance of the telemedicine stations, the performance degradation for a regular station is not much as compared to the regular station in the baseline scheme. This can be inferred from the curve showing the overall average delay experienced by all the telemedicine and the regular stations, which is slightly higher for our scheme as compared to the baseline scheme. The average delay experienced by the telemedicine stations increases when their number is increased from 10% to 20%. This is again due to the fact that at higher loads a telemedicine station will have to wait for a longer period until the telemedicine station, which has gained channel access earlier, has finished its data transmission.

CONCLUSION AND FUTURE WORK

Due to the highly erratic nature of the telemedicine emergency events, designing mechanisms to provide QoS based on traditional methods is not trivial, if not impossible. This is mainly because the traffic pattern based traditional scheduling QoS solutions cannot be used for telemedicine application. If resources are prereserved, it will

Figure 11. Average delay vs. average load



result in a highly inefficient solution. On the other hand, the traditional on-demand QoS approaches are contention-based and may result in unbounded *maximum delay*, and therefore cannot ensure the QoS requirement of the delay intolerant telemedicine applications either. In this article, we proposed a novel mechanism to provide QoS for telemedicine applications. Our proposed dual channel-based QoS approach, where the slim express secondary channel is dedicated for transmitting short Emergency Alert Message (EAM), is an effective and simple method of providing QoS in ubiquitous telemedicine applications. The proposed solution uses the IEEE 802.11e standard as its baseline mechanism (Deng & Haas, 1998). The simulation results show that not only is this mechanism simple and efficient, but also it results in bounded maximum delay for the erratic emergency data transmission in telemedicine applications.

In the future, we will simulate and compare our proposed mechanism with the traditional QoS methods to estimate the overall performance improvements. We will also carry out an analysis on the maximum number of devices for telemedicine and general applications that an AP can accommodate simultaneously, given specific QoS requirements (e.g., maximum tolerable delay) of various telemedicine applications.

REFERENCES

Ada, I., & Castelluccia, C. (2003). Priorities in WLANs. *Computer Networks*, *41*(4), 505-526.

Adamou, M., Khanna, M.S., Lee, I., Shin, I., & Zhou, S. (2001). Fair real-time traffic scheduling over a wireless LAN. In *Proceedings of the 22nd IEEE Real-Time Systems Symposium, RTSS 2001*, London, UK.

Cai, Z., & Lu, M. (2000). SNDR: A new medium access control for multi-channel ad hoc network. In *Proceedings of the IEEE Vehicular Technology Conference (VTC-Spring)* (vol. 2, pp. 966-971).

Deng, J., & Haas, Z.J. (1998). Dual busy tone multiple access (DBTMA): A new medium access control for packet radio networks. In *Proceedings of the IEEE International Conference on Universal Personal Communications (ICUPC '98)* (vol. 2, pp. 973-977).

IEEE Std 802.1 WD3.3. (2002). *Draft Supplement to Standard for Telecommunications and Info-*

mation Exchange Between System - LANIMAN Specific Requirements- Part 1 I: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QOS).

Lawrey, E. (1997). *The suitability of OFDM as a modulation technique for wireless telecommunications, with a CDMA comparison.* Doctoral thesis, James Cook University.

Mico, F., Cuenca, P., & Orozco-Barbosa, L. (2004). QoS in IEEE 802.11 wireless LAN: Current research activities. In *Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering* (vol. 1, pp. 447-452).

Ranasinghe, R.S., Andrew, L.L.H., Hayes, D.A., & Everitt, D. (2001). Scheduling disciplines for multimedia WLANs: Embedded round robin and wireless dual queue. In *Proceedings of the IEEE International Conference Commun. (ICC)* (pp. 1243-1248).

Truong, H.L., & Vannuccini, G. (2003). Performance evaluation of the QoS enhanced IEEE 802.11e MAC layer. In *Proceedings of the 57th IEEE Semiannual Vehicular Technology Conference* (vol. 2, pp. 940-944).

Tsai, Y., & Wang, Y. (2005). MAC protocol for streaming services in wireless multi-channel MANETs. In *Proceedings of the IEEE Vehicular Technology Conference (VTC).*

Vaidya, N.H., Bahl, P., & Gupta, S. (2000). Distributed fair scheduling in a wireless LAN. In *Proceedings of the ACM 6th International Conference on Mobile Computing and Networking (MobiCom 2000)* (pp. 167-178).

Veres, A., Campbell, A.T., Barry, M., & Sun, L.H. (2001). Supporting service differentiation in wireless packet networks using distributed control. *IEEE Journal of Selected Areas in Communications (JSAC), Special Issue on Mobility and Resource Management in Next-Generation Wireless Systems, 19*(10), 2065-2081.

Wu, S., Lin, C., Tseng, Y., & Sheu, J. (2000). A new multi-channel MAC protocol with on-demand channel assignment for multi-hop mobile ad hoc networks. In *Proceedings of the International Symposium on Parallel Architectures, Algorithms and Networks* (pp. 232-237).

Xiao, Y. (2003). A simple and effective priority scheme for IEEE 802.11. *IEEE Communications Letters, 7*(2), 70-72.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 2, Issue 3, edited by J. Tan, pp. 12-30, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

APPENDIX A.

OVERVIEW OF IEEE 802.11E MAC

IEEE 802.11e (IEEE Std 802.1 WD3.3., 2002) has been proposed specifically to support QoS requirements. It introduces the concept of a Hybrid Coordination Function (HCF), which basically consists of two access schemes:

- The Enhanced Distributed Coordination Function (EDCF), which is an extension to the DCF in 802.11 to provide service differentiation via priorities, and
- The Hybrid Coordinator (HC), which is a modification to the existing Point Coordinator (PC) for more efficient polling schemes.

When using EDCF, multiple Access Categories (ACs) have been defined to provide QoS. Each AC uses the DCF; however with different values of the contention parameters DIFS, now called Arbitration IFS (AIFS). ACs with lower parameter values will experience lower mean waiting and backoff times and thus will have a relatively higher access priority to the medium. A station or a node which has multiple ACs in parallel will have internal contention between its own ACs in addition to the access contention with other stations.

According to the draft standard (IEEE Std 802.1 WD3.3., 2002), four different types of services, voice, video, video probe, and best effort, are considered with each of them having their own AC. Table 2 lists the parameter values which have been assigned to each of the four ACs. The values are chosen in such a way that the highest access probability is given to the “voice” AC and the lowest one to the AC supporting the best effort data traffic.

Table 2. Contention parameter values for different ACs

AC	AIFS	CW_{min}	CW_{max}	Service
0	2	15	1023	Best effort
1	1	15	1023	Video probe
2	1	7	15	Video
3	1	3	7	voice

In addition to the EDCF contention mechanism providing service differentiation, a polling-based access mechanism similar to the IEEE 802.11 PCF is also included in the 802.11e draft standard (IEEE Std 802.1 WD3.3., 2002). This polling scheme is controlled by a centralized hybrid coordinator (HC) located at the access point (AP). The HC can start a controlled access period (CAP) whenever needed, in order to poll traffic from the stations.

An important new attribute of the IEEE 802.11e MAC is that the station that obtains the medium access cannot utilize it longer than the duration greater than a specified limit. This attribute is known as a transmission opportunity (TXOP). A TXOP is defined by its starting time and duration. TXOPs obtained via contention-based medium access are referred to as EDCA-TXOPs, and TXOPs obtained by the HC via controlled medium access is referred to as HCCA-TXOP or polled TXOP. The duration of an EDCA-TXOP is limited by a TXOP-limit. This TXOP-limit is distributed regularly by the HC within an information field of the beacon.

Another enhancement is that no station can transmit across the target beacon transmission time (TBTT). This means that a station can send its frame only if it can be completed before the upcoming TBTT. This reduces the expected beacon delay, which gives the HC better control over the medium, especially if the optional contention free period (CFP) is used after beacon transmission.

Chapter 2.15
Designing Clinical Decision Support Systems in Health Care:
A Systemic View

Wullianallur Raghupathi
Fordham University, USA

ABSTRACT

Clinical decision support systems have historically focused on formal clinical reasoning. Most of the systems are rule-based and very few have become fully functional prototypes or commercially viable systems that can be deployed in real situations. The attempts to build large-scale systems without examining the intrinsic systemic nature of the clinical process have resulted in limited operational success and acceptance. The clinical function, another area of medical activity, has emerged rapidly offering potential for clinical decision support systems. This article discusses the systemic differences between clinical reasoning and clinical function and suggests that different design methodologies be used in the two domains. Clinical reasoning requires a holistic approach, such as an intelligent multiagent, incorporating the properties of softness, openness, complexity, flexibility, and generality of clinical decision support systems, while traditional rule-based approaches are sufficient for clinical function applications.

INTRODUCTION

Clinical decision support applications development in the field of medicine historically has been constrained by two factors. First, the design concept was limited to formal clinical reasoning based on expert physicians' rules of thumb. Second, design appears to have been technology driven, for nearly all of the current systems are rule-based, wherein rules in clinical reasoning (e.g., in diagnosis and treatment) are represented as rules in the clinical system. None of the clinical reasoning applications have yet to become fully functional prototypes or commercially viable (Bates, Kuperman, Wang, Gandhi, Kittler, Volk, & et al., 2003; Kaushal, Shojania, & Bates, 2003; Sim, Gorman, Greens, Haynes, Kaplan, Lehmann, & et al., 2001). The limited operational success of these large-scale systems (Kaplan, 2001) is due in large part to the failure to reflect more fully in the design the diverse systemic features of the clinical process. Newer sociotechnical approaches to design, for example, a multiagent approach, and more flexible representational methods are needed to produce viable clinical decision support systems (CDSSs).

A second area of medical activity, which we term here the clinical function, has emerged rapidly as health care participants (e.g., physicians, nurses, HMOs, hospitals, diagnostic labs) increasingly perform their clinical applications work using advanced information technology (IT), such as artificial intelligence-based (AI) systems (Wyatt & Spiegelhalter, 1991) and in-house medical staff. These new AI-based clinical function systems present design issues that differ from those of clinical reasoning systems. For example, these systems typically concern much smaller domains (e.g., alerting to drug interaction, monitoring patient vital signs, reminding to medicate patient) involving structured applications in medicine (Bates et al., 2003; Delaney, Fitzmaurice, Riaz, & Hobbs, 1999; Kaplan, 2001; Kawamoto, Houlihan, Balas, & Lobach, 2005; Ramnarayan & Britto, 2002).

The application of AI (e.g., expert systems) and other computational intelligence techniques (e.g., neural networks) to the field of medicine (Catley, Petriu, & Frize, 2004) has resulted in the attempts to develop CDSSs. Wyatt and Spiegelhalter (1991) have defined medical aids as "active knowledge systems which use two or more items of patient data to generate case-specific advice." CDSSs have the potential to analyze, synthesize and integrate patient-related information to perform complex evaluations and provide that information to clinicians in real time. Over time, as they evolve in sophistication, they offer the prospect of improving the effectiveness and efficiency of patient care by preventing medical errors and enhancing quality (Johnston, Langton, Haynes, & Mathieu, 1994).

Further, the systems can improve preventive care services and help in adhering to recommended care standards (Kawamoto et al., 2005). The overall goal is to improve clinical decision making by focusing on individual patient characteristics and mapping them to a computerized knowledge base of characteristics of similar patients (Garg et al., 2005). They provide a range of levels of decision support, from simple alerts to complex diagnosis. For example, a CDSS can aid a physician in processing complex information to improve prescription writing practices in electronically delivered recommendations (Durieux, Nizard, Ravaud, Mounier, & Lepage, 2000). These types of systems are differentiated from operational decision support systems (DSSs), which are defined as enterprise repositories of clinical and financial information for utilization review, cost evaluation, and performance evaluation (Classen, 1998). In contrast, CDSSs focus on medical decisions (both on making decisions and assisting in making decisions). The key is to use patient specific information that transforms protocols into customized, real-time clinical advice (Kawamoto et al., 2005; Teich, Osheroff, Pifer, Sittig, & Jenders, 2005).

This paper discusses the clinical decision support applications design issues arising from the systemic differences (Churchman, 1971; Van Gigch, 1978, 1991) between clinical reasoning and clinical function systems, and it proposes that different design methodologies be used in the two domains. The paper is organized as follows. First, design issues in clinical reasoning and clinical function are discussed. Second, knowledge representation issues are highlighted. Third, the systemic properties of softness, openness, complexity, generality and purpose are discussed in the context of the two domains. Fourth, domain issues are identified. Next, operational examples of clinical reasoning and clinical function applications are described. Finally, conclusions are offered.

DESIGN ISSUES IN CLINICAL REASONING AND CLINICAL FUNCTION

Clinical reasoning is concerned with the decision processes of medical experts, such as physicians and surgeons, the interpretation of clinical information, and medical diagnostic rules. The outcome of clinical reasoning is some kind of clinical decision (diagnosis). On the other hand, the clinical function has emerged as a differentiated organizational entity. It is defined here as the set of activities that involves the routine and procedural application of medical rules and protocols to the operational needs of the medical care delivery organization. Differences between clinical reasoning and the clinical function are distinguished in terms of several system concepts. First, we discuss clinical knowledge representation and system design.

Clinical Knowledge Representation and System Design

Despite the domain complexity, most expert systems in clinical reasoning have involved the one-to-one mapping of rules of medicine onto rules in the system. This type of representational approach does not provide the flexible basis necessary to develop a sociotechnical clinical reasoning system that affords the openness, softness, complexity, and generality to deal with the dynamics of sociomedical aspects of clinical reasoning. Purely rule-based representation and problem solving are suitable for many clinical function applications that often involve narrow, structured well-designed domains. Knowledge representation for clinical reasoning, however, must include objects, frames, semantic nets, rules, and combinations of these to adequately represent different levels of knowledge, in combination with abstract medical concepts at higher levels and clinical rules at lower primitive levels. Combined with a multiagent systems model for complex problem solving, these representational modes enable interaction between multiple sources and levels of medical knowledge and different types of clinical reasoning, which comprise the problem-solving methodologies. The use of multiagent models will be a major departure from the current limited, rule-based approaches for implementation of AI clinical decision support systems.

Softness and Openness

Natural systems have been characterized as being somewhat hard, mechanistic, and rigid (Van Gigch, 1978). While clinical reasoning does involve some rigidity in terms of medical rules/procedures, the application of clinical reasoning reflects the relatively soft nature of the sociomedical aspects of medicine (e.g., consideration of gender, ethnicity, culture, religion, individual patient customized protocol). Softness considers subjectivity and value as important considerations in the design of clinical reasoning applications. Applications in the clinical function, on the other hand, usually involve specific routine procedures or repetitive tasks. Design models for the clinical function, therefore, should be more rigid and mechanistic in view of the procedural nature of this applied medical work.

Clinical reasoning involves open system properties (Hewitt, 1985), for, as an integral part of society and health care organizations, the medical system continually processes inputs and feedback from the environment, including changes in medical practice, new/modified protocols, new drugs, effect of clinical trials, societal values, replacement of medical experts, organizational perspectives (e.g., hospital mission), and others. The medical systems, as a model at the cusp of hard (pure science) and soft (social science) sciences, exhibits its negentropic nature as clinical reasoning adapts to changing goals (e.g., prevention vs. treatment), new research findings and advances in medicine, different sources of knowledge, and learning from experience. The clinical reasoning system must be capable of arriving at similar decisions equifinally by accepting inputs from multiple experts and applying alternate reasoning paths and different approaches to problem solving (Hewitt, 1985).

While the medical literature emphasizes the open-textured nature of medicine, openness is a feature of the entire medical system. Openness in the design of clinical reasoning systems provides for the flexibility to include several interacting domains in clinical reasoning, accepting inputs from multiple external sources of information and expertise, and accommodating the various reasoning mechanisms that are part of clinical decision processes.

In contrast, applications in the clinical function are essentially closed-system models requiring minimum feedback. Dynamic interaction and flexibility are limited because of the structured nature of clinical function tasks.

Complexity in DSSs

Complexity in CDSSs results from interaction internally between subsystems and externally with various systems in the environment, such as the economic (insurance) and primary care systems, as well as government entities, such as Medicare and Medicaid. Within an organizational health care delivery system (e.g., a hospital), there is also interaction and interdependence with the legislative bodies, the judiciary, the National Institutes of Health, the American Medical Association, medical experts, ethicists, managers, and the general public. The design of expert clinical reasoning systems must include provisions for the various modes of inquiry (Churchman, 1971) and complex interactions among various sources of clinical knowledge, such as medical facts, protocols, judgmental experience of medical experts and health care managers (e.g., nurse managers, discharge planners), customs, and medical precedents. Since societal values also impinge on outcomes, holism in design is essential to deal with the sociomedical complexity.

In the design of CDSSs for clinical reasoning, the trade off has historically been in formulating a computer model that is well structured and rigid in specification as opposed to one that provides for flexibility and adaptability. Strictly reductionaistic, rationalistic approaches to the design of clinical decision-making applications effectively eliminate these essential features. Mutiagent models hold considerable promise for assimilating the variety of inputs for complex clinical reasoning. The clinical function has emerged in health organizations because of increased organizational complexity and environmental uncertainty (e.g., hospitals). The growth of the clinical function was part of a general pattern of organizational elaboration and differentiation of specialized functions. While clinical reasoning is complex by nature, the clinical function is relatively simplistic, and interaction with the environment is minimal. Because of the procedural nature of most clinical function activities, rule-based (procedural) system approaches to application design are adequate and most often used.

Generality and Purpose in CDSSs

Research in AI frequently has emphasized the need for generality in the problem-solving methods. This is an important issue, for clinical reasoning applications should be capable of handling a wide range of medical problems.

Current design efforts for AI-based systems in clinical reasoning, however, tend to be limited to a single domain of medicine or reasoning. Generality in clinical reasoning suggests that the system cannot be limited to a single domain or a particular type of reasoning mechanism. For purposes of design, more general problem-solving models for the clinical reasoning domain have to be investigated. Again the multiagent model (e.g., the blackboard approach) possesses the characteristic of generality, and being conceptual, also can accommodate features that are unique to particular domains. In contrast, clinical function tasks are situation specific and do not require the generality needed by clinical reasoning systems.

Clinical reasoning applications are typically designed to arrive at a clinical decision. Examples of such systems include diagnosis of the cause of patient's chest pain (Hunt, Haynes, Hanna, & Smith, 1998); treatment of infertility (Garg et al., 2005); implementation of clinical guidelines on venous thromboembolism prophylaxis in an orthopedic surgery department (Durieux et al., 2000); analysis of outcome of patient cardiovascular risk (Montgomery, Hahey, Peters, MacIntosh, & Sharp, 2000); interpretation of blood work results (Classen, 1998); and others. Clinical reasoning culminates in decisions (e.g., outcomes—diagnosis of a disease, treatment selection) as the result of the interaction among various sources of knowledge and types of reasoning. The goals for clinical reasoning applications are relatively vague and ill defined. Clearly, a holistic and integrative view of design is needed, rather than the current focus on segmentized, disparate approaches that compartmentalize types of clinical reasoning, including analysis, critiquing, diagnosis, pattern recognition, and treatment recommendation, to name a few (Ramnarayan & Britto, 2002).

Clinical function applications are designed to execute programmed decisions involving narrowly defined medical tasks to achieve specific goals in the health care delivery organization. Some generic examples of such clinical function applications include detection of drug overdose (Hunt et al., 1998); selection of appropriate drug dosage, providing immunization reminders (Garg et al., 2005); alerting to drug interaction and monitoring patient vital signs (Classen, 1998); and others. Typically, the clinical function activities include alerting, recognition, alerting, and monitoring (Kawamoto et al., 2005; Ramnarayan & Britto, 2002).

Domain Issues

The variety of issues faced by clinical decision processes poses significant problems for domain definition in design. Historically, the boundary identification problem in design has been addressed by limiting the system domain to a particular specialty of medicine (e.g., disease), type of reasoning (e.g., rule-based, case-based), or problem-solving approach. Such design limitations have led to the problems with clinical reasoning systems discussed previously. The design of clinical function applications, while not trivial, is less challenging, for the boundaries of individual systems are limited to specific tasks. It is becoming more common for these tasks to be performed by nurse practitioners, nurses, and clinical technicians, who have gained expertise in routine and repetitious tasks in narrow medical domains and who are being increasingly supported by clinical function systems. Examples of clinical function applications include nurse practitioner functions. By and large these are the most readily developed for rule-based applications. Nurse practitioners carry out routine repetitive tasks and develop expertise in such domains as alerting, monitoring, and reminding and, therefore, provide input for clinical decision processes at higher levels. A number of generic applications are currently available to support nurse practitioner tasks. Most of the applications are fairly simple and implemented as rule-based systems. However, they lack the sophistication necessary to support clinical reasoning tasks.

Examples of clinical reasoning tasks include diagnosis, interpretation, and planning. All these high-level clinical decisions are characterized by complexity, ill-structuredness, and uncertainty. Typically, such decisions are collectively made by a group of medical experts. Table 1 summarizes the example of clinical reasoning and clinical function tasks in health organizations. As seen in Table 1, most programmed tasks are performed at lower and middle levels. Rule-based approaches are suitable for (expert system) applications development for these tasks. The higher-level tasks are typically nonprogrammed, ill-structured, and complex. Multiagent models are more appropriate for applications supporting these activities.

Table 1. CDSSs

Health Organization Level	Problem Type	Tasks
Top		
Physicians Surgeons Medical directors Residents Specialists	Ill-structured, complex Nonprogrammed (multiagent model)	Diagnosis Causal analysis Clinical reasoning Recommendation Interpretation Prediction
Middle		
Physician assistants Nurse practitioners Nurses	Semistructured, somewhat Comples, nonprogrammed (single to multiagent)	Alerting Monitoring Clinical reasoning Clinical function Reminding
Lower		
Nurses Clinical assistants Discharge planners	Structured, routine, Programmed (single angent)	Alerting Reminder

Other Issues

The identification of users and experts in the clinical function is straightforward, while deciding whether the physician is the only user or the expert in clinical reasoning is not always so clear. Clinical reasoning involves deeper epistemological and ontological considerations; the development time for a clinical reasoning application is lengthy; the cost is very high; the design and implementation effort is extensive; risks are considerable (e.g., misdiagnosis); and the payoffs are uncertain. In contrast, clinical function applications involve shorter development time, low costs, and considerably less effort. A clinical function application can be built using available technology, such as shells (rule-based expert system). Unlike clinical reasoning applications, these systems do not require extensive medical knowledge on the part of the knowledge engineer. Knowledge acquisition is easier and application validation is more readily achieved. User acceptance of clinical function applications is higher, especially for systems that act as "intelligent clinical assistants" for medical task performance (e.g., drug dosage tracking, monitoring patient vital signs), leading to increased potential for use and commercial viability. Table 2 summarizes the systemic differences between clinical function and clinical reasoning.

Operational Examples of Clinical Reasoning and Clinical Function Systems

With regard to clinical reasoning systems, St. Mary's Medical Center in West Palm Beach, FL, maintains statistics about infants (e.g., vital signs) in an electronic documentation system. This data is analyzed and mined, for example, to identify potential risks and side effects from commonly used medications (Landro, 2006). At Partners Health in Boston, expert systems provide information on medication dosing for specific types of patients, taking into account several factors.

Table 2. Clinical reasoning versus clinical function

Concept	Clinical reasoning	Clinical function
Nature of model	**Open** **Soft** **Flexible** **Dynamic** **Holistic**	**Closed** **Hard** **Inflexible** **Axiomatic** **Sequential**
Solution	**Satisficing**	**Optimal**
Reasoning	**Opportunistic** **Mulitagent**	**Rule-based**
Knoweledge Representation	**Accommodates rules** **Frames, objects, and others** **Modular**	**If-then rules** **Chain of rules**
Control	**Mostly explicit**	**Mostly implicit**
Domain Characteristics	**Ill-structured** **Multiple sources of knowledge** **Multiple reasoning mechanisms** **Nonprogrammed**	**Structured** **Single source** **Single line of reasoning (typically rule-based)** **Programmed**
Decision tasks	**Planning** **Diagnosis** **Interpretation**	**Alerting** **Monitoring** **Reminder**
Knowledge Acquisition	**Modular, Multiple**	**Data intensive** **Integrated**

Also, iLog, a decision support software, enables the sharing of clinical best practices to gather and analyze new or yet untapped data (Mcgee, 2006). IBM and Sloan-Kettering Cancer Center are working with systems to extract lab-result details from text intensive pathology reports to discover correlations by analyzing patterns, for example, of how specific kinds of tumors respond to various therapies (Mcgee, 2006). At West Mead Hospital, Sydney, Australia, Brain Resource in collaboration with IBM is building DSS software that can help analyze data from millions of sources of brain data (Mcgee, 2006). Mayo Clinic and IBM are applying pattern recognition and data-mining techniques to the electronic records of millions of patients to customize medical treatments to individual patients, such as selecting the best chemotherapy for a cancer patient with a particular genetic marker (Mcgee, 2004). In surgery, New York-Presbyterian Hospital is testing Penelope, a robot arm, that uses voice and visual recognition techniques to provide nursing assistance in the operating room (Santora, 2005). A system to analyze lab, pharmacy, patient, and other data, using inference engine software (infection control system), is being used at New York-Presbyterian Hospital. The purpose of this is to study data patterns to alert infectious disease nurses when a situation within the hospital could lead to an outbreak (Havenstein, 2005; Mcgee, 2005). Additionally, AI software is used to create treatment plans for patients in the cardiac intensive care units (using case-based approaches). Furthermore, the Cleveland Clinic and IBM are developing a translational medical platform and infrastructure to apply data-mining techniques to support clinical decision making. In another example, Kaiser Permanente's comprehensive medical database was used (via analysis and pattern recognition) to demonstrate that the arthritis drug Vioxx could cause heart problems (Eweek, 2006). In terms of clinical function applications, Aurora St. Luke's Medical Center, Milwaukee, uses the eICU system, which monitors patients electronically, helping a team of doctors and nurses to keep constant watch on more than 10 ICUs in four different hospitals across eastern Wisconsin (Fischman, 2005). A prototype wristwatch that monitors recovering cardiac patients' vital signs and sends them to their doctors via the cell phone network is being scheduled to be launched in 2006. The system called MDKeeper is being developed by Tadiran Spectralink of Israel (New Scientist, 2005). In Canada, the Global Public Health Intelligence Network (GPHIN) functions as a secure Internet-based early warning system assessing and cross referencing patients to track the length of time spent in a hospital or in quarantine. This type of information assists in providing a big-picture view of the operations of pandemics over large regions (Songini, 2006). Kranhold (2005) reports about a physician-alert system that alerts physicians if they are prescribing a medication that may react with another medication that the patient is taking—the system tracks patients' allergies and warns physicians if a medication might interact badly with the allergy. At University of Pittsburgh Medical Center (UPMC) in western Pennsylvania, a clinical information system delivers treatment information to doctors and nurses at the bedside, providing fast access to lab tests and X-rays, so they do not have to be repeated. The data also can be analyzed for possible outbreaks (Landro, 2005a). Also, another computer-order entry system at the UPMC Children's Hospital automatically calculates appropriate drug doses, based on a child's age and weight, and flags possible allergies or drug interactions (Landro, 2005b). Another application allows doctors to type in a patient's symptoms, and, in response, the system outputs a list of possible causes is reported in Leonhardt (2006). At St. Joseph's Hospital Health Center, Syracuse, NY, a picture archiving and communications system enables physicians and lab technicians to search and access patient information, data, and X-Rays via a Web portal (Mearian, 2005). Children's Memorial Hospital of Chicago is using a new system that confirms that the right medication is being administered at the right time to the right patient by scanning bar codes on the medication bags and sending it to a central computer that performs the matching (Pettis, 2006). The New York City is using a system for early warning of disease patterns called the Syndronic Surveillance System. It uses statistical analysis to calculate risk probabilities for possible disease outbreaks (Perez-Pena, 2003).

CONCLUSIONS

Distinguishing the systemic differences between clinical reasoning and the clinical function can enhance the design of CDSSs by assisting in the selection of the appropriate problem-solving architecture. The automation of clinical reasoning required a holistic approach that incorporates the properties of softness, openness, complexity, flexibility, and generality of medical systems. The trend toward distributed systems, multiagent models, parallel open-system architectures, and general models of intelligence for complex problem solving indicates that these alternatives must be explored in the design of automated clinical reasoning systems. Decision support applications development in medicine is in a primitive stage compared to other domains, but the use of system constructs in design can accelerate its maturing process.

REFERENCES

Bates, D. W., Kuperman, G. J., Wang, S., Gandhi, T., Kittler, A., Volk, L., Spurr, C., Khorasani, R., Tanasijevic, M., & Middleton, B. (2003). Ten commandments for effective clinical decision support: Making the practice of evidence-based medicine a reality. *Journal of the American Medical Informatics Association, 10*(6), 523-530.

Catley, C., Petriu, D. C., & Frize, M. (2004). Software performance engineering of a web service-based clinical decision support infrastructure. ACM SIGSOFT Software Engineering Notes. In *Proceedings of the 4th international workshop on Software and performance WOSP '04, 29 (1)*, pp. 130-138.

Churchman, C. W. (1971). *The design of inquiring systems.* New York: Basic Books.

Classen, D. C. (1998). Clinical decision support systems to improve clinical practice and quality of care. *Journal of the American Medical Association, 280*(15), 1360-1361.

Delaney, B. C., Fitzmaurice, D. A., Riaz, A., & Hobbs, F. D. R. (1999). Can computerized decision support systems delivery improved quality in primary care? *British Medical Journal, 319,* 1-3.

Durieux, P., Nizard, R., Ravaud, P., Mounier, N., & Lepage, E. (2000). A clinical decision support system for prevention of venous thromboembolisim: Effect on physician behavior. *Journal of the American Medical Association, 283*(21), 2816-2821.

Eweek. (2006, March 13). IT for better health [Editorial]. 32.

Fischman, J. (2005, August 1). Can high tech save your life? *U.S. News & World Report*, 45-52.

Garg, A. X., Adhikari, N. K., McDonald, H., Rosas-Arellano, M. P., Devereaux, P. J., Beyene, J., Sam, J., & Haynes, R. B. (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: A systematic review. *Journal of the American Medical Association*, *293*(10), 1223-1238.

Havenstein, H. (2005, August 8). Health care system turns to IT for patient care plans. *Computerworld,* 10.

Hewitt, C. E. (1985). The challenge of open systems. *Byte, 10* (4), 223-242.

Hunt, D. L., Haynes, R. B., Hanna, S. E., & Smith, K. (1998). Effects of computer-based clinical decision support systems on physician performance. *Journal of the American Medical Association, 280*(15), 1339-1346.

Johnston, M. E., Langton, K. B., Haynes, R. B., & Mathieu, A. (1994). Effects of computer-based clinical decision support systems on clinician performance and patient outcome: A critical appraisal of research. *Annals of Internal Medicine, 120*(2), 135-142.

Kaplan, B. (2001). Evaluating informatics applications—clinical decision support systems literature review. *International Journal of Medical Informatics, 64,* 15-37.

Kaushal, R., Shojania, K. G., & Bates, D. W. (2003). Effects of computerized physician order entry and clinical decision support systems on medication safety: A systematic review. *Archives of Internal Medicine, 163*(12), 1409-1416.

Kawamoto, K., Houlihan, C. A., Balas, E. A., & Lobach, D. F. (2004). Improving clinical practice using clinical decision support systems: A systematic review of trials to identify features critical to success. *British Medical Journal, 330,* 765.

Kranhold, K. (2005, June 20). Costly conversions. *The Wall Street Journal,* p. R6.

Landro, L. (2005a, January 17). The high-tech cure. *The Wall Street Journal,* p. R4.

Landro, L. (2005b, June 1). Drug errors show need for tech aid. *The Wall Street Journal,* p. D4.

Landro, L. (2006, January 25). Infant monitors yield new clues. *The Wall Street Journal,* p. D5.

Leonhardt, D. (2006, February 22). Why doctors so often get it wrong. *The New York Times,* p. .

Mcgee, M. K. (2004, August 9). Mayo builds toward customized medicine. *InformationWeek,* 24.

Mcgee, M. K. (2005, January 31). CIOs on call. *InformationWeek,* 28-36.

Mcgee, M. K. (2006, February 13). A pill, a scalpel, a database. *InformationWeek* 39-45.

Mearian, L. (2005, September 26). Hospital set to roll out digitial records system. *Computerworld,* 20.

Montgomery, A. A., Hahey, T., Peters, T. J., MacIntosh, C., & Sharp, D. J. (2000). Evaluation of computer based clinical decision support systems and risk chart for management of hypertension in primary care: Randomized controlled trial. *British Medical Journal, 320,* 686-690.

New Scientist. (2005, July 23). Gizmo. 23.

Perez-Pena, R. (2003, April 4). System in New York for early warning of disease patterns. *The New York Times,* p. 1.

Pettis, A. (2006, April 10). Patient care goes wireless. *Eweek*, N1.

Ramnarayan, P., & Britto, J.(2002) Paediatric clinical decision support systems. *Archives of Disease in Childhood, 87,* 361-362. (www.archdischild.com).

Santora, M. (2005, January 18). On the cutting edge (for surgery, an automated helping hand). *The New York Times,* p. B1.

Sim, I., Gorman, P., Greens, R. A., Haynes, R. B., Kaplan, B., Lehmann, H., & Tang. (2001). Clinical decision support systems for the practice of evidence-based medicine. *Journal of the American Medical Informatics Association, 8*(6), 527-534.

Songini, M. L. (2006, January 30). Catching the flu. *Computerworld,* 32.

Teich, J. M., Osheroff, J. A., Pifer, E. A., Sittig, D. F., & Jenders, R. A. (2005). Clinical decision support in electronic prescribing: recommendations and an action plan. *Journal of the American Medical Informatics Association, 12*(4), 365-376.

Van Gigch, J. P. (1978). *Applied general systems theory*. New York: Harper & Row.

Van Gigch, J. P. (1991). *System design modeling and meta modeling*. New York: Plenum Press.

Wyatt, J., & Spiegelhalter, D. (1991). Field trials of medical decision-aids: Potential problems and solutions. In P. D. Clayton (Ed.), *Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care*(pp. 3-7).. Washington, DC: American Medical Informatics Association.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 2, Issue 1, edited by J. Tan, pp. 44-53, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 2.16
Nonparametric Decision Support Systems in Medical Diagnosis:
Modeling Pulmonary Embolism

Steven Walczak
University of Colorado at Denver and Health Sciences Center, USA

Bradley B. Brimhall
University of Colorado School of Medicine, USA

Jerry B. Lefkowitz
University of Colorado at Denver and Health Sciences Center, USA

ABSTRACT

Patients face a multitude of diseases, trauma, and related medical problems that are difficult to diagnose and have large treatment and diagnostic direct costs, including pulmonary embolism (PE), which has mortality rates as high as 10%. Advanced decision-making tools, such as nonparametric neural networks (NN), may improve diagnostic capabilities for these problematic medical conditions. The research develops a backpropagation trained neural network diagnostic model to predict the occurrence of PE. Laboratory database values for 292 patients who were determined to be at risk for PE, with almost 15% suffering a confirmed PE, were collected and used to evaluate various NN models' performances. Results indicate that using NN diagnostic models enables the leveraging of knowledge gained from standard clinical laboratory tests, specifically the d-dimer assay and reactive glucose, significantly improving overall positive predictive value, compared to using either test in isolation, and also increasing negative predictive performance.

INTRODUCTION

Medical and surgical patients today face a variety of conditions that are both difficult and costly to diagnose and to treat. With ever skyrocketing medical costs (Benko, 2004), the use of information technology is seen as a much-needed means to help control and potentially to reduce medical direct costs (Intille, 2004). Deep vein thrombosis (DVT) and pulmonary embolism (PE) are medical conditions that are particularly difficult to diagnose in the acute setting (Mountain, 2003). Frequent usage of costly clinical laboratory tests to screen patients for further treatment is commonplace. All too commonly, hospitals provide treatment to patients without PE as a preventative measure (Mountain, 2003). Furthermore, patient mortality, morbidity, and both direct and indirect costs for delayed diagnosis of these conditions also may be substantial. Recent studies show that 40% to 80% of patients that die from PE are undiagnosed as having a potential PE (Mesquita et al., 1999; Morpurgo, Schmid, & Mandelli, 1998).

DVT may occur as the result of patient genetic and environmental factors or as a side effect of lower extremity immobility (e.g., following surgery). When a blood clot in the veins of a lower extremity breaks away, it may travel to the lungs and lodge in the pulmonary arterial circulation causing PE. If the clot is large enough, it may wedge itself into the large pulmonary arteries, leading to an acute medical emergency with a significant mortality rate. Approximately 2 million people annually experience DVT, with approximately 600,000 developing PE and approximately 10% of those PE episodes resulting in mortality (Labarere et al., 2004, Mesquita et al., 1999). Documented occurrence of DVT in postoperative surgical populations ranges from 10% (Hardwick & Colwell, 2004) to 28% (Blattler, Martinez, & Blattler, 2004).

Direct costs associated with DVT and PE come from the expensive diagnostic and even more expensive treatment protocols. It may be possible to lower these direct costs, especially when additional testing or treatment may be ruled out due to available knowledge. Nonparametric neural network (NN) systems enable the economic examination (Walczak, 2001) and nonlinear combination of various readily available clinical laboratory tests. Laboratory tests typically performed on surgical patients (e.g., blood chemistry) form the foundation for analysis and diagnostic model development.

One such test is the D-dimer assay that measures patient plasma for the concentration of one molecular product released from blood clots. When blood vessels are injured or when the movement of blood is too slow through veins of lower extremities, blood may begin to clot by initiating a series of steps in which fibrin molecules are cross-linked by thrombin to form a structure that entraps platelets and other coagulation molecules —a blood clot. As healing begins to occur, plasmin begins to break down the clot and releases, among other things, D-dimer molecules. D-dimers are actually small fragments of cross-linked fibrin and provide the basis for assessing blood-clotting activity. Patients with DVT and PE frequently have elevated levels of D-dimer in their plasma. Consequently, many hospitals now employ the D-dimer assay as a first test in the diagnostic pathway for these conditions. Usage of the results of a D-dimer assay effectively reduces the direct costs of DVT by reducing the requirement for downstream testing and treatment, specifically Doppler ultrasound tests (Wells et al., 2003). The use of nonparametric models enables the analysis of laboratory tests without regard to the population distribution, which may be a problematic factor when combining more than one laboratory test for the diagnostic prediction. Additionally, NNs provide nonlinear modeling capabilities that may be beneficial in combining pathology tests.

The research reported in this article will examine the efficacy of using NN models to predict the likelihood of a PE in surgical patient populations. A corollary research question is whether less invasive and less costly diagnostic methods are both available and reliable in predicting PE. The benefit of the reported research is twofold. First, the reported research examines new combinations of laboratory tests in order to determine if a combined model may be more reliable than currently used single variable medical models. Second, the research evaluates the viability of utilizing an NN model to predict PE (and potentially DVT). The NN provides improved positive predictive performance of the combined laboratory test model over the more traditional stepwise logistic regression models that currently are employed in medical modeling (León, 1994; Tran, VanOnselen, Wensink, & Cuesta, 1994; Walczak & Scharf, 2000). The cost effectiveness of utilizing the described neural network pathology tool is determined by examining the direct costs to patients, where direct costs represent the costs of diagnostic workups and the costs of any goods, services, and other resources utilized in any subsequent intervention (Patwardhan et al., 2005; Wildner, 2003). Implementing the described NN-based PE predictive model is capable of reducing patient evaluation direct costs by well over $1,200 per patient suspected of having a potentially life-threatening PE as well as reducing medical risks associated with some of the more advanced diagnostic methods.

The organization of the remainder of this article follows. The next section gives background on some of the issues associated with diagnosing and treating PE and DVT as well as examining the utilization of NN systems in medicine and issues surrounding the design of NN systems. The methodology section describes the patient population and development of the neural network models for predicting PE occurrence. The next section performs an analysis of the NN system's performance and consequent savings in direct costs. The last section concludes with a summary of the research findings and directions for future research.

BACKGROUND AND LITERATURE REVIEW

Diagnostic Difficulties of PE

Various tests exist for attempting to detect the presence of DVT or PE and commonly are administered to patients viewed as being at risk, based on medical history as well as patient signs and symptoms. For example, leg pain or swelling may indicate a patient with DVT; however, such symptoms are far from specific. Venography (x-ray with injected intravenous radioactive contrast dye) of the legs, Doppler ultrasound scans, plethysmography (measuring differential blood pressure between arms and legs), and the D-dimer assay blood test are some of the tests that may be employed in the evaluation of DVT (see http://www.nlm.nih.gov/medlineplus/ ency/article/000156.htm). Venography is the gold standard for diagnosing DVT, which commonly is treated with anticoagulant therapy such as intravenous heparin, requiring hospitalization, or injected low molecular weight heparin combined with oral warfarin medication, which may be continued for six months.

Patients with a sudden onset of chest pain or tachycardia or rapid shallow breathing may have PE, but these are also general signs of a wide variety of medical conditions (Mesquita et al., 1999; Mobley et al., 2005). PEs often are evaluated first using ventilation perfusion scanning (cost is $300 to $500 and involves the injection of radioactive tagged material) and chest x-rays that may be followed by pulmonary angiography (see http://www.nlm.nih.gov/medlineplus/ency/article/000132.htm). If the diagnosis of PE is made, the patient requires emergency treatment, typically thrombolytic therapy in order to dissolve the clot, and hospitalization.

While the pulmonary angiogram is considered the gold standard for determining PE presence (Evander et al., 2003; Mountain, 2003), it has a direct cost between $800 and $1,200 and involves injecting radio contrast die through a catheter that has been threaded into major coronary blood vessels with associated morbidity and mortality risks. If use of the D-dimer assay can determine better the likelihood or absence of PE, then a significant reduction in risks to the patient and costs ensues. The D-dimer assay has a direct cost of less than $7. Another factor that encourages the use of the D-dimer and other blood tests is the timeliness of the test, which averages approximately 30 minutes total time, including acquisition of the sample, transportation, analysis, and delivery of results. Improved detection of PE and early management will decrease the mortality and morbidity associated with PE (Stein, Kayali, et al., 2004).

Currently, the D-dimer assay has strong negative predictive reliability that excludes DVT or PE patients who are exceedingly unlikely to benefit from further invasive testing (e.g., angiography) or very few false negatives. Unfortunately, the test does not appear to have reliable positive predictability (Stein, Hull, et al., 2004), leaving the diagnostic role of the D-dimer uncertain. A potential cause of the lack of positive predictability is that, due to the potential life-threatening outcome of PE, the screening threshold for the D-dimer result is set very low (e.g., 400 or 450) in order to avoid false negative tests. As a result, many patients who do not have PE must undergo a pulmonary angiogram or other expensive testing procedures to determine if PE exists. Additionally, no standard for the D-dimer result threshold has been established, allowing for a wide range of cutoff values between hospitals (Stein, Hull, et al., 2004) and subsequent variation, though usually high, of patient population percentages undergoing unnecessary treatment.

Neural Networks in Medicine

The issues related to the use of D-dimer assay results—lack of standardized cutoff value, lack of positive predictive capability, and other influencing factors—pose a modeling problem. NNs provide strong modeling capabilities when population dynamics of the independent variables are unknown or nonlinear (Smith, 1993; Walczak & Cerpa, 2002), as would happen when using multiple laboratory test results to perform a diagnosis.

Before examining the factors that affect NN development (in the next section), a brief review of NN applications in medical domains is provided in this section. Neural networks or artificial neural networks applied to solve or provide decision support for a variety of medical domain problems started nearly two decades ago (Leese, 1986).

Many physicians are reluctant to utilize artificial intelligence technologies, including NN, especially neural networks that attempt to perform diagnosis or treatment in other than a decision support role (Baxt, 1995; Baxt & Skora, 1996; Hassoun, Wang, & Spitzer, 1994; Hu, Chau, Sheng, & Tam, 1999; Kleinmuntz, 1992; Walczak, 2003). The reluctance of physicians to adopt neural networks has led to the usage of neural networks as primarily image processing and laboratory test analysis tools. Table 1 provides a brief overview of historic and more recent neural network applications in medicine. A full review of all neural network applications in medicine is beyond the scope of this article,[1] and as such, Table 1 only provides a representative sample of research over the past decade.

Although representative and not comprehensive, Table 1 indicates that NN applications in medical domains tend to be used either in imaging or in laboratory settings (Walczak & Scharf, 2000), with diagnostic neural networks occurring rarely. Of interest from Table 1 for the research presented in this article are the three PE-related NN applications. Each of these NN research applications (Evander et al., 2003; Fisher, Scott, & Palmer, 1996; Serpen, Iyer, Elsamaloty, & Parsai, 2003) examines the improvement in identification or validation of PE from chest x-ray images or pulmonary angiogram data. Thus, each of these existing methods still requires the usage of and

Table 1. Types of neural network applications in medicine

Classification	NN applications research
Imaging (detect indicators in images)	Breast disease (Khuwaja & Abu-Rezq 2004, Papadopoulos et al. 2005), Coronary artery disease (Baxt 1991, Baxt & Skora 1996, Dorffner & Porenta 1994, Lapuerta et al. 1995, Scott et al. 2004), Electromyography (Hassoun 1994), Lung disease (Lin et al. 2005, Suzuki et al. 2004), Pulmonary embolism (Evander et al. 2003, Fisher et al. 1996, Serpen et al. 2003), Tomography (Bruyndonckx et al. 2004, Tourassi & Floyd 1995)
Laboratory (produce test results)	Breast disease (Mattfeldt et al. 2004), EEG (Güler et al. 2005b, Nowack et al., 2002, Walczak & Nowack 2001), General blood test pathology (Liparini et al. 2005) Head injury (Erol et al. 2005), Heart disease (Andrisevic et al. 2005, Haraldsson et al. 2004, Mobley et al. 2005), Hematology (Zini 2005), Lung disease (Güler et al. 2005a, Folland et al. 2004),
Resource Planning	Blood transfusions (Covin et al. 2003, Pereira 2004, Ruggeri et al. 2000, Walczak 2005, Walczak & Scharf 2000), Pharmacology (Gaweda et al. 2005, Polak et al. 2004)
Diagnostic	Breast cancer (Übeyli 2005)
Other	Medical cost estimation (Crawford et al. 2005, Goss & Vozikis 2002, Polak et al. 2004), Medical data mining (Petrovsky & Brusic 2004) Patient Outcomes Assessment (Buchman et al. 1994, Frye et al. 1996, Izenberg et al. 1997, Orunesu et al. 2004, Walczak et al. 2003, Yeong et al. 2005)

incurrence of the direct costs associated with these more costly diagnostic methods.

While NNs previously performed image analysis in order to confirm a diagnosis of PE, the use of NNs to predict a PE from standard low direct-cost blood tests is a novel application of this nonparametric modeling paradigm. The next section examines factors impacting the development of NN models, with an emphasis on the medical domain application of predicting PE.

Neural Network Design Issues

Patient treatment information necessarily must be of high quality (Silverman, 1998), since the consequences of prediction error can lead to increased morbidity and even mortality. NNs have been shown to outperform various statistical modeling methods in medical domains with respect to the accuracy of results (Baxt & Skora, 1996; Buchman, Kubos, Seidler, & Siegforth, 1994; Dybowski & Gant, 1995; Lapuerta, Azen, & LaBree, 1995; León, 1994; Stair & Howell, 1995; Walczak, 2005; Walczak & Scharf, 2000; White, 1992). NN design can be problematic and should account for the following influencing design factors (Walczak & Cerpa, 1999, 2002): training algorithm selection, input variable selection, and hidden layer architecture.

Various NN training algorithms exist; however, Alpsan, Towsey, Ozdamar, Ah, & Ghista (1995) have claimed that the backpropagation training algorithm produces classification and prediction models that generalize (on out of sample data) as well as or better than most other NN training algorithms, at least for their specific medical domain problem. Backpropagation-trained NN, in particular have been proven to be robust models of arbitrary complex equations (Hornik, 1991; Hornik, Stinchcombe, & White, 1989; White, 1990). Backpropagation-trained NNs are used commonly to provide solutions to business and engineering problems and also are the most common NN type for medical domain applications, which facilitates cross-research comparison of methodologies (Alpsan et al., 1995; Baxt, 1995; Cherkassky & Lari-Najafi, 1992; Dorffner & Porenta, 1994; Dybowski & Gant, 1995; Montague & Morris, 1994; Rodvold, McLeod, Brandt, Snow, & Murphy, 2001).

A reason for selecting another training algorithm besides backpropagation, such as the radial basis function, is to overcome problems with very noisy input data in the training sets (Barnard & Wessels, 1992; Carpenter, Grossberg, & Reynolds, 1995). Since the proposed NN prediction model utilizes the results of laboratory tests, many that already are controlled by computers and robotics, the resultant error rate in the input data is minimal and should not require a training algorithm that may not generalize well in order to alleviate error-prone data problems. Generally considered the best and most robust training method for NN classification and prediction models, backpropagation works well when low noise input data exist (Walczak & Cerpa, 2002). The backpropagation algorithm has been discussed widely in literature, and readers should see Fu (1994), Hertz, Krogh, and Palmer (1991), Jain, Mao, and Mohiuddin (1996), and Rodvold et al. (2001) for further details on the function of the backpropagation NN training methodology.

A critical step in the design of any NN model and especially in medical applications is the selection of high-quality predictive input (independent) variables that are not highly correlated with each other (Pakath & Zaveri, 1995; Smith, 1993; Tahai, Walczak, & Rigsby, 1998; Walczak & Cerpa, 1999). Domain expert physicians identify all available data that are related to the PE diagnosis problem.

The final influencing factor from the previous in developing NN models is the determination of the hidden node architecture (Smith, 1993; Walczak & Cerpa, 1999). The quantity of hidden nodes has a direct effect on generalization performance (Hung, Hu, Shanker, & Patuwo, 1996) with both too few and too many hidden nodes decreasing performance. The number of hidden layers also affects the smoothness and closeness of fit of the solution surface (Barnard & Wessels, 1992). A common heuristic method to determine the optimal number of hidden nodes is to start with a small quantity of hidden nodes (commonly one-fourth to one-half of the quantity of input nodes) and increment the hidden node quantity by one, until generalization performance begins to deteriorate. This method will safeguard against underfitting and overfitting the model (Walczak & Cerpa, 1999).

The research reported next will evaluate the efficacy of utilizing an NN methodology for developing a diagnostic model of PE that utilizes tests with a lower direct cost.

METHODOLOGY: A NEURAL NETWORK DIAGNOSTIC MODEL FOR PE BASED ON THE D-DIMER

As stated previously, the evaluation mechanisms for determining if a patient has a PE and requires treatment currently have high direct costs (e.g., ventilation perfusion scan that costs from $300 to $500, and pulmonary angiogram that costs from $800 to $1,200). In addition to other demographic and medical risk factors, the purpose of this research is to determine if less costly evaluation mechanisms for PE are available, especially using the D-dimer assay results, which have a direct cost of less than $7 and is used commonly to negatively screen out patients for PE (Stein, Hull, et al., 2004).

Research Population Description

The patient population in this research was 292 surgical patients admitted to a large Midwest teaching hospital during a nine-and-a-half-month period. All patients during the period of the study diagnosed by physicians as being at risk either for DVT or for PE were included in the study with no exclusions. Database records maintained by the hospital clinical laboratory as well as clinical information provided from radiology and chart records provide the data used in the NN input vector (independent variables). Patient records used in this study only come from patients clinically identified as being at risk for a DVT or possible PE by attending physicians based on signs (e.g., swelling in legs or irregular breathing) and symptoms (e.g., report of pain in legs for DVT or chest for PE). Just over 15% of this patient population, 44 out of 292, suffered a confirmed PE, which is consistent with other findings on the rate of DVT occurrence in surgical populations and consequent occurrence of PE (Hardwick & Colwell, 2004; Blattler et al., 2004). A pulmonary angiogram confirmed or negated the presence of PE.

The at-risk population of surgical patients for DVT and a possible PE has some basic statistical measures that are displayed in Table 2. From Table 2, it can be seen that the D-dimer results alone are not significantly different between patients with PE and patients without PE, although the PE patients have a much larger average D-dimer result and also a wider distribution. High D-dimer values may result from a variety of causes in addition to DVT or PE, which further confuses the application of the D-dimer for diagnosing PE.

NN Design of a PE Diagnosis Model

The focus of this study is on the use of the D-dimer assay results in possible combination with other values to improve the overall positive predictive performance to reduce false positives, which leads to unnecessary PE evaluation direct costs. As discussed in the "Neural Network Design Issues" section, the small error associated with laboratory tests and the reported robustness of the backpropagation training algorithm implied selection of backpropagation as the training algorithm for implementing the NN-based PE diagnosis model. Radial basis function (RBF) neural networks and stepwise logistic regression (LR) models also were implemented simultaneously in order to produce multiple models for comparison and to evaluate if a backpropagation-trained neural network would produce the optimal model. Using the model selection perspective (Swanson & White, 1995), comparison of each of the independent model types yielded the backpropagation NN model as the optimal model. The backpropagation models, with regard to overall accuracy (the sum of true positives and true negatives divided by all cases) outperformed both of the other model types ($p <$ .01 against the RBF-trained NN and $p < 0.10$ for the LR models).

Domain expert physicians identified blood test results with minimal direct costs that would be available for suspected DVT and PE patients prior to the performance of an angiogram. Since the research question of trying to determine if one or more of these test results could improve the PE diagnosis accuracy, a correlation matrix of all blood chemistry and other blood test results is performed to determine those variables that appear to be correlated with a positive diagnosis of PE. The variables include D-dimer, reactive glucose (Glu-R), blood urea nitrogen, serum creatinine, serum sodium, serum potassium, serum chloride,

Table 2. Typical d-dimer surgical population values

	D-dimer minimum	D-dimer maximum	D-dimer μ (average)	D-Dimer σ (std dev)	Age μ Age σ	Gender
Patients w/ confirmed PE	145	35450	4218.7	5862.75	48.45 22.38	47.5 % male
Patients w/o PE	124	22375	1782.63	1264.16	53 16.64	38.34 % male

serum carbon dioxide level, serum calcium, white blood cell count, red blood cell count, hemoglobin, hematocrit, mean corpuscular volume, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration, red cell distribution width, and platelet count. Other values, including mean platelet volume, absolute neutrophil count (and percent), absolute lymphocyte count (and percent), absolute monocyte count (and percent), absolute eosinophil count (and percent), and absolute basophile count (and percent), were considered for use but eliminated due to missing data for many of the patients (i.e., the tests had not been ordered by their attending physician). The predictive correlation of different blood test results measured by a Pearson's correlation matrix indicated that the D-dimer had the highest correlation with a validated PE for those test results included in the matrix and that randomized glucose (Glu-R) and CO_2 also both had significant ($p < 0.05$) predictive correlations with the PE-dependent variable. The age and sex of the patient also are included as independent variables to determine if these may be predictive when combined with blood test results.[2] Different backpropagation-trained NN models use various combinations of these five variables: age, sex, D-dimer, Glu-R, and CO_2 as the input vector of independent variables, producing 23 different sets of input variables and corresponding diagnosis models.

NN models for each of the 23 different input vectors start with a single hidden processing node. Both one- and two-hidden-layer architectures are evaluated, since the complexity of the diagnosis solution surface is unknown (Fu, 1994). When a second hidden layer is implemented, it also is started with a single hidden processing node. The quantity of processing nodes in each layer is incremented independently of the other layer and by a quantity of one. Reinitialization of each new larger model produces random interconnection weights before training occurs. Hidden node expansion stops, once the generalization performance begins to deteriorate and the model with the best generalization performance is kept.

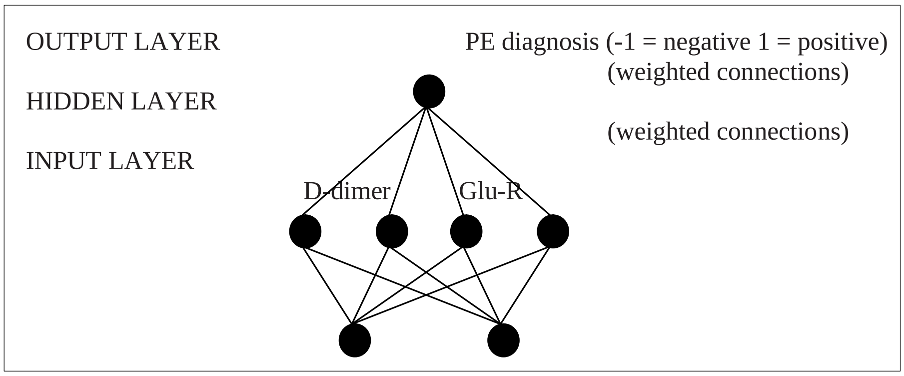
Training of each NN model uses one sample of data, and then validation uses a separate holdout sample. Due to the relatively small size of the population data, we use an eightfold cross validation technique in order to increase the quantity of holdout samples analyzed in the results to the full population of 292. The majority of the NNs evaluated performed similarly to just using the D-dimer result alone, meaning they had very good specificity (identifying positive PE cases) but very poor sensitivity (identifying negative PE cases). The finding that the sex-independent variable did not influence the performance of the various neural supports the findings of Stein et al. (2003b), who found that sex was not a factor in the diagnosis of PE.

The best performing NN model utilized both the D-Dimer and the GLU-R result for each patient. The name of the NN PE diagnosis model is PEDimeNet, since it utilizes the D-dimer laboratory test in combination with another test in order to maximize positive PE-predictive performance. The hidden node architecture for this backpropagation-trained NN is two input nodes representing the D-dimer and Glu-R results, four hidden nodes, and one output node, as shown in Figure 1. This NN is the one discussed in the next section.

RESULTS AND DISCUSSION

Backpropagation-trained NNs provide a real number output between -1 and 1 (though some NN shell tools allow this range to be extended) and not a true or false result for predicting the occurrence of a PE in a patient. These real-number output values require interpretation in order to determine the optimal cutoff point for partitioning the solution space into the two desired groups: probable PE and unlikely PE. An ROC (Receiver Operating Characteristic) curve analysis is performed commonly by those involved in medical diagnostics in order to determine the optimal threshold cutoff value for maximizing positive predictive performance

Figure 1. PEDimeNet NN architecture



on the generalization sample (Kamierczak, 1999; Obuschowsk, Lieber, & Wians, 2004), and this technique is used with the NN training sample output values in order to determine the optimal real number threshold value for performing the classifications. A similar and independent ROC curve analysis for the RBF and LR models reported in the previous section determines each model's cutoff value. The area under the ROC curve for the training data was 0.79. Once the ROC threshold value is determined, it is then applied universally to all out-of-sample output values produced by the NN in order to determine the NN's PE diagnosis for the corresponding patient.

Table 3 displays the prediction results of the resulting NN model. Similar to increasing the D-dimer cutoff value to include more of the patient population in the group that undergoes further evaluation via an angiogram or other PE verification method, a tradeoff exists between the positive predictive accuracy and the negative predictive accuracy using the NN model. Lowering the PE cutoff point causes the positive predictive accuracy to increase and creates a corresponding decrease in the negative predictive accuracy. Due to the possible outcome of mortality for PE patients, a positive predictive performance of 90% sensitivity is desired, as specified by the domain expert physicians (where the TP rate in Table 3 is the sensitivity, and the TN rate is the specificity). The overall diagnosis accuracy ([TP+TN]/Total) performance of the PEDimeNet NN model is 62.33% with almost 91% of all PE patients identified as well as nearly 57% of patients that do not have a PE and, therefore, do not require any additional testing for PE with subsequent reduction in direct costs.

Recall that traditional usage of the D-dimer with a relatively low threshold value produces a false negative (FN) rate very close to 0%. As a result, a negative PE result usually means that the patient truly will not have a PE. The false positive rate for the standard D-dimer test is very high, which leads to increased medical costs through unnecessary costly procedures for follow-up evaluation (e.g., pulmonary angiogram).

Table 3. NN diagnostic prediction results for PE

	NN positive PE	NN negative PE
Actual PE (positive)	TP = 40 (90.91%)	FN = 4 (9.09%)
No PE (negative)	FP = 107 (43.15%)	TN = 141 (56.85%)

(TP = True Positive, FN = False Negative, FP = False Positive, and TN = True Negative)

For the test population used in the research reported in this article, a simulation of adjusting the D-dimer cutoff value to produce a false negative rate identical to the proposed NN model produced a false positive rate of 91.53% (227 patients that require unnecessary evaluation tests), meaning that only 21 of 248 patients did not have further costly PE evaluation procedures recommended. Whereas, the PEDimeNet NN was able to save additional laboratory testing for 141 patients, as opposed to just 21 when using the traditional D-dimer only diagnosis. Therefore, an NN model of PE that utilizes both D-dimer and Glu-R test results provides a high-quality positive predictor. The NN PE prediction model exceeds using a non-NN D-dimer-based evaluation mechanism with respect to true negatives (patients without PE) by well over 600%, resulting in significant cost savings for the patients. Direct cost savings would be approximately $96,000 to $144,000 for the nine-and-a-half-month period of the research study, which translates to a $121,000 to $182,000 direct-cost annualized savings.

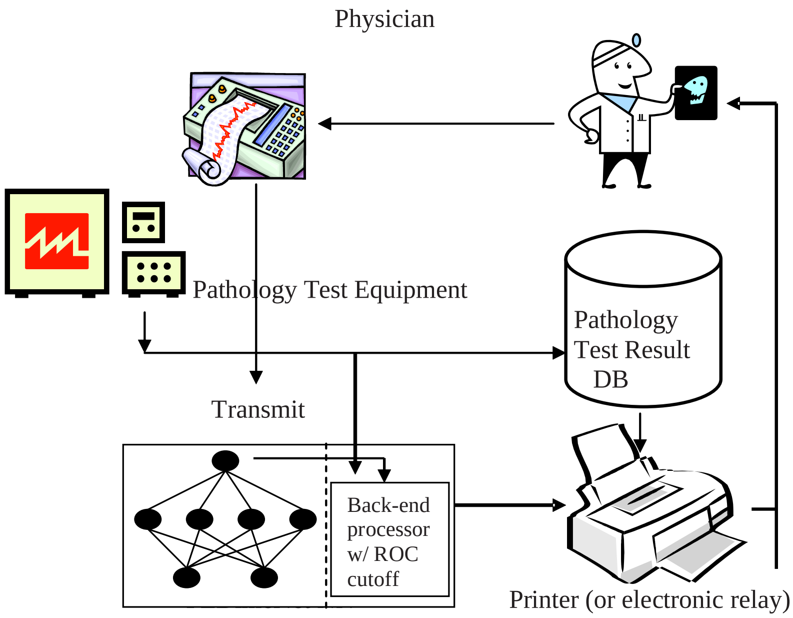
A limitation of the reported research is that the NN diagnosis model only models the population at a single hospital. While this hospital is one of the larger providers in a major metropolitan area and, as such, has a demographically diverse population, differences in the correspondence between the D-dimer and Glu-R values with an actual PE may exist in different geographic populations (Stein, Kayali, et al., 2004). As such, the methodology described would require each hospital to independently train their own NN model and to establish their own ROC cutoff threshold value for interpreting the generalization (out-of-sample) diagnosis values of the NN.

Utilizing the proposed NN system would require only very minor workflow changes in the emergency department and clinical laboratory of the studied hospital and would integrate similarly at most hospitals that follow similar diagnostic methodologies. Medical user acceptance of new decision support technology is highly dependent on two factors: (1) that the decision support technology not be perceived as attempting to make diagnosis (Baxt, 1995; Baxt & Skora, 1996; Walczak & Scharf, 2000) and (2) that the new information technologies not disrupt normal workflows (Kirkley & Stein, 2004; Walczak, 2003). Physicians receive laboratory results at most hospitals either electronically or on paper after encoding and storing the results in a pathology database. The PEDimeNet NN-based decision support system would be positioned architecturally in the pathology departments' IT framework, as displayed in Figure 2. This integration of the PEDimeNet NN diagnostic tool then would appear seamless to physicians, integrating with their existing workflow simply by adding the analysis to the pathology report already received by the physician.

The PEDimeNet PE diagnosis decision support tool is composed of the PEDimeNET NN, an interface that allows the entry of D-dimer and Glu-R laboratory test results (which may be handled electronically through an EDI interface or performed manually at a keyboard), a backend program that contains the ROC curve analysis-based decision cutoff value (which also may be read from a database to allow for movement in the cutoff value over time, perhaps due to new pathology equipment sensitivity) and produces the go/no-go diagnostic recommendation for probable PE in a patient. Future enhancements would include a database or agent-oriented Web browser to recommend additional tests, if necessary, and treatment protocols.

The following scenario presents an illustration of the functioning of the proposed PEDimeNet system within a hospital setting. A physician would request a diagnostic workup, including a patient blood draw. The clinical laboratory receives the patient's blood sample. Laboratory tests are completed using the existing laboratory equipment. The results are delivered to the physician either electronically (with EDI requiring a small hardware modification to interface with

Figure 2. PEDimeNet relative to other pathology IT systems



the existing pathology instrumentation) or on paper (after entering the results into the laboratory database with a simultaneous transmission to the PEDimeNet). A software interface to the database or other peripheral devices (in C# or Java) encodes the received value into the format required by the NN. The PEDimeNet diagnostic decision support system then would interpret the NN's results, based on a cutoff value established by applying the ROC curve analysis determined threshold. The backend analysis of the NN output provides a report of the interpreted NN prediction in either electronic or printed format for the attending physician, who then would determine if any further evaluation was required. Other than the step in the lab necessary to transmit the results to the PEDimeNet decision support system, the automated process is similar to standard decision-making processes for evaluating the need to screen for a PE. From the physician's perspective, there is no change in workflow from the traditional receipt of the original noninterpreted D-dimer assay results.

CONCLUSION

The research presented in this article demonstrates the efficacy of utilizing NNs for predicting the occurrence of PE in patients presenting to the emergency department and those at risk of PE following surgery. The reported results were able to match the very low false negative rate of a more traditional test that relied on the D-dimer alone. Although the usage of a D-dimer screening test is questioned in cases of PE (Stein, Hull, et al., 2004), it has proved reliable for evaluating the presence of DVT (Wells et al., 2003), a leading cause of PE. Furthermore, over a 600% increase in true negative predictions with a corresponding decrease in associated medical risks to patients and decreased medical costs of hundreds of thousands of dollars per hospital per year demonstrates the benefits of using a nonlinear and nonparametric modeling technique such as the backpropagation train NN.

Another benefit derived from this research is the identification of the superior positive predictive performance for PE when using the combination of the D-dimer result value in combination with the Glu-R result value. The D-dimer individually was not able to distinguish reliably between positive PE cases and negative ones, creating a high corresponding false positive assessment rate so that only a very small number of actual PEs would be misdiagnosed. The utilization of an NN as the statistical analysis component of the PEDimeNet NN decision support system enabled the rapid combination of input values in nonlinear ways without regard to population distribution characteristics. A potential drawback of using NNs as a model determination mechanism in order to evaluate the contribution of various independent variables is the need to determine the ideal architecture. While the current research utilized a brute force incremental process to analyze both single- and two-hidden-layer networks, new neural network shell tools are able to automate some or all of the NN node architecture optimization process.

While the research reported in this article focused on the prediction of PE, the relationship between PE and DVT and previous research has indicated a strong positive correlation between the D-dimer assay results and accurate DVT diagnosis (Wells et al., 2003), which indicates that a similar NN model would be able to accurately predict DVT. This would permit the operation of two side-by-side (or possibly one) NNs in the PEDimeNet decision support system in order to evaluate both PE and DVT simultaneously, which is a topic of future research.

In addition to the specific research benefits just described, generalizing the research method and results provides implications for both researchers and practitioners. Future research efforts will benefit by realizing that nonparametric modeling techniques in general and neural networks specifically are a viable method for analyzing complex and potentially ill-defined (unknown population and error distributions) problems. Researchers should embrace the most advanced modeling techniques available, especially since neural networks repeatedly have demonstrated increased performance over more traditional parametric statistical methods. The data set for the presented NN PE diagnosis application is relatively error-free. Researchers should note that designing optimal neural network models has many implementation factors, and careful attention in selecting input variables, selecting the training algorithm, and determining the hidden-layer and hidden-node architecture will help to produce the optimal NN model (Smith, 1993; Walczak & Cerpa, 1999, 2002).

An implication for practice is to challenge existing protocols and heuristics. The addition of a low direct-cost variable to the PE diagnosis model significantly increases the accuracy and reduces direct costs. Other medical diagnostic and assessment problems may derive benefit from examining new sets of variables. The NN modeling method provides an inexpensive and reliable way to rapidly assess the impact of adding new variables or removing existing decision variables (Walczak, 2001). Previous research (Walczak et al., 2003) demonstrates this by showing that a commonly used variable for determining acute pancreatitis patient outcomes, RANSON score, detracts from the predictive performance of other readily available variables.

The research reported in this article demonstrates the impact of NN diagnosis models through the reduction of direct costs. Future research is needed to realize the overall benefit gained from NN diagnosis models by examining the impact on indirect costs, such as reduced patient healing times and overall quality of care received.

REFERENCES

Alpsan, D., Towsey, M., Ozdamar, O., Ah, C. T., & Ghista, D. N. (1995). Efficacy of modified backpropagation and optimization methods on a real-world medical problem. *Neural Networks, 8*(6), 945-962.

Andrisevic, N., Ejaz, K., Rios-Gutierrez, F., Alba-Flores, R., Nordehn, G., & Burns, S. (2005). Detection of heart murmurs using wavelet analysis and artificial neural networks. *Journal of Biomechanical Engineering, 127*(6), 899-904.

Barnard, E., & Wessels, L. (1992). Extrapolation and interpolation in neural network classifiers. *IEEE Control Systems, 12*(5), 50-53.

Baxt, W. G. (1991). Use of an artificial neural network for the diagnosis of myocardial infarction. *Annals of Internal Medicine, 115*(11), 843-848.

Baxt, W. G. (1995). Application of artificial neural networks to clinical medicine. *The Lancet*, *346*, 1135-1138.

Baxt, W. G., & Skora, J. (1996). Prospective validation of artificial neural network trained to identify acute myocardial infarction. *The Lancet, 347*, 12-15.

Benko, L. B. (2004). Going retro. *Modern Healthcare, 34*(33), 14-15.

Blattler, W., Martinez, I., & Blattler, I. K. (2004). Diagnosis of deep venous thrombosis and alternative diseases in symptomatic outpatients. *European Journal of Internal Medicine, 15*(5), 305-311.

Bruyndonckx, P., Léonard, S., Tavernier, S., Lemaître, C., Devroede, O., Wu, Y., et al. (2004). Neural network-based position estimators for PET detectors using monolithic LSO blocks. *IEEE Transactions on Nuclear Science, 51*(5), 2520-2525.

Buchman, T. G., Kubos, K. L., Seidler, A. J., & Siegforth, M. J. (1994). A comparison of statistical and connectionist models for the prediction of chronicity in a surgical intensive care unit. *Critical Care Medicine, 22*(5), 750-762.

Carpenter, G. A., Grossberg, S., & Reynolds, J. H. (1995). A fuzzy ARTMAP nonparametric probability estimator for nonstationary pattern recognition problems. *IEEE Transactions on Neural Networks, 6*(6), 1330-1336.

Cherkassky, V., & Lari-Najafi, H. (1992). Data representation for diagnostic neural networks. *IEEE Expert, 7*(5), 1992, 43-53.

Covin, R., O'Brien, M., Grunwald, G., Brimhall, B., Sethi, G., Walczak, S., et al. (2003). Factors impacting transfusion of fresh frozen plasma, platelets, and red blood cells during elective coronary artery bypass graft surgery. *Archives of Pathology & Laboratory Medicine, 127*(4), 415-423.

Crawford, A. G., Fuhr, J. P., Jr., Clarke, J., & Hubbs, B. (2005). Comparative effectiveness of total population versus disease-specific neural network models in predicting medical costs. *Disease Management, 8*(5), 277-287.

Dorffner, G., & Porenta, G. (1994). On using feedforward neural networks for clinical diagnostic tasks. *Artificial Intelligence in Medicine, 6*(5), 417-435.

Dybowski, R., & Gant, V. (1995). Artificial neural networks in pathology and medical laboratories. *The Lancet, 346*, 1203-1207.

Erol, F. S., Uysal, H., Ergün, U., Barişçi, N., Serhathoğlu, & Hardalaç, F. (2005). Prediction of minor head injured patients using logistic regression and MLP neural network. *Journal of Medical Systems, 29*(3), 205-215.

Evander, E., Holst, H., Järund, A., Ohlsson, M., Wollmer, P., Åström, K., et al. (2003). Role of ventilation scintigraphy in diagnosis of acute pulmonary embolism: An evaluation using artificial neural networks. *European Journal of Nuclear Medicine and Molecular Imaging, 30*(7), 961-965.

Fisher, R. E., Scott, J. A., & Palmer, E. L. (1996, March). Neural networks in ventilation-perfusion imaging. *Radiology, 198*, 699-706.

Folland, R., Hines, E., Dutta, R., Boilot, P., & Morgan, D. (2004). Comparison of neural network predictors in the classification of tracheal-bronchial breath sounds by respiratory auscultation. *Artificial Intelligence in Medicine, 31*(3), 211-220.

Frye, K. E., Izenberg, S. D., Williams, M. D., & Luterman, A. (1996). Simulated biologic intelligence used to predict length of stay and survival of burns. *Journal of Burn Care and Rehabilitation, 17*(6), 540-546.

Fu, L. (1994). *Neural networks in computer intelligence*. New York: McGraw-Hill.

Gaweda, A. E., Muezzinoglu, M. K., Aronoff, G. R., Jacobs, A. A., Zurada, J. M., & Brier, M. E. (2005). Individualization of pharmacological anemia management using reinforcement learning. *Neural Networks, 18*(5/6), 826-834.

Goss, E. P., & Vozikis, G. S. (2002). Improving health care organizational management through neural network learning. *Health Care Management Science*, *5*(3), 221-227.

Güler, I., Polat, H., & Ergün, U. (2005a). Combining neural network and genetic algorithm for prediction of lung sounds. *Journal of Medical Systems, 29*(3), 217-231.

Güler, N. F., Übeyli, E. D., & Güler, I. (2005b). Recurrent neural networks employing Lyapunov exponents for EEG signals classification. *Expert Systems with Applications, 29*(3), 506-514.

Haraldsson, H., Edenbrandt, L., Ohlsson, M. (2004). Detecting acute myocardial infarction in the 12-lead ECG using Hermite expansions and neural networks. *Artificial Intelligence in Medicine, 32*(2), 127-136.

Hardwick, M. E., & Colwell, C. W., Jr. (2004). Advances in DVT prophylaxis and management in major orthopaedic surgery. *Surgical Technologist, 12*, 265-268.

Hassoun, M. H., Wang, C., & Spitzer, A. R. (1994). NNERVE: neural network extractions of repetitive vectors for electromyography — Part II: Performance analysis. *IEEE Transactions on Biomedical Engineering, 41*(11), 1053-1061.

Hertz, J., Krogh, A., & Palmer, R. (1991). *Introduction to the theory of neural computation*. Reading, MA: Addison Wesley.

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks, 4*, 251-257.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks, 2*(5) 359-366.

Hu, P. J., Chau, P. Y. K., Sheng, O. R. L., & Tam, K. Y. (1999). Examining the technology acceptance model using physician acceptance of telemedicine technology. *Journal of Management Information Systems, 16*(2), 91-112.

Hung, M. S., Hu, M. Y., Shanker, M. S., & Patuwo, B. E. (1996). Estimating posterior probabilities in classification problems with neural networks. *International Journal of Computational Intelligence and Organizations, 1*(1), 49-60.

Intille, S. S. (2004). A new research challenge: Persuasive technology to motivate healthy aging. *Plant Physiology*, *136*(1), 235-237.

Izenberg, S. D., Williams, M. D., & Luterman, A. (1997). Prediction of trauma mortality using a neural network. *American Surgeon, 63*(3), 275-281.

Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. *IEEE Computer, 29*(3), 31-44.

Kamierczak, S. C. (1999). Statistical techniques for evaluating the diagnostic utility of laboratory tests. *Clinical Chemistry and Laboratory Medicine, 37*, 1001-1009.

Khuwaja, G. A., & Abu-Rezq, A. N. (2004). Bimodal breast cancer classification system. *Pattern Analysis & Applications, 7*(3), 235-242.

Kirkley, D., & Stein, M. (2004). Nurses and clinical technology: Sources of resistance and strategies for acceptance. *Nursing Economic$, 22*(4), 195, 216-222.

Kleinmuntz, B. (1992). Computers as clinicians: An update. *Computers in Biology and Medicine, 22*(4), 227-237.

Labarere, J., Bosson, J. L., Brion, J. P., Fabre, M., Imbert, B., Carpentier, P., & et al. (2004). Validation of a clinical guideline on prevention of venous thromboembolism in medical inpatients: A before-and-after study with systematic ultrasound examination. *Journal of Internal Medicine, 256*(4), 338-348.

Lapuerta, P., Azen, S. P., & LaBree, L. (1995). Use of neural networks in predicting the risk of coronary artery disease. *Computers in Biomedical Research, 28*(1), 38-52.

Leese, S. (1986). Computer advances towards the realization of a cybernated health-science diagnostic system. *American Journal of Psychotherapy, 40*(3), 321-323.

León, M. A. (1994). Binary response forecasting: Comparison between neural networks and logistic regression analysis. In *Proceedings of the World Congress on Neural Networks, Vol. 2*, San Diego, CA (pp. 244-247).

Lin, D-W., Yan, C-R., & Chen, W-T. (2005). Autonomous detection of pulmonary nodules on CT images with a neural network-based fuzzy system. *Computerized Medical Imaging & Graphics, 29*(6), 447-458.

Liparini, A., Carvalho, S., & Belchior, J. C. (2005). Analysis of the applicability of artificial neural networks for studying blood plasma: Determination of magnesium ion concentration as a case study. *Clinical Chemistry & Laboratory Medicine, 43*(9), 939-946.

Mattfeldt, T., Kestler, H. A., & Sinn, H-P. (2004). Prediction of the axillary lymph node status in mammary cancer on the basis of clinicopathological data and flow cytometry. *Medical & Biological Engineering & Computing, 42*(6), 733-739.

Mesquita, C. T., Morandi Júnior, J. L. B., Perrone, F. T., Oliveira, C. S., Barreira, L. J., Nascimento, S. S. C. A., et al. (1999). Fatal pulmonary embolism in hospitalized patients—Clinical diagnosis versus pathological confirmation. *Arquivos Brasileiros de Cardiologia, 73*(3), 255-258.

Mobley, B. A., Schechter, E., Moore, W. E., McKee, P. A., & Eichner, J. E. (2005). Neural network predictions of significant coronary artery stenosis in men. *Artificial Intelligence in Medicine, 34*(2), 151-161.

Montague, G., & Morris, J. (1994). Neural-network contributions in biotechnology. *Trends in Biotechnology, 12*(8), 312-324.

Morpurgo, M., Schmid, C., & Mandelli, V. (1998). Factors influencing the clinical diagnosis of pulmonary embolism: Analysis of 229 postmortem cases. *International Journal of Cardiology, 65*(Supplement 1), S79-S82.

Mountain, D. (2003). Diagnosing pulmonary embolism: A question of too much choice? *Emergency Medicine, 15*(3), 250-262.

Nowack, W. J., Walczak, S., & Janati, A. (2002). Clinical correlate of EEG rhythmicity. *Journal of Clinical Neurophysiology, 19*(1), 32-36.

Obuschowski, N. A., Lieber, M. L., & Wians, F. H. (2004). ROC curves in clinical chemistry: Uses, misuses, and possible solutions. *Clinical Chemistry, 50*, 1118-1125.

Orunesu, E., Bagnasco, M., Salmaso, C., Altrinetti, V., Bernasconi, D., Del Monte, P., et al. (2004). Use of an artificial neural network to predict Graves' disease outcome within 2 years of drug withdrawal. *European Journal of Clinical Investigation, 34*(3), 210-217.

Pakath, R., & Zaveri, J. S. (1995). Specifying critical inputs in a genetic algorithm-driven decision support system: An automated facility. *Decision Sciences, 26*(6), 749-779.

Papadopoulos, A. Fotiadis, D. I., & Likas, A. (2005). Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machines. *Artificial Intelligence in Medicine, 34*(2), 141-150.

Patwardhan, M. B., Matchar, D. B., Samsa, G. P., McCrory, D. C., Williams, R. G., & Li, T. T. (2005). Cost of multiple sclerosis by level of disability: A review of literature. *Multiple Sclerosis, 11*(2), 232-239.

Pereira, A. (2004). Performance of time-series methods in forecasting the demand for red blood cell transfusion. *Transfusion, 44*(5), 739-746.

Petrovsky, N., & Brusic, V. (2004). Virtual models of the HLA class I antigen processing pathway. *Methods, 34*(4), 429-435.

Polak, S., Skowron, A., Mendyk, A., & Brandys, J. (2004). Artificial neural network in pharmacoeconomics. *Studies in Health Technology and Information, 105*, 241-249.

Rodvold, D. M., McLeod, D. G., Brandt, J. M., Snow, P. B., & Murphy, G. P. (2001). Introduction to artificial neural networks for physicians: Taking the lid off the black box. *The Prostate, 46*(1), 39-44.

Ruggeri, A., Comai, G., Belloni, M., & Zanella, A. (2000). A simulation study for the design of a control system for the blood concentration process in autotransfusion. *Annals of Biomedical Engineering, 28*(4), 470-482.

Scott, J. A., Aziz, K., Yasuda, T., & Gewirtz, H. (2004). Integration of clinical and imaging data to predict the presence of coronary artery disease with the use of neural networks. *Coronary Artery Disease, 15*(7), 427-434.

Serpen, G., Iyer, R., Elsamaloty, H. M., & Parsai, E. I. (2003). Automated lung outline reconstruction in ventilation-perfusion scans using principal component analysis techniques. *Computers in Biology and Medicine, 33*(2), 119-142.

Silverman, B. G. (1998). The role of Web agents in medical knowledge management. *M.D. Computing, 15*(4), 221-231.

Smith, M. (1993). *Neural networks for statistical modeling*. New York: Van Nostrand Reinhold.

Stair, T. O., & Howell, J. M. (1995). How long will it take? How much will it cost? Multiple regression and neural network programs at ED triage. *American Journal of Emergency Medicine, 13*, 118-119.

Stein, P. D., Hull, R. D., Patel, K. C., Olson, R. E., Ghali, W. A., Alshab, A. K., et al. (2003a). Venous thromboembolic disease: Comparison of the diagnostic process in blacks and whites. *Archives of Internal Medicine, 163*(15), 1843-1848.

Stein, P. D., Hull, R. D., Patel, K. C., Olson, R. E., Ghali, W. A., Alshab, A. K., et al. (2003b). Venous thromboembolic disease: Comparison of the diagnostic process in men and women. *Archives of Internal Medicine, 163*(14), 1689-1694.

Stein, P. D., Hull, R. D., Patel, K. C., Olson, R. E., Ghali, W. A., Brant, R., et al. (2004). D-dimer for the exclusion of acute venous thrombosis and pulmonary embolism. *Annals of Internal Medicine, 140*(8), 589-602.

Stein, P. D., Kayali, F., & Olson, R. E. (2004). Estimated case fatality rate of pulmonary embolism, 1979 to 1998. *American Journal of Cardiology, 93*(9), 1197-1199.

Suzuki, K., Horiba, I., Sugie, N., & Nanki, M. (2004). Extraction of left ventricular contours from left ventriculograms by means of a neural edge detector. *IEEE Transactions on Medical Imaging, 23*(3), 330-339.

Swanson, N. R., & White, H. (1995). A model-selection approach to assessing the information in the term structure using linear models and artificial neural networks. *Journal of Business and Economic Statistics, 13*(3), 265-275.

Tahai, A., Walczak, S., & Rigsby, J. T. (1998). Improving artificial neural network performance through input variable selection. In P. Siegel, K. Omer, A. deKorvin, & A. Zebda (Eds.), *Applications of fuzzy sets and the theory of evidence to accounting II* (pp. 277-292). Stamford, CT: JAI Press.

Tourassi, G. D., & Floyd, C. E. (1995). Lesion size quantification in SPECT using an artificial neural network classification approach. *Computers in Biomedical Research, 28*(3), 257-270.

Tran, D. D., VanOnselen, E., Wensink, A., & Cuesta, M. A. (1994). Factors related to multiple organ system failure and mortality in a surgical intensive care unit. *Nephrology Dialysis Transplantation, 9*(4 Supp.), 172-178.

Übeyli, E. D. (2005). A mixture of experts network structure for breast cancer diagnosis. *Journal of Medical Systems, 29*(5), 569-579.

Walczak, S. (2001). Neural networks as a tool for developing and validating business heuristics. *Expert Systems with Applications, 21*(1), 31-36.

Walczak, S. (2003). A multiagent architecture for developing medical information retrieval agents. *Journal of Medical Systems, 27*(5), 479-498.

Walczak, S. (2005). Artificial neural network medical decision support tool: Predicting transfusion requirements of ER patients. *IEEE Transactions on Information Technology in Biomedicine, 9*(3), 468-474.

Walczak, S., & Cerpa, N. (1999). Heuristic principles for the design of artificial neural networks. *Information and Software Technology, 41*(2), 109-119.

Walczak, S., & Cerpa, N. (2002). Artificial neural networks. In R. A. Meyers (Ed.), *Encyclopedia of physical science and technology* (3rd ed.) (Vol. 1; pp. 631-645). San Diego, CA: Academic Press.

Walczak, S., & Nowack, W. J. (2001). An artificial neural network approach to diagnosing epilepsy using lateralized bursts of theta EEGs. *Journal of Medical Systems, 25*(1), 9-20.

Walczak, S., Pofahl, W. E., & Scorpio, R. J. (2003). A decision support tool for allocating hospital bed resources and determining required acuity of care. *Decision Support Systems, 34*(4), 445-456.

Walczak, S., & Scharf, J. E. (2000). Reducing surgical patient costs through use of an artificial neural network to predict transfusion requirements. *Decision Support Systems, 30*(2), 125-138.

Wells, P. S., Anderson, D. R., Rodger, M., Forgie, M., Kearon, C., Dreyer, J., et al. (2003). Evaluation of d-dimer in the diagnosis of suspected deep-vein thrombosis. *New England Journal of Medicine, 349*(13), 1227-1235.

White, H. (1990). Connectionist nonparametric regression: Multilayer feedforward networks can learn arbitrary mappings. *Neural Networks, 3*, 535-549.

White, H. (1992). Consequences and detection of misspecified nonlinear regression models. In H. White (Ed.), *Artificial neural networks: Approximations and learning theory* (pp. 224-258). Oxford, UK: Blackwell.

Wildner, M. (2003). Health economic issues of screening programmes. *European Journal of Pediatrics, 162*(Supp.), S5-S7.

Yeong, E-K., Hsiao, T-C., Chiang, H. K., & Lin, C-W. (2005). Prediction of burn healing time using artificial neural networks and reflectance spectrometer. *Burns, 31*(4), 415-420.

Zini, G. (2005). Artificial intelligence in hematology. *Hematology, 10*(5), 393-400.

ENDNOTES

1. A database inquiry on EBSCO using the terms *neural network* AND (medicine OR medical) produced 997 articles. A similar query on PubMed, the NIH article database, produced 6,120 articles.
2. The occurrence of PE between blacks and whites has been found to be comparable (Stein et al., 2003a) and, as such, is not included in the variables considered for the NN input vector. Additionally, this variable is subjective, being either self-reported by patients or reported as an observation by the admitting nurse in the studied hospital and, as such, may have a higher than acceptable error rate for use in a backpropagation-trained NN.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 1, Issue 2, edited by J. Tan, pp. 65-82, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 2.17
A Cross-Cultural Framework for Evaluation

Pekka Turunen
Shiftec, Finland

ABSTRACT

The evaluation of information systems (ISs), especially in the healthcare field, is a complex task. Evidently, there is a need for better understanding of the different aspects evaluation. While in the general IS science field there have been several attempts to build frameworks and models to better understand the evaluation of ISs, in healthcare similar frameworks have been almost nonexistent. Unfortunately, general frameworks cannot be exactly applied for the cross-cultural evaluation of healthcare ISs, because they do not recognise the specific nature of the medicine. Based on works in different areas, this chapter represents an attempt at to combine them to conceptual frameworks for the evaluation of healthcare ISs.

INTRODUCTION

The evaluation of health information systems (ISs) has proved to be an especially difficult task. Evaluation projects are often interdisciplinary by nature and designed by both information technology people and medical professionals (Heathfield, Pitty, & Hanka, 1998; Turunen & Talmon, 2000). Different parties have difficulties in understanding each other because, for example, a lack of common tools (Nykänen, 2000).

Among others, one problem is the lack of framework for conceptual understanding of IS impacts and their evaluation. Frameworks and models, used for that purpose, are mainly drawn from different research areas and, unfortunately, are inadequate for this specific field. The frameworks for the evaluation of information systems are not able to describe the specific nature of the area (e.g., golden standards). At the same time, the frameworks for evaluation in the healthcare field do not usually recognise the difficulties of measuring impacts of ISs, the need for external validity or they are too general for the evaluation of ISs. Yet, it has been suggested that a new evaluation paradigm is needed (Shaw, 2002).

This chapter presents well-known models from those different areas and combines them into an appropriate framework for the evaluation of health ISs. The framework is based directly on the previous works and is a logical extension of the historical development of such models and frameworks.

The purpose of the frameworks is to formulate a conceptual guide of evaluation and emphasise the connection (noncausal by nature) between the different impacts of healthcare ISs. Thus, the framework may be useful in identifying relationships among the success variables of ISs. It assists in taxonomising existing evaluation results of ISs and, therefore, in comparing different evaluations. Equally, the framework should aid in making an overall judgment based on different evaluation methods and results of the healthcare IS. Furthermore, the frameworks may give some hints for the measurement of different impacts.

The article presents frameworks at two levels. A general framework describes healthcare IS evaluation at a general level. In addition, a specific framework has been developed for diagnostic ISs. A procedure to minimise evaluation is also presented.

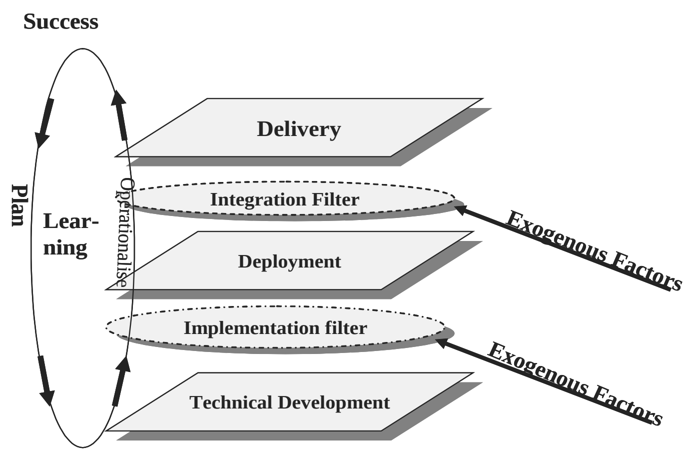
FRAMEWORKS FOR THE EVALUATION OF INFORMATION SYSTEMS

The best known and most widely used framework for the evaluation of ISs is probably DeLone and McLean's (1992) IS success model (Iivari & Ervasti, 1994; Jurison 1996). This model has brought about a good amount of fair and unfair criticism and, therefore, has encouraged further development in the area (Ballantine et al., 1996; Bonner, 1995; Grover, Jeong, & Segars, 1996; Kangas & Manwani, 1998; Manwani & Kangas, 1998; Pitt, & Watson, 1994; Seddon, 1997; Seddon, Staples, Patnayakuni, & Bowtell, 1998). One of the most important further developments is the 3-D model of IS's success (Balantine et al., 1996).

The 3-D model has been divided into three main elements: a) development, b) deployment, and c) delivery (see Figure 1). The development element includes such things as technology, system type, quality of data, IS-professional skills, and so forth. The second element contains variables such as user satisfaction, task impact, personal impact, and so on (e.g., alignment of individual business objectives, resources, and use of the output are included in the delivery element). In addition to these three elements researchers also refer to a *fourth element* of political, social, and economic impacts. The meaning of these aspects will be increased while implementing wide Internet-based ISs. However, private-sector focus of evaluation is still often at the organisational level.

Filters among elements try to illustrate that an impact at one level does not automatically cause

Figure 1. The simplified version of 3-D model for information systems success. (Source: Balantine et al, 1996)



an impact at the next level. One of the most important new features that Ballantine et al. (1996) have added to the original idea is the learning loop. Learning addresses that the evaluation is always a two-way process.

There are three main features common to almost all the models, guidelines, taxonomies, or frameworks, which try to grasp the principal idea of information flow and evaluation of that process as a whole. All significant aspects can be seen, for example, in the 3-D model (see Figure 1 and the previous text).

First, the division of ISs into three basic constructs or elements are typically a) technology, b) user, and c) organisation (Balantine et al., 1996; Cronk, 2000; DeLone & McLean, 1992; Enning et al., 1997; Friedman & Cornford, 1989; Hamilton & Chervany, 1981; Rowley, 1998; Salmela, 1997; Scott, 1994). In different studies slightly different terms, such as *individual* instead of *user*, are used for the same elements. Some of the studies and schools have concentrated in making more accurate analyses of the main elements (e.g., Chismar & Wiley-Patton, 2003; Venkatesh & Morris, 2000).

Second, the frameworks cover a wide range of different kinds of both subjective and objective evaluation methods (e.g., Ballantine et al., 1996; Cronk, 2000; DeLone & McLean 1992). Third, most frameworks are based on the premise that an impact on one level leads most likely, but not always, to an impact on the next level.

According to Nykänen's (2000) longitudinal research of several evaluation cases, ISs' evaluation models do not take into account healthcare environmental variables. Medicine, for example, tends to focus on patients, medical evaluation methods considered "golden standards" and attention is also given for physicians.

FRAMEWORKS IN THE FIELD OF HEALTHCARE

A Framework for Assessing Quality of Care

Donabedian's (1980) framework is aimed at assessing quality of care. Quality assessment is built on the concept of three elements: a structure followed by a process and an outcome (see Figure 2).

In Donabedian's (1980) framework, the *structure* can mean human, physical, and financial resources. For example, medical tools are a typical example of the structure. The *process* is a set of activities (e.g., treatment plans that go on within and between practitioners and patients). The *outcome* is the change in a patient's health status that can be attributed to given care. The health status is defined broadly to include, for example, patient satisfaction, a patient's knowledge of health-related issues, and health-related behavioural change. In such a chain, each element is, to some extent, a cause of the element that follows. The structure of the framework and the content of the three elements are very much the same as in frameworks that are used in the IS science (Hebert, 2001).

A Framework for Evaluation of Telehealth

The framework for telehealth evaluation has been based on DeLone and McLean's (1992) IS success

Figure 2. The functional relationship among three elements in Donabedian's (1980) framework for assessing the quality of care

model and Donabedian's (1980) framework for assessing quality of care (Hebert, 2001; see also Figure 3). The framework for telehealth evaluation is one rare attempt to combine different aspects of the cross-cultural research area.

In the framework, individual and organisational elements have been innovatively divided into structures and outcomes, a taxonomy that the other researchers have not used. According to Hebert (2001), Donabedian's (1980) structure is equal DeLone and McLean's (1992) first element, "information and system quality." Nevertheless, in the telehealth framework the structure has been connected with DeLone and McLean's "individual and organisational impact," which actually should mean Donabedian's outcome, not structure (Hebert, 2001). Probably as a result of these, we use only three basic elements (structure, process, and outcome) for categorising two case studies. However, the telehealth framework is a first proper framework to cover the whole picture of the evaluation of telehealth.

Models for Evaluation of Medical Technology

Fineberg, Bauan, and Sosman (1977) first presented the idea of modeling efficacy of clinical diagnostic technology (see Figure 4). Their model greatly influenced physicians' thinking (e.g., Chizeck & Katona, 1985; Cook, Brun-Buisson, Guyatt, & Sibbald, 1994; Maisey, 1996; Sailly et al., 1985; Szczepura & Stilwell, 1988; Ure, Spangenberger, Lefering, Dietrich, & Troidl, 1992) and it has been used for the evaluation of telehealth systems (Ohinmaa & Reponen, 1997). This model emphasises the clinical-care process, medical technology, and the significance of physicians (making diagnoses). Even though the clinical efficacy model is narrower than models in the IS science, the principal idea of the model is very similar to that of IS models, because they all present a Western scientific way of thinking.

Guyatt, Tugwell, Feeny, Drummond, and Haynes (1986) have developed the clinical ef-

Figure 3. A framework for evaluation of telehealth (Source: Hebert, 2001)

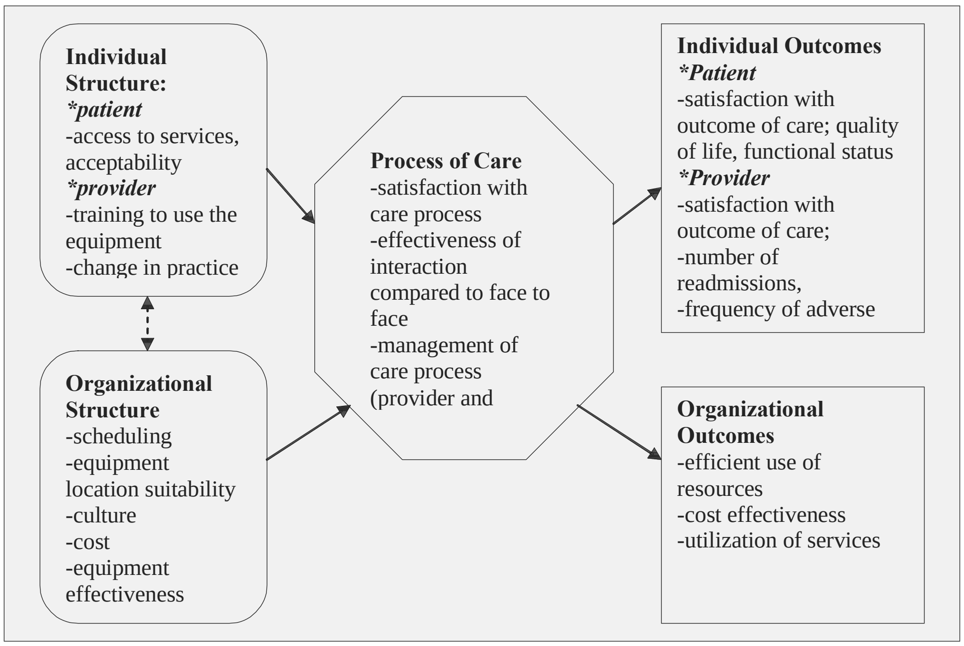
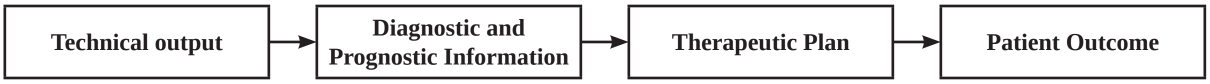


Figure 4. Clinical efficacy model (Source: Fineberg, et al., 1977)



ficacy model in two ways. They have extended the evaluation of patient outcome to include the economic evaluation. They have also added the therapeutic process as a new process parallel to the diagnostic process. Their framework also emphasises randomised controlled trial (RCT) in the evaluation of medical technology. The new features are welcome extensions in the direction of a framework that covers all the impacts and evaluation methods of medical technology.

Normally, stakeholders with a strict medical background approach the evaluation of health ISs through Fineberg et al.'s (1977) or Guyatt et al.'s (1986) pattern. The problem is that patterns can address only a minority of effects that medical technologies are causing and only a small minority that healthcare ISs are causing. Evaluation methods, such as the RCT, can hardly evaluate complex ISs in the real-life situation in which effects of the system may be indirect, intangible, or nonquantified (Davey, 1994; Feinstein, 1983; Heathfield et al., 1998; Hebert, 2001; Kaplan, 2001; Shaw, 2002).

Traditional medicine has also favoured validation of outcome. However, the development of ISs is usually also a main issue in the evaluation of ISs. Unfortunately, a summative evaluation does not incorporate a development perspective (Hamilton & Chervany, 1981).

A CONCLUSION OF PRIOR RESEARCH: FRAMEWORKS FOR EVALUATION OF HEALTHCARE INFORMATION SYSTEMS

The idea of evaluation frameworks is to combine the key elements of both research areas in such a way that frameworks are able to cover the healthcare information area and can possibly be accepted by the different stakeholders. Another significant issue is to present nonmedical evaluation methods to stakeholders with a medical background. It is important try to use a language that looks familiar to stakeholders (Nykänen, 2000). The frameworks may provide new measurements for the evaluation of healthcare ISs, when it is not possible or reasonable to use traditional clinical measurements.

Three basic elements familiar from the general IS area can be combined equally with elements of medical technology. The main combined elements are a) technical quality, b) user impact, and c) organisational and patient impact (see Figure 5). One has to take into consideration the role of societal impacts while increasing the amount of Internet applications and integration process in nationwide ISs. However, evaluating the impacts is still impossible or at least extremely difficult.

In Figure 5, the arrows showing effects of the IS and evaluation (they could also be called a learning loop) are shown above the figure. The use of a broken line is designed to emphasise the noncausal nature of ISs and the evaluation of their impact.

According to Nykänen (2000), IS people are not able to consider the environmental variables of healthcare. For them the specific subframework emphasises, and hopefully, crystallise, the core process of medical treatment and the importance of evaluating it (see Figure 6). The specific framework is especially designed for diagnostic ISs or clinical decision support systems and can also be seen as a model for physicians.

Figure 5. A general framework for the evaluation of impacts of health information systems

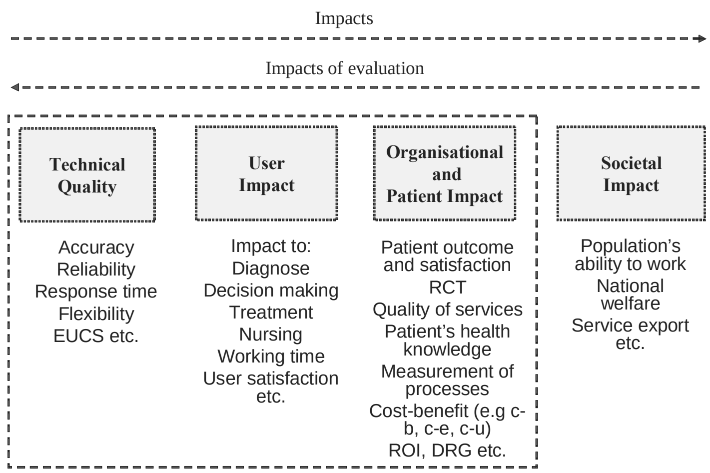
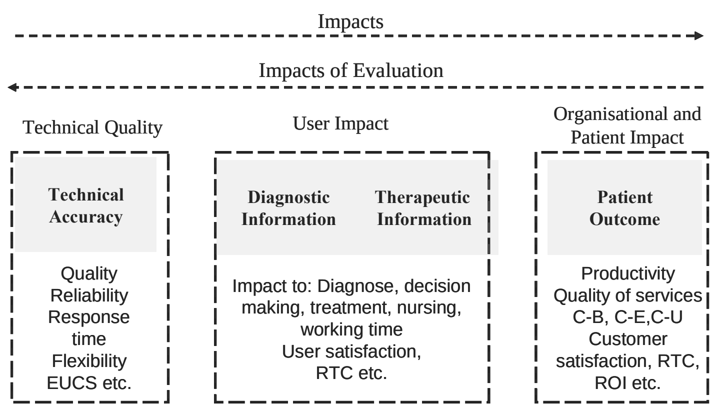


Figure 6. A framework for the evaluation of diagnostic information systems



The Use of Frameworks to Minimise Evaluation Set

Resources are rarely available to enable the launching of massive evaluation programs (Friedman & Wyatt, 1997). Following aspects may ease formulating evaluation plans in order to minimise the evaluation set:

1. What are the expected relationships between the different elements or variables (Fineberg et al., 1977; Hebert, 2001)?
2. How does use of new IS compare to the process of traditional care or previous IS (Hebert, 2001)?

This procedure may help to formulate a minimum amount of evaluation trials to prove an impact of a system or to find the potential cases that may affect the outcome (e.g., improved medical data to lead better decision making and better outcome).

The framework may help to classify evaluation results and thus, to compare evaluation studies. Because classification of evaluation studies seems to be in some sense a subjective task (see Figure 5), this framework is offered rather as a conceptual guide for classifying, not as a straightjacket.

REFERENCES

Ballantine, J., Bonner, M., Levy, M., Martin, A., Munro, I., & Powell, P. L. (1996). The 3-D model of information systems success: The search for the dependent variable continues. *Information Resources Management Journal, 9*(4).

Bonner, M. (1995) DeLone and McLean's model for judging information system success: A retrospective application in manufacturing. *Proceedings of the Second European Conference on IT Evaluation*, Henley Management College, UK.

Chismar, W. G., & Wiley-Patton, S. (2003). Does the technology acceptance model apply to physicians. *Proceedings of the 36th Hawaii International Conference on System Sciences*.

Chizeck, H. J., & Katona, P. G. (1985). Closed loop control. In Carson & Cramp (Eds.), *Computers and control in clinical medicine*. New York: Plenum Press.

Cook, D. H., Brun-Buisson, C., Guyatt, G. H., & Sibbald, W. J. (1994). Evaluation of new diagnostic technologies: Bronchoalveolar lavage and the diagnosis of ventilator-associated pneumonia. *Critical Care Medicine, 22*(8).

Cronk, M. C. (2000). Understanding complex information system constructs through holistic construal. *Proceedings of Seventh European Conference on the Evaluation of Information Technology*, Dublin, Ireland.

Davey, B. (1994). The nature of scientific research. In K. McConway (Ed.), *Studying health and disease*. Oxford, England: The Open University Press.

DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research, 1,* 60-95.

Donabedian, A. (1980). *Explorations in quality assessment and monitoring. Vol. I: The definition of quality and approaches to its assessment.* MI: Health Administration Press.

Enning, J., Fabianne, E., Talmon, J., Vissers, M., Nykänen, P., Roine, R., et al. (1997). Guidelines for evaluation of telematics applications in medicine, Deliverable D03.2b. Retreived December 1, 2001, from *http://www-vatam.unimaas.nl*

Feinstein, A. R. (1983). An additional basic science for clinical medicine: II. The limitations of randomized trials. *Annals of Internal Medicine, 99*, 544-550.

Fineberg, H. V., Bauan, R., & Sosman, M. (1977). Computerized cranial tomography: Effect on diagnostic and therapeutic plans. *Journal of the American Medical Association, 3,* 244-227.

Friedman, A. L., Cornford, D. S. (1989). *Computer systems development: History, organization and implementation.* Chichester: Wiley.

Friedman, C. P., & Wyatt, J. C. (1997). *Evaluation methods in medical informatics.* New York: Springer.

Grover, V., Jeong, S. R., & Segars, A. H. (1996). Information systems effectiveness: The construct space and patterns of application. *Information & Management, 4,* 117-191.

Guyatt, G. H., Tugwell, P. X., Feeny, D. H., Drummond, M. F., & Haynes, B. R. (1986). The role of before-after studies of therapeutic impact in the evaluation of diagnostic technologies. *Journal of Chronic Diseases, 39*(4), 295-304.

Hamilton, S., & Chervany, N. L. (1981). Evaluating information system effectiveness. Part I: Comparing evaluation approaches. *MIS Quarterly, 3,* 55-69.

Heathfield, H., Pitty, D., & Hanka, R. (1998). Evaluating information technology in healthcare: Barriers and challenge. *British Medical Journal, 316,* 1959-1961.

Hebert, M. (2001, September 2-5). Telehealth success: Evaluation framework development. *Proceedings of Medinfo 2001*, London.

Iivari, J., & Ervasti, I. (1994). User information satisfaction, IS implementability and effectiveness. *Information & Management, 27*(4), 205-220.

Jurison, J. (1996). The temporal nature of IS benefits: A longitudinal study. *Information & Management, 30,* 75-79.

Kangas, K., & Manwani, S. (1998). Package implementation within a European MNC: Case of a Finnish project serving as a pilot. *Proceedings of the Sixth European Conference on Information Systems,* Aix-en-Provence, France.

Kaplan, B. (2001). Evaluating informatics applications-some alternative approaches: Theory, social interactionism, and call for methodological pluralism. *International Journal of Medical Informatics, 64,* 39-56.

Maisey, M. (1996). Evaluating the benefits of nuclear cardiology. *Quarterly Journal of Nuclear Medicine, 40*(1), 47-54.

Manwani, S., & Kangas, K. (1998). IT project evaluation in multinational companies: A multiple stakeholder viewpoint of a Finnish project within a European MNC. *Proceedings of the 1998 Information Resources Management Association International Conference*, Boston, MA.

Nykänen, P. (2000). *Decision support systems in health informatics perspective. A-2000-10 Acta Electronica Universitatis Tamperensis* (Vol. 55). Tampere.

Ohinmaa, A., & Reponen, J. (1997). A model for the assessment of telemedicine and a plan for testing of the model within five specialities. *FinOHTAn Report, 5.*

Pitt, L. F., & Watson, R. T. (1994). Longitudinal measurement of service quality in information systems: A case study. *Proceedings of the 15th International Conference on Information Systems*, Vancouver, British Columbia, Canada.

Rowley, J. (1998). Towards a framework for information management. *International Journal of Information Management, 18*(5).

Sailly, J. C., Lebrun, T., Vincent, P., Eeckhoudt, L., Izard, D., Leclerc, H., et al. (1985) The medical and economic consequences of automation in bacteriology: A case study in a French university hospital. *Social Science & Medicine, 10,* 1163-1166.

Salmela, H. (1997). From information systems quality to sustainable business quality. *Information and Software Technology, 39*(12), 819-826.

Scott, J.E. (1994) The Measurement of Information Systems Effectiveness: Evaluating A Measuring Instrument. In *Proceedings of the 15th ICIS*. Vancouver (pp.111-128).

Seddon, P. B. (1997). A respection and extension of the DeLone and McLean model of IS success. *Information Systems Research*, *8*(3), 240-253.

Seddon, P. B., Staples, D. S., Patnayakuni, R., & Bowtell, M. (1998). The IS effectiveness matrix: The importance of stakeholder and system in measuring IS success. *Proceedings of the 19th International Conference on Information Systems*, Helsinki, Finland.

Shaw, N. T. (2002). CHEATS: A generic information communication technology (ICT) evaluation framework. *Computers in Biology and Medicine, 32,* 209-220

Szczepura, A. K., & Stilwell, J. A. (1988). Information for decision makers at hospital laboratory level: An example of a graphical method of representing costs and effects for a replacement automated technology in a haematology laboratory. *Social Science & Medicine, 7,* 715-725.

Turunen, P., & Talmon, J. (2000, September 28-29). Stakeholder groups in the evaluation of medical information systems. *Proceedings of Seventh European Conference on the Evaluation of Information Technology*, Dublin, Ireland.

Ure B. M., Spangenberger W., Lefering R., Dietrich A., & Troidl H. (1992). Routine gastroscopy before laparoscopic cholecystectomy: Evaluation of the technology in 376 patients. *Z Gastroenterol, 30*(8), 529-533.

Venkatesh, V., & Morris, M. G. (2000). Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behaviour. *MIS Quarterly, 24*(1).

Chapter 2.18
BACIIS:
Biological and Chemical Information Integration System

Zina Ben Miled
Indiana University Purdue University, USA

Nianhua Li
Indiana University Purdue University, USA

Omran Bukhres
Indiana University Purdue University, USA

ABSTRACT

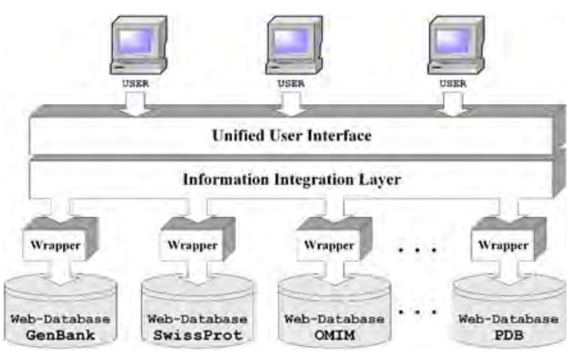
Life science Web databases are becoming increasingly essential in conducting everyday biological research. With more than 300 life science Web databases and the growing size of the life science data, searching and managing these complex data requires technology beyond that of traditional database systems. The open research issues include the interoperability of geographically distributed autonomous databases, which are generally only Web-accessible, and the seamless semantic-based integration of these databases with total transparency to the user. In this paper, the implementation of a Biological and Chemical Information Integration System (BACIIS) is presented. BACIIS supports the integration of multiple heterogeneous life science Web databases and allows the execution of global applications that extend beyond the boundaries of individual databases. This paper discusses the architecture of BACIIS. It also discusses the techniques used to extract and integrate data from the various life science Web databases.

INTRODUCTION

The confederation of geographically distributed life science databases is a critical challenge facing biological and biomedical research. Databases are the intermediaries between experimental observation and the ability to extract biological knowledge. The challenge in managing life science data is due to the large volume of data, the rate at which the data are increasing in size, and the heterogeneity and complexity of the data format. There are currently more than 300 life science Web databases (Baxevanis, 2002) providing access to scientific data and literature via the Web. These databases use different nomenclatures, file formats, and data access interfaces. Furthermore, they may include redundant and conflicting data.

This paper presents a system (BACIIS) that uses the mediator-wrapper approach to support the integration of life science Web databases. BACIIS constructs a tightly coupled federation of databases and uses ontology as a global schema. Currently, BACIIS (Figure 1) integrates seven life science Web databases: GenBank (Benson, 1998), SWISS-PROT (Bairoch, 1998), PIR (Baker, 1998), PROSITE (Sigrist, 2002), ENZYME (Bairoch, 2002), PDB (Berman, 2000), and OMIM (McKusick, 1998). GenBank is an annotated collection of all publicly available DNA sequences, which includes approximately 20 million DNA sequences. SWISS-PROT is an annotated, non-redundant protein sequence database that includes high-quality annotation of proteins, such as the description of the function of a protein, its domain structure, post-translational modifications, and variants. PIR is a protein sequence and structure database. ENZYME is a database that contains information relative to the nomenclature of enzymes. PROSITE is a database of protein families and domains. It consists of biologically significant sites, patterns, and profiles that help to identify to which known protein family a new sequence belongs. PDB is a database of 3-D biological macromolecular structure data. The OMIM database is a catalog of human genes and genetic disorders, which contains textual information, pictures, and reference information.

Figure 1. Information integration for life science Web database



This paper specifically shows how a mediator-wrapper approach can be successfully used to integrate Web databases in the context of the restrictions imposed by the life science domain. These restrictions include:

- The databases are autonomous and only Web accessible. Therefore, data extraction has to be performed following the data presentation protocol that is adopted by the component databases.
- Access to the databases can change often; different databases return their result in a different format, and there is a large number of life science databases. Therefore, any data extraction mechanism needs to take these aspects into consideration in such a way that it easily can adapt to changes in the databases, and databases can be added to the integration system easily.
- Resolving heterogeneity is an important aspect in the integration of life science databases and has to be addressed with minimal user intervention in order for the integration system to be useful to the scientists.

BACIIS is a tightly coupled federation of life science Web databases. Its architecture, shown in Figure 2, is a three-tier software architecture (Kossmann, 2000) that consists of a mediator, wrappers (Ambite, 1998; Levy, 1998; Wiederhold, 1992), and an ontology that is used as a global schema. The mediator transforms data from its format in the component database to the internal format used by the integration system (Ambite, 1998). A wrapper extracts information from the component database (Levy, 1998). Each component database is described by using domain ontology terms. This information is stored in a data source schema file, and each life science Web database is associated with a data source schema file. The ontology and the data source schema files for the component databases make up the domain knowledge base of BACIIS.

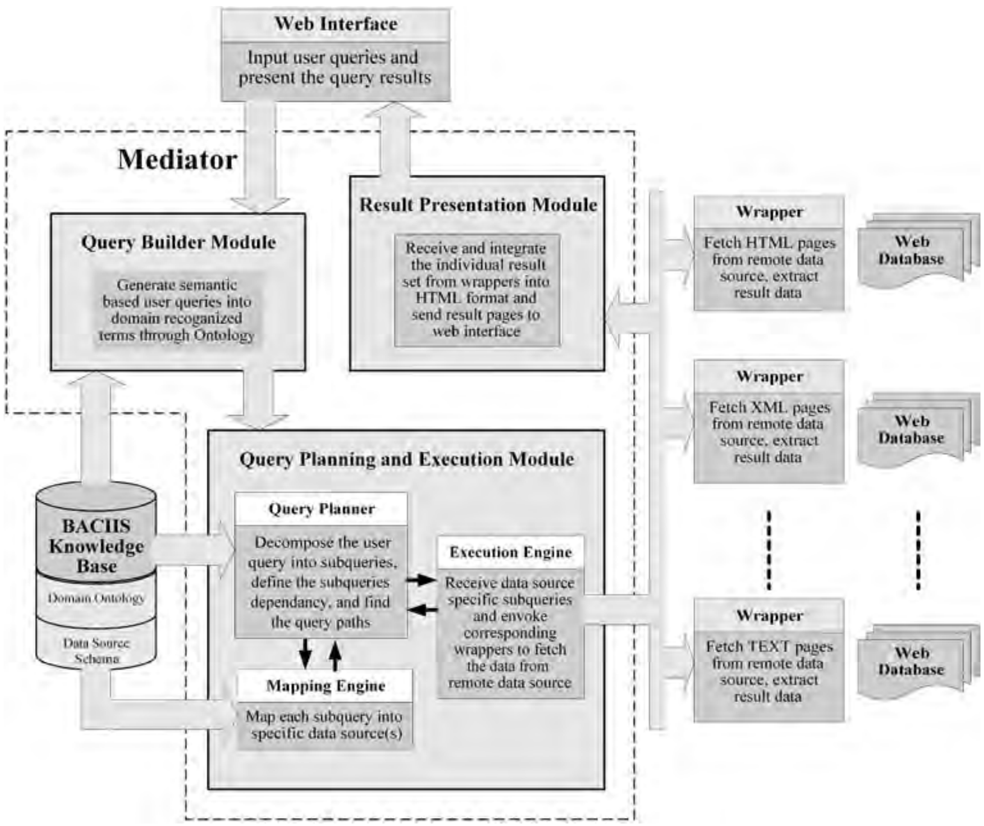
In the remainder of the paper, the approaches that we developed and the design decisions that we adopted to integrate life science Web databases, while respecting the above-mentioned restrictions, are discussed. The three tiers of the BACIIS software architecture are presented in the following sections. The last three sections of the paper present an example execution of a query in BACIIS, discuss related work, and summarize the major contributions of this paper.

MEDIATOR

The mediator in BACIIS identifies the relations between the query input and output data types by using the domain knowledge. Based on these relations, it generates a query execution plan, using the appropriate Web databases. After the wrapper fetches the results from the Web databases, the mediator needs to integrate the results returned from the Web databases according to the object relations defined in the ontology.

The mediator (Figure 2) includes three modules: (1) a query builder module, (2) a query-planning module, and (3) a result presentation module. The query builder module translates a multi-database query entered by the user into domain-recognized terms from the ontology. Queries translated by the query builder module are forwarded to the query-planning module.

Figure 2. BACIIS architecture



This module consists of the query planner, the mapping engine, and the execution engine. The query planner receives a multi-database query and decomposes it into subqueries. The mapping engine uses the data source schema to find the appropriate Web database, which can accept the set of input data types and generate the set of output data types for a given subquery. Using the list of Web databases established by the mapping engine, the query planner constructs an execution path for the entire multi-database query, according to the dependencies among the subqueries. The execution engine receives the query execution path from the query planner and invokes the wrapper associated with each Web database that corresponds to a given subquery.

This mediator behaves like a traditional mediator (Kossmann, 2000). The novelty introduced in BACIIS is the ability to identify the Web databases that can answer the subquery without user intervention by matching the query input and output types of the query to the input and output types recorded in the data source schema files of the component databases. This departs from previously proposed systems (Decker, 2001) that require the user to specify the query execution plan and the desired semantic mapping.

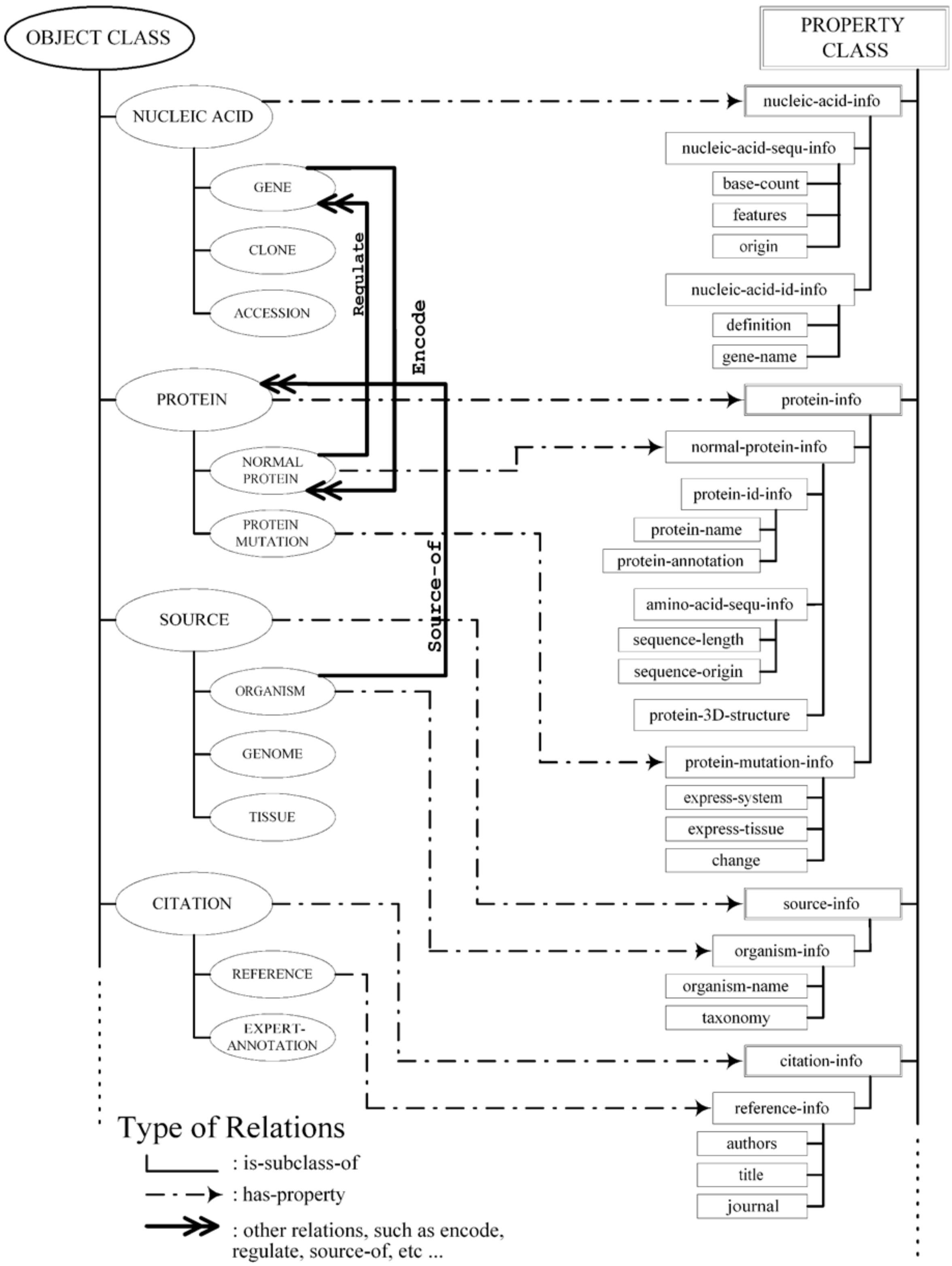
DOMAIN ONTOLOGY

As stated in Lakshmanan (1996), when integrating heterogeneous databases, three issues have to be addressed: semantic, syntactic, and system variability. Resolving semantic variability consists of adequately interpreting and cross relating information stored in different databases (Lakshmanan, 1996). In general, syntactic variability arises from the heterogeneity of the database schema, data models, and query processing approaches. System heterogeneity refers to the use of different operating systems and communication protocols (Lakshmanan, 1996). In BACIIS, the databases are accessed via their Web interfaces. Therefore, system heterogeneity is not critical. However, both semantic and syntactic variability are critical and have to be addressed before the integration system is of any use to the scientists.

Domain ontologies (Bouguettaya, 1999; Gruber, 1993; Ouzzani, 2000) have been used successfully as domain models in several general-purpose information integration systems, such as the Information Manifold (Levy, 1998), Observer (Mena, 1998), Ontobroker (Decker, 1999), and SIMS (Arene, 1996). The domain ontology in BACIIS serves as a global schema and plays a major role in resolving both semantic and syntactic variability. It is used for the reconciliation of synonyms across the various databases. For example, to refer to the literature reference, the database GenBank uses “reference,” and the databases PDB (Protein Data Bank) and PIR (Protein Information Resource) use “citation” and “bibliography,” respectively. The ontology is used to map all the different terms used by the remote databases to a single term. In addition, the ontology in BACIIS provides the domain knowledge necessary for resolving syntactic heterogeneities. For example, the reference information in the database SWISS-PROT is a single attribute. In the database GenBank, reference information consists of multiple attributes including author, title, journal, and so forth. By using the ontology and the data source schema, these heterogeneous data types are all mapped into appropriate domain concepts. The data extracted from the Web databases is internally reformatted to adhere to the unified data representation specified by the ontology.

Different from the traditional two-dimensional approach, the ontology developed in BACIIS has three dimensions: object, property, and relation. A portion of this three-dimensional ontology structure is shown in Figure 3. For example, *nucleic acid* and *protein* in Figure 3 are object classes. In the same figure, *nucleic-acid-sequ-info* and *amino-acid-sequ-info* are property classes. In the ontology of BACIIS, the properties of an

Figure 3. Portion of the BACIIS domain ontology



object are defined as a property class, which occupies a position in the property hierarchy tree. In ontology, a relation refers to the association between two concepts. In BACIIS, classes under the object and property dimensions are arranged into a hierarchy tree, based on the relation "is-a-subclass-of." For example, GENE is a subclass of the class NUCLEIC ACID in the object hierarchy. Similarly, "base-count" is a subclass of "nucleic-acid-sequ-info" in the property hierarchy. Also, relations under the relation dimension link concepts between the two dimensions property and object as well as within each of these dimensions. For example, the relation "has property" indicated by "—·→" in Figure 3 is between the object classes and property classes, as in the case of the class NUCLEIC ACID in the object hierarchy and the class "nucleic-acid-info" in the property hierarchy. The relation "is-a-subset-of" is used for property classes or object classes with parent-descendant relationships. It is indicated by "|—" in Figure 3. The relations "regulate" and "source-of" indicated by "→>" are used for object classes that are neither parents nor descendants of each other.

The three-dimensional structure of the ontology makes it possible to differentiate between two relations without affecting the traditional definition of a relation. Relations between two classes from different dimensions, object and property, are object-property relations. Relations between two classes within the same dimension are either property-property relations in the case of the property dimension or object-object relations in the case of the object dimension. This design choice makes the ontology more flexible and extendible. For example, changes to a property class will not have an impact on the definition of the object classes.

In order to resolve semantic and syntactic heterogeneities, in particular for tightly-coupled federated databases, ontology has recently emerged as an efficient framework for capturing biological knowledge (Goble, 2001). Ontology is basically a formal description of the concepts and entities in a given domain. A core part of the ontology is a vocabulary of terms and their meanings (Schulze-Kremer, 2002). Therefore, it is ideally suited for resolving both syntactic and semantic variations in the biological domain. Semantic variations are resolved by expressing the concepts in ontology at different degrees of granularity (3rd Millennium Inc., 2002). Furthermore, by expressing the relationships among the various concepts, the ontology organizes the concepts in a structured way, thus reducing the number of possible interpretations of these concepts (Schulze-Kremer, 2002).

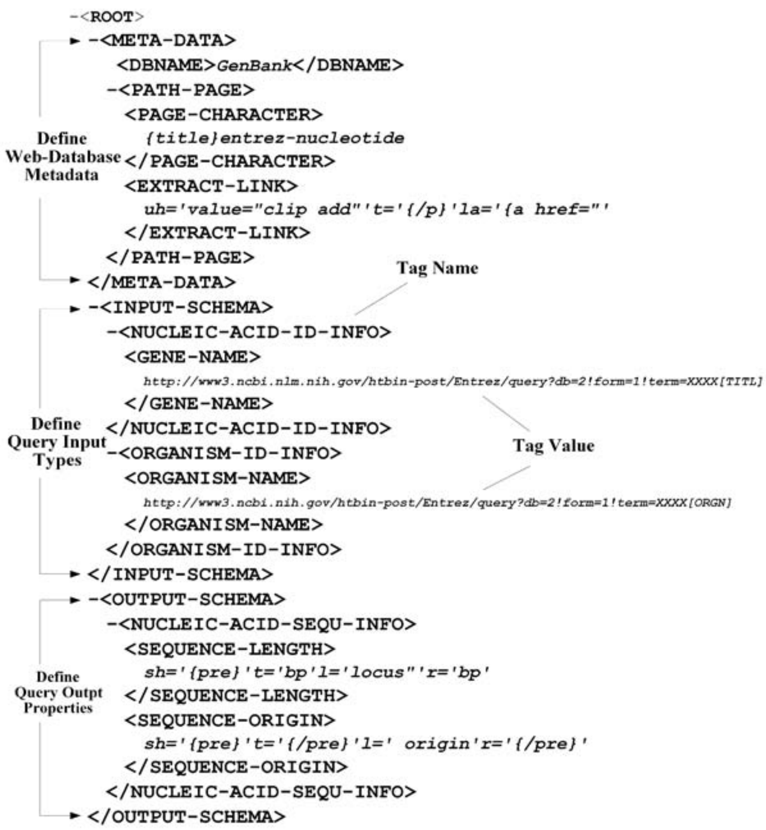
In BACIIS, we improve on this paradigm by making the ontology independent of the individual data schema of the component databases. Because of this independence, the ontology will change only when the biological domain evolves to include new discoveries and knowledge concepts rather than when a component database changes or a new component database is added. This independence is important in light of the fact that the life science databases change their Web interfaces often, and the fact that there is a large number of databases in the life science domain.

DATA SOURCE SCHEMA

The data source schema is used to describe each database participating in the integration. Each data source schema consists of the input data types accepted and the output data types generated by the corresponding Web database. These data types are expressed by using ontology terms and are used, as previously mentioned, to select the component databases that can answer a given subquery. Also, the expression of the input and output data types in terms of ontology terms facilitates synonym reconciliation and the resolution of presentation heterogeneity, since the schema of the different component databases are mapped to the same global schema.

Figure 4 shows a portion of the data source schema of the database GenBank. This schema consists of three sections: metadata, input data types, and output data types. The metadata section includes general information about the Web database, such as the database name. The input data types section includes all the input data types that are accepted by the Web database — in this case, GenBank. These input data types correspond to the property classes in the ontology, and they adhere to the same hierarchy established between the property classes in the ontology. For example, "nucleic-acid-id-info" corresponds to a property class in the ontology (Figure 3). This property class has the property subclass "gene-name." This parent-descendant relationship is preserved in the data source schema of Figure 4. The property classes in the ontology correspond to tag names in the data source schema. Some of the property classes also have tag values. These property classes correspond to input data types for the Web database, and they also correspond to property classes that are leaf nodes in the property dimension of the ontology. These leaf nodes have tag values that correspond to the URL that can be used to query the database by using the corresponding input data type (or tag name). For example, gene-name is a leaf node in the property hierarchy, and it is also an input data type for the database GenBank. Its tag value in the data source schema of GenBank is http://ww3.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db=2!form=1!term=XXXX[TITL], which can be used to query GenBank, using gene name as input. A given database may not accept all the property classes defined in the ontology. Therefore, only the property subtree that is induced by the input data types accepted by the database is included in the

Figure 4. Example of data source schema defined in XML format



data source schema. The output data type section in the schema file has the same characteristics as the input data type section, with the exception that the tag names and the tag values correspond to output data types that can be retrieved from the Web database and the extraction rules needed to extract this data from the pages returned by the Web database, respectively.

By using the data source schema to describe each Web database and the ontology to describe the domain knowledge, the integration system can identify the content of each database, determine which database(s) can answer a given query or subquery, retrieve and extract the information from this database, and integrate it with data from other databases. All of these operations can be performed by using the proposed approach, while preserving the autonomy of the databases and without any user intervention.

WRAPPERS

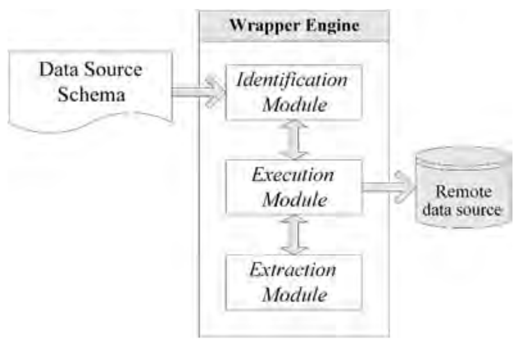
The wrappers are responsible for retrieving the data from remote databases. They have been used widely to extract information from Web databases (Decker, 2001; Lage, 2002). When integrating traditional databases, the wrapper usually resides on the server side. However, in the case of Web databases and because of their autonomy, the wrappers in BACIIS reside on the client side.

Because of the large number of databases and the fact that their Web interfaces change often, the wrapper approach used in BACIIS is different from a traditional one. The wrappers in BACIIS are constructed on the fly by using a generic wrapper engine and the data source schema of the target Web databases. Database-specific information, such as the URL addresses used to construct the subqueries for each input data type and the extraction rules needed to extract the required data from the returned result page for each output data type, are included in the data source schema.

The wrapper engine has three modules (Figure 5): the identification module, the execution module, and the extraction module. The identification module first constructs an empty result container for the result data for each query step. The result container is organized according to the relation hierarchy defined by the ontology. The execution module submits the database specific subquery to the remote database and fetches the return pages. The extraction module then parses the returned pages, extracts the desired data by using the proper extraction rules, and puts the extracted data into the result containers constructed by the identification module. The wrapper engine extracts all the database-specific information that it needs from the data source schema file. In a sense, the wrapper engine acts like a compiler that interprets the instructions included in the data source schema file. Since the wrapper engine does not include any database-specific information, adding a database requires only creating the corresponding data source schema. Furthermore, changes to the Web interfaces are only reflected in the data source schema. This is different from wrappers in other integration systems, which construct a wrapper for each database independently (Decker, 2001).

The data container is constructed by using a tree structure derived from the property hierarchy in the ontology, which includes all the descendants of the output data type. The mediator collects the

Figure 5. Wrapper



data containers from all the wrappers and organizes them according to the property hierarchy defined in the ontology. The query results are organized in a tree structure. The information belonging to the same property is stored in the same node. The hierarchy structure and concept definition imposed by the ontology allows data from different component databases to be combined semantically.

EXAMPLE QUERY

This section illustrates the execution of the following query in BACIIS: "Find all the literature references and DNA sequences for the mouse gene RAS and also return the amino acid sequences and protein 3-D structures for the proteins that are encoded by this gene." The query will be expressed as follows: "select *nucleic-acid-sequ-info*, *references*, *amino-acid-sequ-info*, *protein-3D-structure*, where *organism-name*=mouse and *gene-name*=RAS." The user expresses the multi-database query by using query input data types and query output data types from the ontology. In this example, the output data types *nucleic-acid-sequ-info*, *references*, *amino-acid-sequ-info*, *protein-3D-structure* and the input data types *organism-name* and *gene-name* are all ontology terms.

From the Web interface, the user can compose this query by specifying the input and output data types and values. The input and output data types displayed to the user are obtained directly from the ontology. The user specifies only the top-level property classes as output data types. The subclasses or descendants of these property classes are retrieved from the ontology during query processing. The property classes "reference," "nucleic-acid-sequ-info," and "amino-acid-sequ-info" are all top-level property classes. Figures 6 and 7 show how the user selects the query input data types and query output data types from the ontology in order to compose this query.

Figure 6. Compose query input

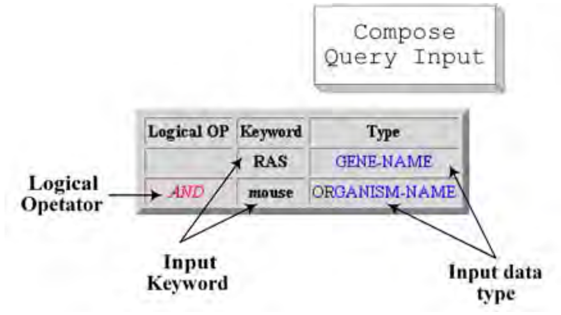
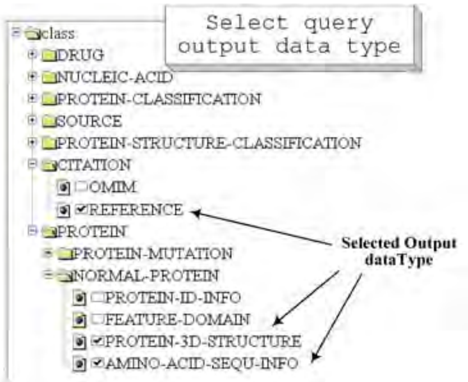


Figure 7. Select query output types



More specifically, in order to build the query input part, the user enters "RAS" as a keyword and selects "gene-name" as the input data type. The user then adds the second input by entering "mouse" as a keyword and selecting "organism-name" as the input data type. For the output part, the user can select the object class "PROTEIN" from the ontology. When the user selects "NORMAL-PROTEIN," which is an object subclass of the object class "PROTEIN," the property classes that are associated with this object subclass will be automatically retrieved from the ontology. For example, the object class "NORMAL-PROTEIN" is linked to the property class "normal-protein-info" by the relation "has-property" (shown by "—·→" in Figure 3). Thus, when "NORMAL-PROTEIN" is selected, the ontology property subtree rooted at "normal-protein-info" will be displayed to the user. The user then can select the output data types "amino-acid-sequ-info" and "protein-3D-structure," which are top-level property classes of this property subtree. The top-level property class "amino-acid-sequ-info," for example, has multiple descendants, including "origin," "base-count," "version-no," and so forth, as shown in Figure 3. In order to complete the query, the user can add the other output data types first by selecting the object "nucleic-acid," and then by choosing the property "nucleic-acid-sequ-info" from the property list returned by the ontology for the object "nucleic-acid." In order to specify reference information for the gene "RAS," the user needs to choose one more property, "reference," from the property list of the "citation" object in the ontology. The result of the example query is shown in Figure 8.

RELATED WORK

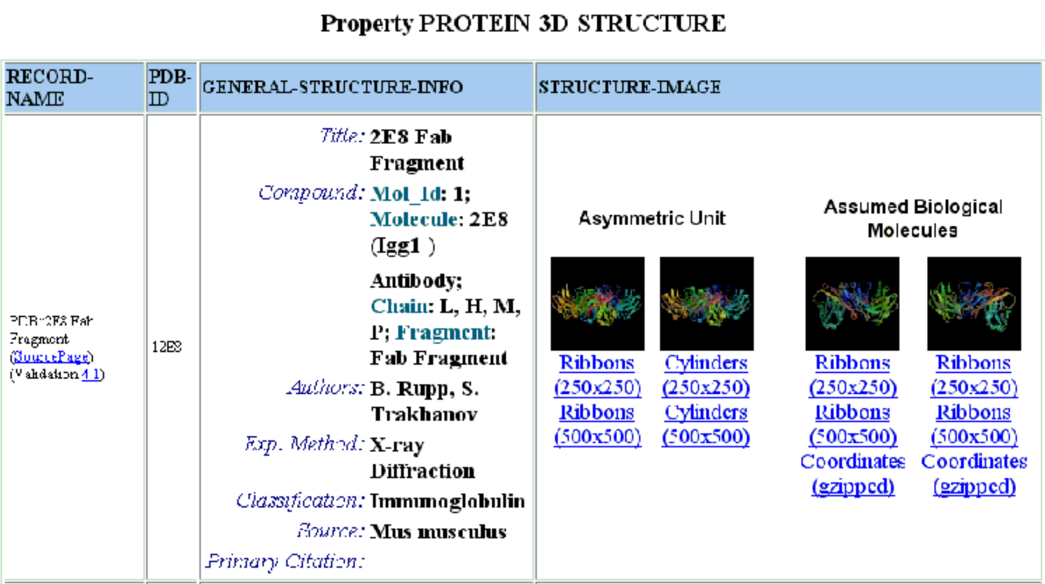
One of the main advantages of federated databases is that they do not require that component databases be physically combined into one database (Karp, 1995). Database federation is used to integrate autonomous, heterogeneous, and distributed databases (Busse, 1999). This class of integration systems can be divided further into two subclasses: tightly coupled and loosely coupled.

In a tightly coupled federated database system, a global data schema is constructed, and the queries are expressed on this unified schema (Yu, 1998). The global schema integrates all the component schemas. Therefore, the component databases are transparent to the user. TAMBIS (Goble, 2001; Paton, 1999) is an example of tightly coupled federated databases. It uses ontology as a conceptual model to integrate multiple biological Web databases.

In a loosely coupled federated database system, there is no global schema, and the user needs to specify the component databases in the query by using a multi-database query language (Busse, 1999). One of the disadvantages of loosely coupled federated databases is that the component databases and any heterogeneity that may be present among their schema are not transparent to the user. This approach is used by BioKleisli (Davidson, 1997).

As previously mentioned, both BACIIS and TAMBIS use the tightly coupled federated databases approach with an ontology-based global schema. However, they differ in the way the

Figure 8. Partial result page of the example query



Property PROTEIN 3D STRUCTURE

RECORD-NAME	PDB-ID	GENERAL-STRUCTURE-INFO	STRUCTURE-IMAGE
PDB:2E8 Fab Fragment (SourcePage) (Validation 4.1)	12E8	*Title:* 2E8 Fab Fragment *Compound:* Mol_Id: 1; Molecule: 2E8 (Igg1) Antibody; Chain: L, H, M, P; Fragment: Fab Fragment *Authors:* B. Rupp, S. Trakhanov *Exp. Method:* X-ray Diffraction *Classification:* Immunoglobulin *Source:* Mus musculus *Primary Citation:*	Asymmetric Unit: Ribbons (250x250) Ribbons (500x500); Cylinders (250x250) Cylinders (500x500) Assumed Biological Molecules: Ribbons (250x250) Ribbons (500x500) Coordinates (gzipped); Ribbons (250x250) Ribbons (500x500) Coordinates (gzipped)

wrappers are generated. In TAMBIS, wrappers are database-specific and are constructed by using functions from the CPL language. A similar approach is used by BioKleisli.

CONCLUSION AND FUTURE WORK

The contribution of the BACIIS system is that it integrates geographically distributed heterogeneous Web databases with total transparency, while preserving the autonomy of the databases. The life science domain is unique in the sense that it counts Web databases in the hundreds. Furthermore, these databases are not accessible via their DBMSs, and they do not adhere to a common standard interface or data model. The BACIIS system overcomes these limitations through a scalable design with an ontology-centric approach. Most of the previously developed life science integration systems do not hide the heterogeneity of the databases from the users.

The use of an ontology-centric approach of a three-dimensional ontology and of independent data source schema that contain database specific information reduces the overhead associated with adding new databases or changing the existing databases, while addressing the heterogeneity of the component databases.

While the current BACIIS successfully integrates multiple life science Web databases, the implementation of BACIIS has one limitation that needs to be addressed: The extraction rules used stored in the data source schema files are constructed manually. These extraction rules need to be modified every time the interfaces of the Web databases change. Although this is an improvement on traditional wrappers, which are entirely hand-coded, the manual update of the data source schema files is still tedious. The development of an inductive wrapper algorithm that will generate the extraction rules automatically is being considered. We are currently designing a method similar to the one proposed by Lage (2002) for simple Web interfaces. However, in the case of BACIIS, the wrapper induction approach has to be able to handle complex Web interfaces.

BACIIS is available at http://baciis.engr.iupui.edu/

AKNOWLEDGMENT

This work is supported in part by NSF CAREER DBI-0133946 and NSF DBI-0110854.

REFERENCES

3rd Millennium Inc. (2002). *Practical data integration in biopharmaceutical R&D: Strategies and technologies.* White paper.

Ambite, J.L., et al. (1998). ARIADNE: A system for constructing mediators for Internet sources. *Proceedings of the ACM Sigmod International Conference on Management of Data*, (pp. 561-563).

Arene, Y., Knoblock, C.A., & Shen, W.M. (1996). Query reformulation for dynamic information for dynamic information integration. *International Journal of Intelligent and Cooperative Information System, Special Issue on Intelligent Information Integration*, *6*(2/3), 99-130.

Bairoch, A. (2000). The ENZYME database in 2000. *Nucleic Acids Research*, *28*, 304-305.

Bairoch, A., & Apweiler, R. (1998). The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1998. *Nucleic Acids Research*, *26*, 38-42.

Baker, P.G., & Brass, A. (1998). Recent developments in biological sequence databases. *Current Opinion Biotechnology*, *9*, 54-58.

Baxevanis, A.D. (2002). The molecular biology database collection: 2002 update. *Nucleic Acids Research*, *30*, 1-12.

Benson, D.A., Goguski, M.S., Lipman, D.J., Ostel, J., & Ouellette, B.F. (1998). GenBank. *Nucleic Acids Research*, *26*, 1-7.

Berman, H., et al. (2000). The protein data bank. *Nucleic Acids Research*, *28*, 235-242.

Bouguettaya, A. (1999). *Ontology and databases*. Kluwer Academic Publishers.

Busse, S., Kutsche, R.D., Leser, U., & Weber, H. (1999). Federated information systems: Concepts, terminology and architectures. *Technical report*, *99*, 9.

Davidson, S.B., Overton, G.C., Tannen, V., & Wong, L. (1997). BioKleisli: A digital library for biomedical researchers. *Journal of Digital Libraries*, *1*(1), 36-53.

Decker, K., Zheng, X., & Schmidt, C. (2001). A multi-agent system for automated genomic annotation. *Proceedings of the Fifth International Conference on Autonomous Agents*, (pp. 433-440).

Decker, S., Erdmann, M., Fensel, D., & Studer, R. (1999). Ontobroker: Ontology based access to distributed and semi-structured information. In Semantic issues in multimedia systems. *Proceedings of ds-8*, (pp. 351-369).

Goble, C.A., et al. (2001). Transparent access to multiple bioinformatics information sources. *IBM Systems Journal*, *40*(2), 532-552.

Gruber, T.R. (1993). Towards principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, *43*, 907-928.

Karp, P. (1995). A strategy for database interoperation. *Journal of Computational Biology*, *2*(4), 573-586.

Kossmann, D. (2000). The state of the art in distributed query processing. *ACM Computing Surveys*, *32*(4), 422-469.

Lage, J.P., Silva, A.S., Golgher, P.B., & Laender, A.H.F. (2002). Collecting hidden Web pages for data extraction. *Proceedings of Fourth International Workshop on Web Information and Data Management*, (pp. 69-75).

Lakshmanan, L.V.S., Sadri, F., & Subramanian, I.N. (1996). SchemaSQL: A language for interoperability in relational multidatabase systems. *Proceedings of the 22nd International Conference on Very Large Databases*, (pp. 239-250).

Levy, A.Y. (1998). The information manifold approach to data integration. *Proceedings of the IEEE Intelligent System*, (pp. 1312-1316).

McKusick, V.A. (1998). *Mendelian inheritance in man. Catalogs of human genes and genetic disorders (12th ed.).* Baltimore, MD: Johns Hopkins University Press.

Mena, E., Kashyap, V., Illarramendi, A., & Sheth, A. (1998). Domain specific ontologies for semantic information brokering on the global information infrastructure. *Proceeding of the International Conference on Formal Ontology in Information Systems 1998*, (pp. 269-283).

Ouzzani, M., Benatallah, B., & Bouguettaya, A. (2000). Ontological approach for information discovery in Internet databases. *Distributed and Parallel Databases, An International Journal*, *8*(3), 367-392.

Paton, N.W., et al. (1999). Query processing in the TAMBIS bioinformatics source integration system. *Proceedings of the Eleventh International Conference on Scientific and Statistical Databases (SSDBM)*, (pp. 138-147).

Schulze-Kremer, S. (2002). Ontologies for molecular biology and bioinformatics. *Silico Biology*, *2*(3), 179-193.

Sigrist, C.J., et al. (2002). PROSITE: A documented database using patterns and profiles as motif descriptors. *Brief Bioinform*, *3*(3), 265-274.

Wiederhold, G. (1992). Mediators in the architecture of future information systems. *IEEE Computer*, 38-49.

Yu, C.T., & Meng, W. (1998). *Principles of database query processing for advanced applications*. Morgan Kaufmann.

This work was previously published in Journal of Database Management, Vol. 16, No. 3, edited by K. Siau, pp. 72-85, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 2.19
Macroscopic Modeling of Information Flow in an Agent-Based Electronic Health Record System

Ben Tse
University of Regina, Canada

Raman Paranjape
University of Regina, Canada

ABSTRACT

This chapter presents an architecture, or general framework, for an agent-based electronic health record system (ABEHRS) to provide health information access and retrieval among different medical services facilities. The agent system's behaviors are analyzed using the simulation approach and the mathematical modeling approach. The key concept promoted by ABEHRS is to allow patient health records to autonomously move through the computer network uniting scattered and distributed data into one consistent and complete data set or patient health record. ABEHRS is an example of multi-agent swarm system, which is composed of many simple agents and a system that is able to self-organize. The ultimate goal is that the reader should appreciate the benefits of using mobile agents and the importance of studying agent behaviors at the system level and at the individual level.

INTRODUCTION

Health information is the heart of any electronic health record (EHR) application, and the management of this growing and complex information has become a huge and complicated task for every health care organization. Moreover, much of this health information is fragmented and scattered among different medical service facilities, mainly because a patient is usually seen by many different health care providers in many different medical service facilities. So, the patient is likely to have at least one relevant health document or record in each medical facility he/she visits. Over time, the patient's medical record becomes fragmented, inaccurate, and inconsistent. This is the complex environment in which EHR management must operate.

Fortunately, a large part of today's health information is already available in electronic form, and the primary task of the EHR system is to communicate this information electronically anywhere in the world. Based on this vision, the need to move the patient's health information among different health care facilities grows in parallel with the need to centralized health information across numerous EHR systems within the health care domain (Silverman, Andonyadis, & Morales, 1998). Therefore, in order to allow rapid and accurate access of patients' health record, the creation of a digital health information network is essential to our modern health care system.

Mobile agent technology can be an essential tool in creating this system, because of fundamental mobile agent attributes including: autonomy, mobility, independence, pro-activity, reactivity, communicability, swarm, and group intelligence. Because of these attributes, mobile agents can provide solutions to address the challenges of health information systems in the health care domain, as suggested in Nealon and Moreno (2003) and Crow and Shadbolt (2003). Later section in the chapter, we will present an architecture or general framework for ABEHRS to provide health information access and retrieval among different medical services facilities.

As shown in Lerman and Galstyan (2001), a swarm agent system (simple agents with a distributed controller) that exhibits collective behaviors has demonstrated advantages in robustness, scalability, and cost, over a traditional agent system (intelligent agents with a centralized controller) that exhibits individualized behaviors. Moreover, self-organization is one of the interesting collective behaviors that appears in a swarm system. This concept of self-organization manifests itself in our ABEHRS, which allows the patient's health record to self-organize. In the background section, basic definitions of swarm and intelligence will be presented, and the definition of self-organization will also be presented. Later sections in the chapter will present the reasons why the swarm paradigm is chosen over the traditional paradigm and the process of self-organization used in our ABEHRS.

In order for our ABEHRS to be useful, a study of the system-level behavior must be performed. This allows a greater understanding of the system dynamics, so that if such a system was actually implemented, the dangers of unexpected system behavior could be limited. Smith, Paranjape, and Benedicenti (2001) show that the full potential of individual agents is not obtained during unexpected/unwanted agent-group behaviors (for example, oscillation of agents between hosts). Therefore, we believe that the behavior of agents in a multi-agent system must be completely understood before the completion of actually system design, so as to minimize the chances of system failure and to achieve superior system and individual agent performance. In the proceeding sections, we present a study of agent behaviors in ABEHRS using simulation and mathematical models, as well as the comparison between the simulation and numerical results.

BACKGROUND

The Electronic Health Record

As defined in Grimson, Grimson, and Hasselbring (2000) and the Web site of the Healthcare Information and Management Systems Society (http://www.himss.org), an electronic health record is an electronic version of patient's health record, which includes prescriptions, lab results, evaluations by doctors, and so forth. Since the EHR is in an electronic format, so it can be easily accessed by a health information system within an electronic health network. Around the world, people are recognizing the value of EHR systems. Their advantages include increasing effectiveness and efficiency of clinical staff and health practitioner, providing seamless care for patients with timely, appropriate, and secure access to the right information whenever and wherever they are, and increasing the profitability of the practice or facility.

Although EHR appears to hold great promise, there are many challenges that need to be addressed before they can be fully integrated in a health care system. As described in Nealon & Moreno (2003), Huston (2001), and Walczak (2003), these challenges include scattered and fragmented health information, security and confidentiality, delay in availability of information, complexity of medical data, heterogeneous health information system, and network bandwidth consumption.

Mobile Agent Technology

A mobile agent (MA) is a software object that can migrate to different computers over an Internet protocol (IP) network, or the Internet, to perform user-assigned and self-initiated tasks in heterogeneous computing environments (Gray, Kotz, Nog, Rus & Cybenko, 1997). An MA is a representation of its owner, and it inherits the owner's authority and acts on behalf of the owner. By adding artificial intelligence modules to a MA, it can have the ability to select the best set of activities to perform until it reaches the goal specified by the owner, or the ability of reacting to specific events in its surrounding environment to provide flexible, dynamic, and intelligent solutions. Mobile agents exhibit a number of characteristics, and they include:

- **Modularity:** Mobile agents are diverse in nature; different services can be distributed among a set of agents or encapsulated into a single complex agent. In this way, the required algorithms can be embedded into the agent system to perform the service
- **Mobility:** Agents can migrate to the resource they need instead of accessing these resources remotely, which reduces network bandwidth usage tremendously
- **Interactivity:** Agents can communicate to other agents using a specified message format
- **Autonomy:** Agents are typically independent entities and have the ability to choose the best actions to perform to achieve specific goals

Mobile agents do not operate independently; they required extensive supporting infrastructure to provide communication services, mobility services, management services, security, and monitoring services. This infrastructure is commonly called an agent execution environment, or AEE. The AEE must be installed and running on a host computer before the mobile agent program can run. TEEMA (TRLabs execution environment for mobile agents) is an AEE that was developed by the faculty, staff, and graduate students at TRLabs Regina and Electronic Systems Engineering at the University of Regina. More information related to TEEMA can be found in Gibbs (2000) and Martens (2001). TEEMA is the experimental platform used in this work.

There are various software applications that are uniquely suited for development using mobile agent technology. Typically, these applications take advantages of the intrinsic characteristics of mobile agents, including autonomy, re/proactivity, modularity, interaction, and/or mobility. And our ABEHRS is one of the applications that take advantage of these intrinsic characteristics of agent to facilitate the process of retrieving/accessing fragmented and scattered patient health information across different medical service sites. So, each mobile agent has the ability to "move" around different medical service facilities across

IP networks, the ability to "automate" tasks on behalf of the patient, the ability to act in a "re/pro-active" bases to respond to the patients or medical providers, the ability to "interact" with other agents, and the ability to "embed" partial or whole patient information as a function. After the health information is encapsulated by the agent, it can be moved to different medical sites within our ABEHRS, we can define these behaviors/actions as information flow within ABEHRS.

Swarm and Its Intelligence

The idea of an intelligent agent swarm takes its inspiration from insect societies in biology and nature. Swarm consists of a large group of small insects (such as bees or ants), and each insect has a simple role, with limited capability (for example, moving around). The swarm as a whole (such as a colony of ants or a hive of bees), however, is able to do many complex and sophisticated tasks (such as finding the shortest distance between the nest and the food source). Furthermore, each insect works autonomously, and essentially their teamwork is self-organized. Coordination and cooperation between individual insects arises from the interactions between insects, or the interactions between insects and the environment. Through these interactions (such as stigmergy), collective behavior emerges, and efficient and effective outcomes are generated.

As defined in Bonabeau, Dorigo and Theraulaz (1999), swarm intelligence is the concept of using the collective behavior arising from a group of relatively simple objects in a decentralized and self-organized system to solve complex and distributed problems. Artificial swarm intelligence can be demonstrated using relatively simple and autonomous agents. The agents are distributed, self-organized, and there is no central controller to direct how each agent should behave. Agents may be dynamically added or removed, making the system highly flexible and scalable. Examples of agent swarm systems are Truszkowski, Hinchey, Rash, and Rouff (2004), Wu (2003), and Charles, Menezes, and Tolksdorf (2004).

Self-Organization Definition

According to Camazine et al. (2001), self-organization is a key feature in a swarm system. It refers to the process that describes a system as it moves from a chaotic state to a well-ordered state by using the interactions at the lowest-level components of the system. Camazine et al. also pointed out the four characteristics of self-organization in a swarm system, and they are: positive feedback, amplification of fluctuations, multiple interactions, and negative feedback. When the system is in the chaotic state, the positive feedback signals will influence certain properties of the system, and these properties will be amplified by the many random interactions between the individual components. These interactions can be seen as a combination of exploitation and exploration of new solutions from the components. The negative feedback serves as a regulatory control mechanism of the system's properties, and is used to direct the system back to a well-ordered state. In fact, the self-organization of a system can be seen as a formulated feedback control loops between the individual components in the system. An example of a self-organized swarm system is Kadrovach and Lamont (2002).

By combing the concept of self-organization with agent technology, we are able to allow individual health record to self-organize themselves in our ABEHRS.

Agent Behaviours Modeling

In order to ensure that a useful and effective mobile agent system is constructed, it is important to study, examine, and analyze the system-level behaviors, as well as the agent-level behaviors. There are several approaches used to study behavior in multi-agent systems; most include simulation and mathematical modeling. Using the simulation approach to evaluate agent behaviours in multi-agent systems is a traditional approach for agent designers and researchers. In this approach, the simulation consists of a set of agents that encapsulate the behaviours (for example, rules of movement, strategies) of various individuals that interacted with the local environment and with other individuals. Usually, these behaviours, as well as external forces (for example, resource fluctuation), are represented using different stochastic processes, and so, by executing these behaviours during the simulation, a global dynamic model of the system and of the agents can be created. The simulation model provides reasonable design ideas to the agent designers when developing a real multi-agent system. Simulations of agent behaviour in multi-agent system are shown in Tecchia, Loscos, Conroy, and Chrysanthou (2001), Smith et al. (2001), and Heine, Herrier and Kirn (2005).

The mathematical modeling approach is an alternate way of evaluating agent behaviours. The mathematical model usually consists of a set of equations that describe the process of the behaviours being modeled, and these behaviours are represented by a set of stochastic probabilities that are used in the equations. In order for the mathematical model to be useful and realistic, relevant details of the process must be accounted for. Generally, two different levels of mathematical modeling are used to describe a multi-agent system, and the levels are microscopic and macroscopic. In general, a microscopic model describes the agent-level behaviours in a multi-agent system, for example, interactions between two agents. On the other hand, a macroscopic model describes the system level behaviours in a multi-agent system, for example, collective or global behaviour emerged from the agent's interactions. Studies of agent behaviours using microscopic or macroscopic mathematical models are shown in Agassounon and Martinoli (2002), Lerman and Galstyan (2001), Lerman and Shehory (2000), Sugawara and Sano (1997), and Wang, Liu, and Jin (2003).

In this work, we choose the approach of agent system evaluation using simulation and mathematical modeling. For the simulation, a prototype of ABEHRS model was developed using TEEMA platforms, and simulation results were obtained. For the mathematical modeling, we choose a macroscopic model to analyze system behavior in ABEHRS. The reasons to use a macroscopic model are that global dynamic behaviors of the system can be readily observed, and macroscopic modeling is computationally more efficient than microscopic model, since only a few parameters are used. Moreover, among the mathematical modeling approaches mentioned above, Lerman and Galstyan (2001) have presented a general mathematical macroscopic approach to analyze the global dynamic behaviors of multi-agent swarm system. In our mathematical model, a set of differential equations based on the Lerman and Galstyan (2001) model are determined and numerical results are obtained by solving these equations. We can compare and contrast these results in the context of our ABEHRS. It is our belief that a simple multi-agent swarm system will exhibit complex and unexpected behaviors, and unless we model these behaviors using a macroscopic mathematical model and simulation, unpredictable system behavior will result.

THE AGENT-BASED ELECTRONIC HEALTH RECORD SYSTEM

Our main objective in this work is to propose a general framework or architecture that incorporates the use of mobile agent technology to facilitate the flow of health records among various medical service facilities, and combines all scattered and fragmented health records into one consistent and complete data set.

To accomplish the system objective, agents must exhibit the following characteristics: autonomy, re/pro-activity, mobility, modularity, and interaction. With the presence of autonomy and re/pro-activity, agents can monitor any changes in the patient's electronic health record. Whenever a change is identified to the patient's electronic record, the agent will automatically perform necessary tasks to ensure the electronic record of the patient is up-to date and is consistent. Thus the ABEHRS agents are highly autonomous. Another important characteristic of agents is mobility. With this characteristic, agents are able to migrate to different medical service facilities to perform necessary data collection tasks. Interaction between agents is mandatory in our system, and it is used in identity verification of patients and agents at different medical service facilities. It is also used in the co-ordination of other agents' activities. Agents can encapsulate health information, communication protocols, or medical data en/decoding formats as embedded functions. By using mobile agents and their fundamental and intrinsic characteristics, many of the challenges of the EHR systems can be overcome. In the subsequent sections, we will present system concepts, design concept, and the system components for the ABEHRS.

System Concepts

By using mobile agent technology, we add mobility and autonomy to the health record, which allows the record to automatically move anywhere within the health information network. This system can be colloquially described as "putting a mobile agent wrapper" around an electronic health record and instructing the agent to move to other medical facilities for its owner (the patient).

There are two important aspects of the system, which are shown below:

1. A complete health record is defined as every pieces of information in a patient's health record (which may be generated by different medical facilities within the health care system) united into one consistent and complete set of information. The health record may include doctor's evaluations, test results, prescriptions, and so forth.
2. The ABEHR system is described as self-completing, self-regulating, and self-organizing. We define these capabilities for the system as the mechanism by which it automatically (without any supervision or guidance) makes each patient health record complete.

Therefore, we wish to apply the idea of self-organization through agents to facilitate the concept of self-completing patient's health records. The following subsections will provide an overview of how self-organized health record is achieved, and how agent cloning is used in ABEHRS.

Self-Organized Health Record

As defined above, self-organization refers to the process of a system moving from a chaotic state to a well-ordered state by using the interactions between the lowest-level components of the system. From the perspective of the individual patient's health record, an individual health record is in a well-ordered state, when the record is complete and consistent. If the health record is not complete and consistent (for example, newly created record), then the record is in a chaotic state. Mobile agents are effectively the lowest-level component, which autonomously interact with other agents in the system to move a chaotic health record into a well-ordered state (for example, update health record in HRC and in patient's health storage device). Once initiated, agents perform the necessary actions to have the health records system self-organize. All the actions and interactions related to self-organized health record performed by the agents are presented in the section of Typical Actions of Patient Agent in Different Medical Sites.

Agent Cloning

In order to facilitate the process of self-organization and to provide needed information in a timely fashion, agent cloning is used in our ABEHRS. By using the agent-cloning approach, the agents are able to perform parallel tasks at different medical sites. Also, when inheriting sensitive information (for example, health card number, access control permissions) from the initiator (the agent who initiates the cloning process), more secure control of data may be achieved than passing the information to other agents using messaging. Therefore, cloning of mobile agents is the process typically used in the ABEHRS. The events that trigger the cloning process depend on the medical services required by the patient. Medical services could be record update/retrieval services, prescription-filling services, lab test services, and so forth. Using this agent-cloning approach, a group of agents can be easily generated, and their direct and group actions work, to keep the health record complete and consistent for the individual patient.

Design Concepts

As previously stated, multi-agent systems are either developed using the traditional paradigm with centralized control and deliberative agents or the swarm paradigm with distributed control and simple agents. There are various factors that influence the decisions as to which design paradigm to adopt for the development of ABEHRS, and Lerman and Galstyan (2001) had shown us that the swarm paradigm is a more suitable approach than the traditional paradigm for the development of our ABEHRS. Reasons supporting this decision are the following:

- Health information systems are diverse in nature. This diversity makes it almost impossible to develop a centralized controller for a large group of health information systems.
- Deliberative agents are smarter, autonomous, and more individualized than swarm agents. Deliberative agents contain deeper knowledge about the system and have higher capabilities to perform more complicated tasks. In order to get the full potential out of the deliberative agents, agent designers must have detailed coordination and complete plans of actions for the agents. These robust mechanisms simply do not exist. Conversely, agents in the swarm are simple and task-oriented, so agent designers can easily develop mechanisms to coordinate and plan the actions of agents. Also, swarm agents require only the knowledge necessary to perform the specified task. Thus, designing and coding swarm agents is more simple and less error prone than deliberative agents.
- In general, health information organizing is conceptually simple but it requires collective effort from a large group of agents to accomplish. When the task is performed in a collective way, the operation is more effective and efficient than if an individual complex agent attempts to perform it. Coalition formation of deliberative complex agents is not easily achieved, while for swarm agents, coalition forming can scale up to thousands of agents easily.
- The health care industry is very dynamic and rapidly changing (for example, standard of health information, medical protocols). The adaptation to a changing process for swarm agents is less complex than the adaptation for deliberative agents. This is mainly due to the deliberative agents containing a large amount of knowledge about the system that is hard to substitute. Swarm agents, on the other hand, contain only the necessary knowledge to perform a specific task, and coordination between swarm agents is self-directed. So a swarm agent system will be able to tolerate changes in the health care industry more effectively.

Thus, we see that the properties of swarm agents (simple, interchangeable, distributed, dynamically add/delete) and the features of the agent swarm system (self-organization, self-regulation, collective action, ant forging) can be effectively applied to real-world problems in the health care domain.

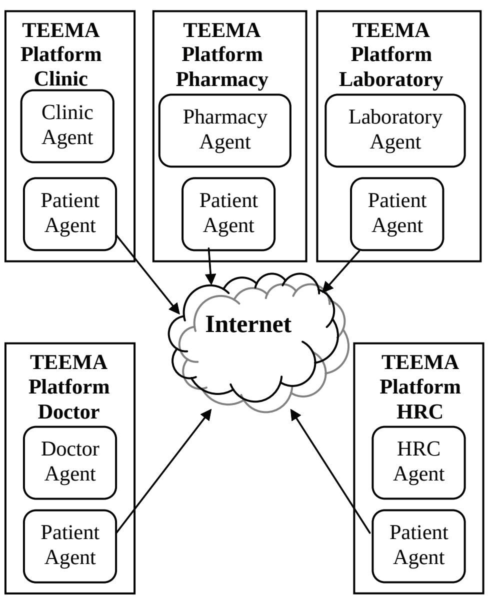
System Components

The agent-based electronic health record system is developed using the TEEMA platform. A system overview is shown in Figure 1. Each TEEMA platform represents a certain physical location, such as a clinic or laboratory. In each platform, there are a single stationary and mobile agent. The system contains five different types of stationary agents and one mobile agent.

Stationary agents

- Clinic agent (CA) is responsible for verifying the identity of the patient, and creating a mobile patient agent if a patient arrives at a clinic and requests medical services. It can be seen as an agent interface between the patient, clinical staff, and the patient agent.
- Doctor agent (DA) is responsible for managing the doctor's evaluation of the patient. It can be seen as an agent interface between the patient agent and the doctor.
- Pharmacy agent (PhA) is responsible for verifying the identity of the patient agent, retrieves prescriptions from the patient agent, and sends notification to the patient agent when the patient or owner wishes to pickup his/her prescription. It can be seen as an agent interface between the patient agent and the pharmacist.
- Lab agent (LA) is responsible for verifying the identity of the patient agent, retrieves lab work orders from the patient agent, sends notification to the patient agent when the patient or owner arrives at the lab, and sends notification to the patient agent when the test result for its owner is prepared. It can be seen as an agent interface between the patient agent and the laboratory technicians.
- Health record central agent (HRCA) is responsible for controlling access to the health record database that holds the personal and medical information of all patients. It can be seen as an agent interface between the patient agent and the central database.

Figure 1. An overview of the ABEHRS architecture



Mobile agents

- Patient agent (PA) is responsible for updating the patient health record that is stored in the patient's storage device card, transferring records to the health record central, and if there is a prescription and/or lab test needed, the patient agent will clone itself and migrate to pharmacy and/or laboratory. It can migrate to all of the different platforms in the simulation.

Typical Actions of Patient Agent in Different Medical Sites

The process begins when a patient visits the medical clinic. The patient places his/her health storage device card (SC) into a card reader, and the clinic agent performs verification (for example, fingerprint) on the patient's identity. If the verification is successful, a patient agent is created, and the PA copies the patient's entire health record from the SC into its payload area. If the verification fails or any other incidents (for example, the patient forgets to bring his/her SC) take place, then the clinical staff will be involved to resolve the incidents. The first operation of the PA is to review the medical service request log (for example, request lab work order) from the patient's record. By doing this, the PA will have enough data to decide what actions to perform next. If the patient's record is not up to date (for example, missing lab test result), then the PA will clone itself. The PA clone will migrate to health record central (HRC), retrieve the latest record from HRC database, migrate back to the clinic with the record, send the new record to the original PA, insert the record into the patient's SC, and finally destroy itself. After the cloning process, the original PA will migrate to the platform of the doctor that has been chosen by the patient, and inform the doctor agent about its owner/patient who needs the consultation service. The original PA will then wait at the doctor platform for the examination to be finished.

During the examination of the patient, which is performed by the doctor, the doctor is able to review the patient's completed and consistent health record. This access to the complete health record will allow the doctor to provide a diagnosis quickly and effectively. After the doctor finishes the examination of the patient, the DA will save the results of the medical evaluation into a specific location on the doctor's storage device, and then the DA will send a notice to the PA to retrieve the medical evaluation (for example, diagnosis, prescription form, lab work order form) of the patient from that specific location. After the PA retrieves the evaluation, it will migrate back to the clinic platform to perform actions required to serve the patient.

Once the PA migrates back to the clinic, it will check the doctor's evaluation for any medical services requested by the doctor. If the medical services are requested, then the PA will clone itself, and the clone will migrate to the platform of the medical service provider that is chosen by the patient. For example, if the requested service is filling a prescription, then the PA clone will migrate to the platform of the pharmacy that is chosen by the patient. This cloning process will be repeated until all requested medical services have been looked after. At the end of the process, the original PA will save the medical evaluation to the patient's SC. Finally, the original PA migrates to HRC and appends the newest medical evaluation to the complete set of the patient's health record in the database, and then destroys itself.

When the cloned PA arrives at the platform of the patient's chosen pharmacy, it will first verify itself to the pharmacy agent at the pharmacy service site. If the verification is successful, then the PA will send the prescription content to the PhA and hibernate. If the verification fails, then the PA will destroy itself. When the patient visits the pharmacy and wishes to pickup his/her prescription, then the PhA will inform the PA about the arrival of the patient, and this will trigger the PA to de-hibernate. Once the PA de-hibernates, it will perform verification on the patient's identity and the prescription content inside the patient's SC, and the cloned PA will destroy itself after a successful verification. If the verification fails, then a notice will send to PhA, which in turn sends a notice to pharmacy staff. And the process of verification will be repeated until a successful verification occurs or a verification override is executed by the facility staff. At the end, the PA will be forced to act as if there is a successful verification and destroys itself.

When the cloned PA arrives at the platform of the patient's chosen laboratory, it will first verify itself to the laboratory agent at the laboratory service site. If the verification is successful, then the PA will hibernate. If the verification fails, then the PA will destroy itself. When the patient visits the laboratory and wishes to go through the lab testing procedures, the LA will inform the PA about the arrival of the patient, and it will trigger the PA to de-hibernate. Once the PA de-hibernates, it will perform verification on the patient's identity and the work order form in the patient's SC. If the verification fails, then a notice will be sent to the LA, which will in turn send a notice to laboratory staff. And the process of verification will be repeated until a successful verification occurs or verification override is executed by the facility staff. After a successful verification, the PA will send the lab work order to LA and re-hibernate until the lab test results are ready to be retrieved. After the lab technician generates the test results for the patient, the LA will inform the PA, and it will trigger the PA to de-hibernate again. Once the PA is de-hibernated, it will retrieve the results and go through the process of cloning. The clone of the PA will migrate to HRC, update/append the lab test result to the complete set of the patient's health record, and then destroy itself. After the cloning process, the PA will migrate back to the clinic where it was originally formed, store the test results into a specific location, inform the doctor (whom the patient consulted with) via the doctor's DA about the test results and the location of where the test result is stored, and finally destroy itself.

Once the PA arrives at the HRC, it will verify itself to the heath record central agent before it can access the centralized health information database. If the verification fails, the PA will destroy itself. If there is a successful verification, the PA can either retrieve or update health information from or to the complete set of patient's health record. If the operation of PA is updating, then it will destroy itself after completion. If the operation of PA is to retrieve information, then after the information is retrieved, it will migrate back to the clinic's platform where it originated.

The actions of PA in different locations within ABEHRS can be summarized by the following flowcharts, as shown in Figures 2 through 6. Based on the above actions and interactions, the concepts of self-completing, self-regulating and self-organizing can be achieved in our proposed system architecture.

SIMULATION SETUP

In order to demonstrate that the architecture of ABEHRS presented above can achieve the concept of health record self-organization, a simulation approach has been applied to ABEHRS. We chose to use the simulation approach because actual physical prototype will have a greater potential to exhibit unexpected or unplanned system behaviors related to the self-organization processes. At the same time, we examine the global system dynamics for the self-organized processes, so agent population at each medical site can be measured. The parameters of the simulation also help formulate our macroscopic mathematical model of the system. Therefore, the simulation results will be used as a contrasting representation for the comparison to the numerical results we generate from the mathematical model. In this section, the structure and conditions of the simulation are presented, as well as the results and the discussion.

Structure and Conditions

A simple computer environment was setup to simulate the necessary scenario. Several computers were used in our simulation, and they are interconnected via 100Mbps Ethernet. Each computer executed at least one TEEMA platform, and each TEEMA platform represented one specific physical medical site; the number of

Figure 2. Clinic actions flowchart (flowchart representation of the actions of patient agents at the clinic platform)

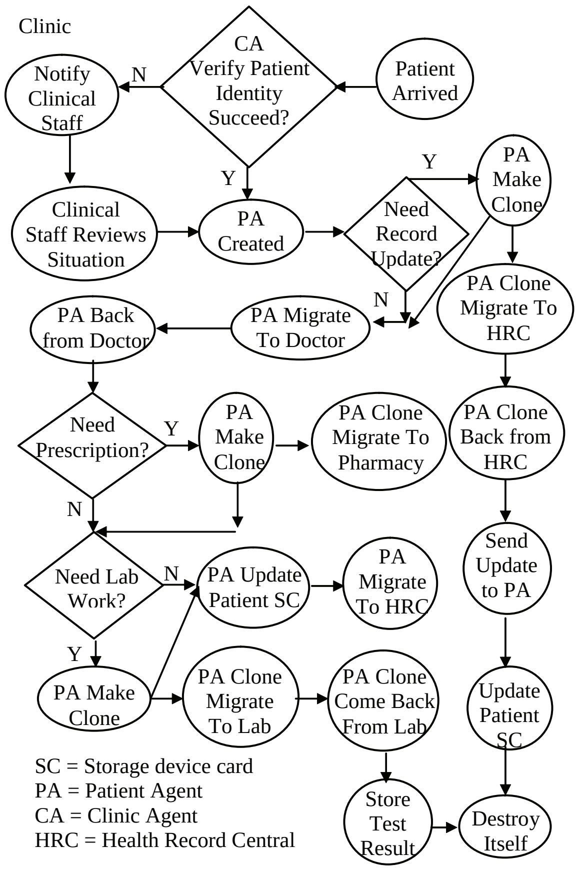


Figure 3. Doctor actions flowchart (flowchart representation of the actions of patient agents at the doctor platform)

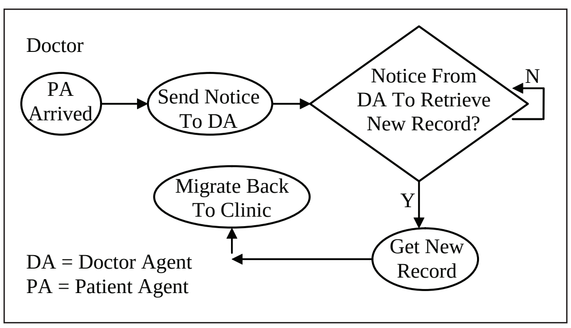


Figure 4. Pharmacy actions flowchart (flowchart representation of the actions of patient agents at the pharmacy platform)

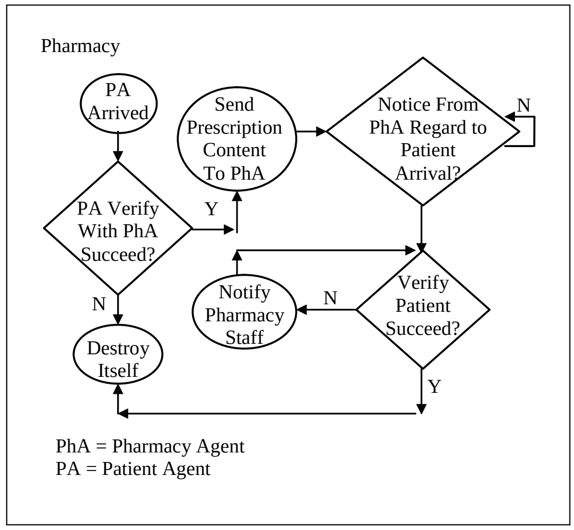


Figure 5. Lab actions flowchart (flowchart representation of the actions of patient agents at the lab platform)

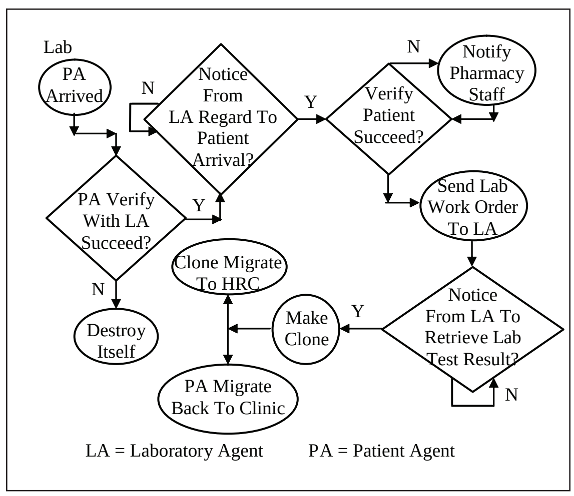
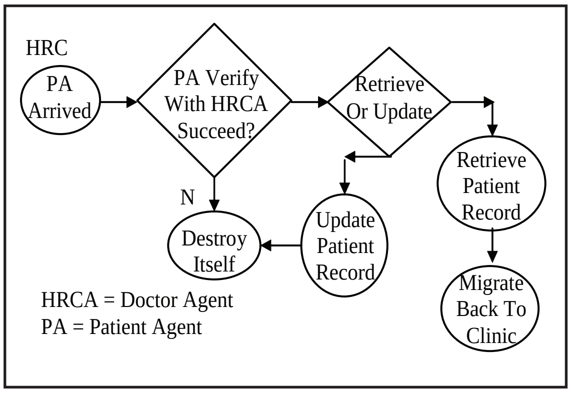


Figure 6. Health record central actions flowchart (flowchart representation of the actions of patient agents at the health record central platform)



platforms used depends on the structure of the simulation. For this simulation run presented in this work, we used: one clinic, two doctors in the clinic, one pharmacy, one laboratory, and one HRC; thus a total of six TEEMA platforms were used, and they are executed on two computers: TEEMA platforms for one clinic and two doctors are executed on the first computer, and TEEMA platforms of one HRC, one lab, and one pharmacy are executed on the second computer.

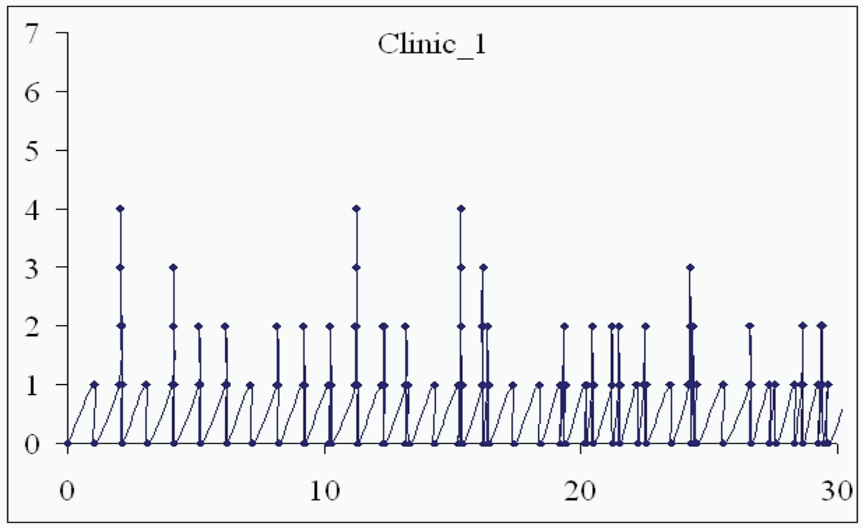
Basic conditions and assumptions used in the simulation are listed next:

1. Doctor evaluation, prescription content, and lab test results are predefined to be fixed-length text. The combination of this content is considered the health record of a patient.
2. Even though the health records are randomly generated, records in the storage device card and records in the HRC are correlated. If the record in the storage device card indicates there is an absent of lab test result, then the record in the HRC must contain the lab test result, and it must be ready for the patient agent to retrieve.
3. There are several random behaviours used in the simulation, and they are represented by different kinds of random numbers. The behaviours are shown below:
   - **Patient preference behavior:** This behavior describes a patient's wish to choose a specific doctor. For simplicity, a uniform distributed random number is used to represent this behavior.
   - **Patient necessitated behavior:** This behavior describes the need for a specific medical action. This includes the need for prescriptions and lab work. A Bernoulli random number was used to describe this type of behavior. Since the need for a prescription/lab work is binary, the chance that a patient will need this type of medical service when he or she visits the clinic is 50/50.
   - **Patient arrival behavior:** This behavior describes the rate of patient arrival at the clinic. For simplification, a constant mean rate of arrival was used and set to one patient arrival at the clinic every minute.
   - **Professional service behavior:** This behavior describes the service time of any medical services provided to the patients. This includes physicians, pharmacists, and lab technician patient processing time. For simplicity, a uniform distributed random numbers between one and five was used to represent this type of behavior in the experiment.

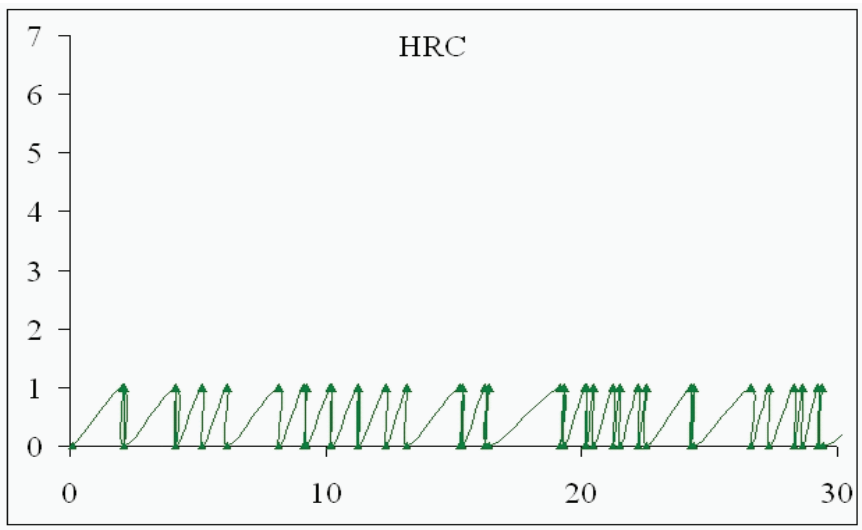
Simulation Result

The results of a typical simulation run are shown in Figure 7. From Figure 7, we see that there are many events occurring in each platform, especially for Clinic_1. These events are the indication of agent actions/interactions at different platforms or medical service sites. To show that ABEHRS

Figure 7a-b. Evolution of agent populations. Typical traces showing movements of patient agents in the ABEHRS between six TEEMA platforms for the first 30 minutes (x-axis is time, y-axis is agent population).

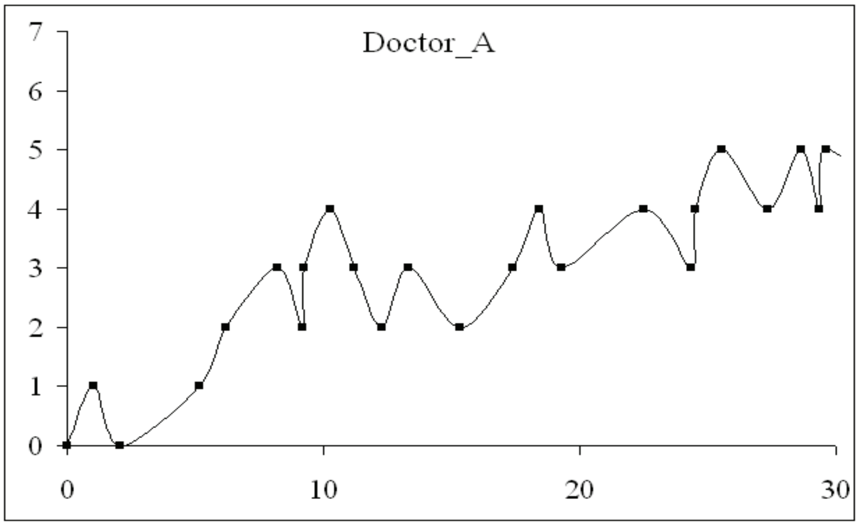


(a) Evolution of Agent Population in Clinic_1 Platform

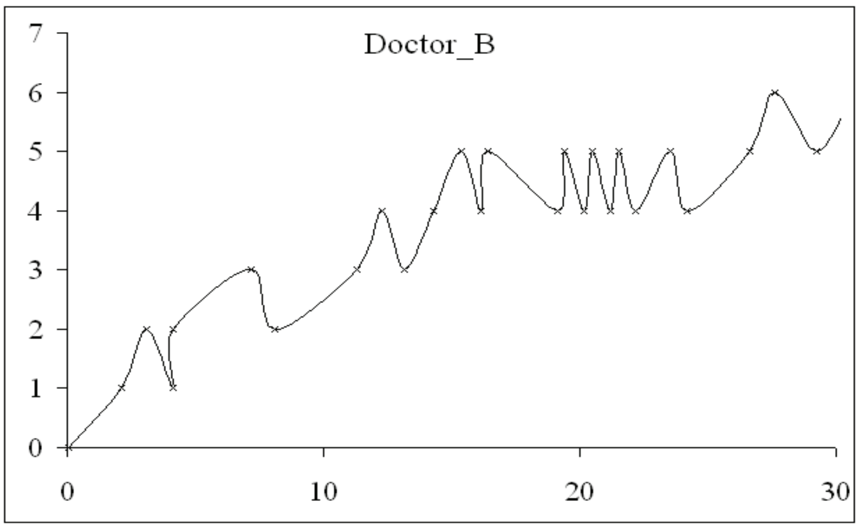


(b) Evolution of Agent Population in HRC Platform

Figure 7c-d. Evolution of agent populations. Typical traces showing movements of patient agents in the ABEHRS between six TEEMA platforms for the first 30 minutes (x-axis is time, y-axis is agent population).



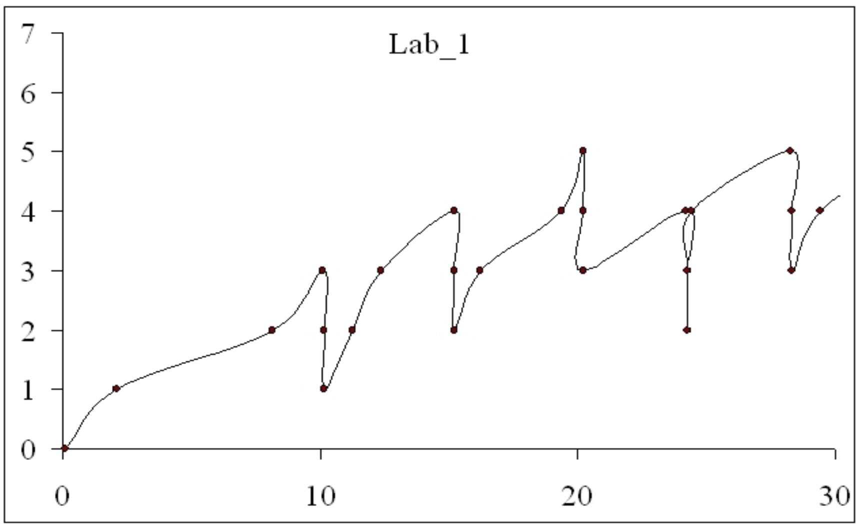
(c) Evolution of Agent Population in Doctor_A Platform



(d) Evolution of Agent Population in Doctor_B Platform

Figure 7e-f. Evolution of agent populations. Typical traces showing movements of patient agents in the ABEHRS between six TEEMA platforms for the first 30 minutes (x-axis is time, y-axis is agent population).

(e) Evolution of Agent Population in Pharmacy_1 Platform



(f) Evolution of Agent Population in Lab_1 Platform

does achieve health record self-organization as defined in the section of self-organized health record, we first examine the health information files in HRC at the end of the simulation and all the self-organization related events occurring during the simulation. By reviewing the files, we are able to see that each patient's health record does contain the newly created record, and it was appended to the complete set of patient history of health information. This indicated that every piece of health record generated at different medical sites did indeed combine into one complete and consistent data set for the patient.

Although the results in Figure 7 do not represent the entire range of behaviours exhibited by the system, it is possible to identify some relevant observations, and they are:

- We notice there are several cluster formations at Doctor_A, Doctor_B, Pharmacy_1 and Lab_1 platforms. It seems that the cluster size of PAs at Doctor_B has the same cluster size of PAs at Doctor_A, which is ~ 5. Cluster formations at the Pharmacy_1 and the Lab_1 platforms are also similar to the cluster sizes at both doctor platforms. Each PA arriving at those medical service platforms waits for an event (for example, patient arrives at the lab) to trigger the action of the next task, and this causes the size of the cluster to grow continuously. Even though there are events (for example, patient picks up his/her prescription) to cause the PAs to either destroy themselves or migrate to other platforms, the occurrences of these events were not high enough to limit the cluster growth due to the arrival of PAs at those platforms. Clearly, we can solve this problem by changing the patient's arrival rate and/or the service rates of these facilities to overcome the increase of the cluster size. Conversely, we can take advantage of this growing cluster size of PAs, since cluster formation is a valuable behaviour in a system where a group of agents can perform a task more effectively than a single agent can. So, a large cluster size means a large number of PAs grouping together, and there could be tremendous gains for group or batch processing from the agent cluster.
- We observe the total number of clones produced by the PAs during simulation period, since an agent-cloning approach is used to replicate agents into a group of agents to perform the self-organizing health record-related tasks, as mentioned in the section on agent cloning. Generally, each PA can clone itself zero, one, two, or four times during the process of self-organization. Maximum clones produced for each PA are four, which is counted as follows: one for HRC update, one for prescription, and two for lab work (the clone created in the lab platform will be included as the clone produced by the original PA). So, for the period of 30 minutes, 30 unique PAs are created (since the patient's arrival rate is ~1 per each minute, PA's creation rate is also ~1 per each minute), and there could be a maximum of 120 clones produced during the simulation period. As stated above for patient necessitates behaviours, if there are 30 PAs, then the number of clones produced for pharmacy and lab should be 15 and 15, respectively. Therefore, the cloning effect can be easily model by a Bernoulli random variable, which in turn we can apply it to our mathematical model. Moreover, by observing the number of clones produced, we are able to observe any overpopulation of PAs at each medical service site.

DYNAMIC MODEL

Using the flowcharts as a guide, we are able to construct a macroscopic model that treats agent population at each medical site as the fundamental unit, hence directly describing the characteristic of the system. The mathematical model captures the dynamics of the self-organization process by examining agent population at each platform. This model contains a set of coupled rate equations that describe how the agent population at each platform evolves over time. The following is our general mathematical model of ABEHRS that contains n number of clinics, m number of doctors in each clinic, x number of pharmacies, y number of laboratories, and one health record central. See Box 1.

The dynamic variables in the model are:

- $N_{Cn}(t)$ is the number of agents in clinic n, where $n \geq 1$
- $N_{CnDm}(t)$ is the number of agents in the doctor m in clinic n, where $m \geq 1$
- $N_{Px}(t)$ is the number of agents in pharmacy x, where $x \geq 1$
- $N_{Ly}(t)$ is the number of agents in laboratory y, where $y \geq 1$
- $N_{HRC}(t)$ is the number of agents in the health record central

And the definitions of the parameters used in the model are:

- λ is the patient's arrival rate at the clinic, which is equivalent to the creation rate of agent
- δ is the rate of agent cloning occur at the clinic platform, which is $\sim \lambda * \beta_A + (\Sigma (1/\tau_{CnDm})*(\beta_B+\beta_C))*$
- τ_{CnDm} is the examination time of a doctor on a patient
- τ_{AvgD} is the average of all τ_{CnDm}
- τ_{Px} is the service time for an individual agent in a pharmacy service site (prescription fill time + prescription pickup time)
- τ_{Ly} is the service time for an individual agent in a lab service site (time for a patient to come to the lab + test result production time).
- τ_{SPC} is the service time for an individual agent to perform necessary task in the clinic platform
- τ_{SPL} is the service time for an individual agent to perform necessary task in the lab platform
- τ_{HRC} is the service time for an individual agent in a HRC

Box 1.

$$\frac{d}{dt}N_{Cn}(t) = \lambda+\delta+\sum_{i=1}^{m} 1/\tau_{CnDm}+\beta_A*\alpha_{CnR}*N_{Cn}(t-\tau_{HRC}) +\beta_C*\alpha_{CnLy}*N_{Cn}(t-\tau_{AvgD}-\tau_{Ly})*\theta(t-\tau_{AvgD}-\tau_{Ly}) - \sum_{i=1}^{m}1/(\tau_{CnDm}+\tau_{SPC}) - \alpha_{CnD}*N_{Cn}(t+\tau_{SPC}) -\beta_A*\alpha_{CnR}*N_{Cn}(t) -\beta_B*\alpha_{CnPx}*N_{Cn}(t-\tau_{AvgD}) - \beta_C*\alpha_{CnLy}*N_{Cn}(t-\tau_{AvgD}) -\beta_A*\alpha_{CnR}*N_{Cn}(t-\tau_{HRC}+\tau_{SPC}) - \beta_C*\alpha_{CnLy}*N_{Cn}(t-\tau_{AvgD}-\tau_{Ly}+\tau_{SPC})* \theta(t-\tau_{AvgD}-\tau_{Ly}) \quad (1)$$

$$\frac{d}{dt}N_{CnDm}(t) = \alpha_{CnD}*N_{Cn}(t)* \beta_{Dm} - 1/\tau_{CnDm} \quad (2)$$

$$\frac{d}{dt}N_{Px}(t) = (\beta_{PPx}* \sum_{i=1}^{n} \alpha_{CiPx}*N_{Ci}(t-\tau_{AvgD})* \beta_B - 1/\tau_{Px})* \theta(t-\tau_{AvgD}) \quad (3)$$

$$\frac{d}{dt}N_{Ly}(t) = (\beta_{PLy}* \sum_{i=1}^{n}\alpha_{CiLy}*N_{Ci}(t-\tau_{AvgD})* \beta_C + 1/(\tau_{Ly}-\tau_{SPL}) - 2/\tau_{Ly})* \theta(t-\tau_{AvgD}) \quad (4)$$

$$\frac{d}{dt}N_{HRC}(t) = \sum_{i=1}^{n}\alpha_{CiR}*N_{Ci}(t)* \beta_A + \theta(t-\tau_{AvgD}-\tau_{Ly}) * \sum_{j=1}^{y} 1/\tau_{Lj} - 1/\tau_{HRC}+ \sum_{j=1}^{n} \sum_{i=1}^{m} 1/(\tau_{CjDi}+\tau_{SPC}) \quad (5)$$

- β_A, β_B, β_C are the probability of a patient who needs an update or a prescription or a lab work, respectively
- β_{Dm} is the probability of a specific doctor being chosen by a patient; it is set to 1/(# of doctors in the clinic), since we assume each doctor is to be chosen equally
- β_{PPx} is the probability of a specific pharmacy chosen by a patient; it is set to 1/(number of pharmacies in the system), since we assume each pharmacy is to be chosen equally
- β_{PLy} is the probability of a specific lab chosen by a patient; it is set to 1/(# of labs in the system) and we assume each lab is to be chosen equally
- α is the transition rates of agents between different medical platforms, for example, α_{CnPx} is the rate at which PAs leave the clinic n platform to the pharmacy x platform. α_{CnR} is the rate at which PAs leave the clinic n platform to the HRC platform
- $\theta_{(t-\tau)}$ is a unit step function to ensure certain variables are zero during $t < \tau$

Instead of providing a detailed explanation of what each team in each equation means, we will only provide a general explanation. For the terms with a minus sign in front, they are describing the patient agents (PAs) that leave the platform for other platforms. The terms having the plus sign in front describe PAs who enter the platform from other platform. In addition, in most of the terms, there is a time delay inside a dynamic variable (for example, $N(t-\tau)$), which illustrates that some of the agents that entered a platform at $(t-\tau)$ time are now exiting this platform and migrating to other platform. For example, $N_{Cn}(t-\tau_{HRC})$ illustrates that there are several patient agents that entered the HRC platform at $(t-\tau_{HRC})$ time and which are now exiting the HRC platform and migrating to the clinic platform.

The number of terms in each equation in the above model is too large and the equations are too complex to be solved numerically. In order to further minimize the number of terms, we assume τ_{SPC} and τ_{SPL} are negligibly small. So, the equations in the above model will change to a general minimized model, as shown in Box 2.

Using the minimized model above, we can represent the same number of platforms as in our simulation, and we have six dynamic variables, $N_{C1}(t)$, $N_{C1D1}(t)$, $N_{C1D2}(t)$, $N_{P1}(t)$, $N_{L1}(t)$, and $N_{HRC}(t)$. The equations in Box 2 will be changed to those in Box 3.

The equations in Box 3 illustrate the self-organized health record process in the sense that each term in the equations is modeling part of the process that are required to achieve self-organized health record.

Numerical Result

This is an initial investigation of the model focused on the self-organized health record by examining the agent population at each medical site. The results have shown that mathematical analysis is not only robust but yields obvious results. For simplicity, we assume the following when solving the equations:

- All α to be uniformly distributed in some space, which is set to 1.
- All β to be a constant value 0.5, except for β_{PP1} and β_{PL1}, which is set to 1.
- τ_{C1D1} and τ_{C1D2} are set to be a constant value of 3, τ_{P1} and τ_{L1} are set to a constant value of 6. These values are the expected value of the uniform distributed random number in our parameters used in the simulation. τ_{HRC} is set to be a constant of 0.787.
- $\lambda = 1$ and $\delta = 7/6$.

We use a built-in function named “dde23” in Matlab to solve the time-delay differential equations in our model. The following is the solution of solving the differential equations (11-16) in our macroscopic model, and the solution is shown in Figure 8.

Box 2.

$$\frac{d}{dt}N_{Cn}(t) = \lambda + \delta - \alpha_{CnD}*N_{Cn}(t) - \beta_A*\alpha_{CnR}*N_{Cn}(t) - \beta_B*\alpha_{CnPx}*N_{Cn}(t-\tau_{AvgD}) - \beta_C*\alpha_{CnLy}*N_{Cn}(t-\tau_{AvgD}) \quad (6)$$

$$\frac{d}{dt}N_{CnDm}(t) = \alpha_{CnD}*N_{Cn}(t)*\beta_{Dm} - 1/\tau_{CnDm} \quad (7)$$

$$\frac{d}{dt}N_{Px}(t) = (\beta_{PPx}*\sum_{i=1}^{n}\alpha_{CiPx}*N_{Ci}(t-\tau_{AvgD})*\beta_B - 1/\tau_{Px}) * \theta(t-\tau_{AvgD}) \quad (8)$$

$$\frac{d}{dt}N_{Ly}(t) = (\beta_{PLy}*\sum_{i=1}^{n}\alpha_{CiLy}*N_{Ci}(t-\tau_{AvgD})*\beta_C - 1/\tau_{Ly}) * \theta(t-\tau_{AvgD}) \quad (9)$$

$$\frac{d}{dt}N_{HRC}(t) = \sum_{i=1}^{n}\alpha_{CiR}*N_{Ci}(t)*\beta_A + \theta(t-\tau_{AvgD}-\tau_{Ly}) * \sum_{j=1}^{y}1/\tau_{Lj} - 1/\tau_{HRC} + \sum_{j=1}^{n}\sum_{i=1}^{m}1/(\tau_{CjDi}) \quad (10)$$

Box 3.

$$\frac{d}{dt}N_{C1}(t) = \lambda+\delta-\alpha_{C1D}*N_{C1}(t)-\beta_A*\alpha_{C1R}*N_{C1}(t)-\beta_B*\alpha_{C1P1}*N_{C1}(t-\tau_{AvgD})-\beta_C*\alpha_{C1L1}*N_{C1}(t-\tau_{AvgD}) \quad (11)$$

$$\frac{d}{dt}N_{C1D1}(t) = \alpha_{C1D}*N_{C1}(t)*\beta_{D1} - 1/\tau_{C1D1} \quad (12)$$

$$\frac{d}{dt}N_{C1D2}(t) = \alpha_{C1D}*N_{C1}(t)*\beta_{D2} - 1/\tau_{C1D2} \quad (13)$$

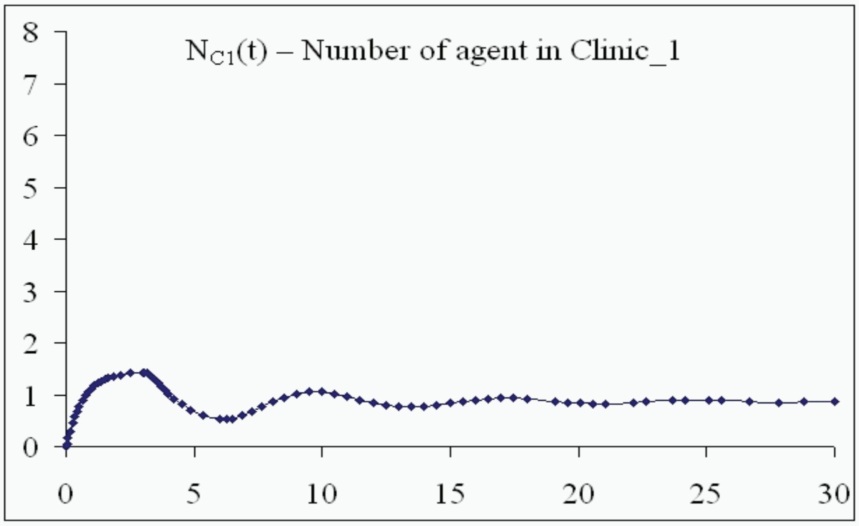
$$\frac{d}{dt}N_{P1}(t) = (\beta_{PP1}*\alpha_{C1P1}*N_{C1}(t-\tau_{AvgD})*\beta_B - 1/\tau_{P1}) * \theta(t-\tau_{AvgD}) \quad (14)$$

$$\frac{d}{dt}N_{L1}(t) = (\beta_{PL1}*\alpha_{C1L1}*N_{C1}(t-\tau_{AvgD})*\beta_C - 1/\tau_{L1}) * \theta(t-\tau_{AvgD}) \quad (15)$$

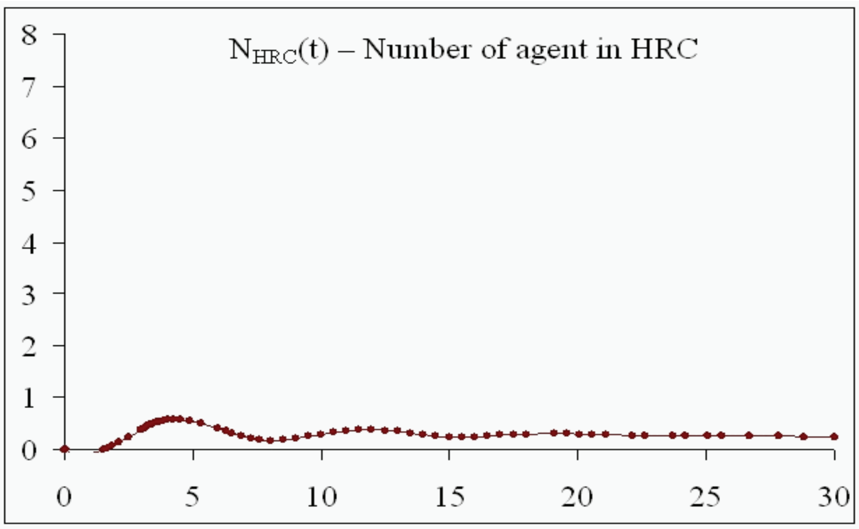
$$\frac{d}{dt}N_{HRC}(t) = \alpha_{C1R}*N_{C1}(t)*\beta_A + \theta(t-\tau_{AvgD}-\tau_{L1})* 1/\tau_{L1} + 1/\tau_{C1D1} + 1/\tau_{C1D2} - 1/\tau_{HRC} \quad (16)$$

From Figure 8, we see that there is a small oscillatory effect that occurs in the value of $N_{C1}(t)$ at the beginning of the result. The reason may be that there are many PAs being created, and they are cloning themselves at the same time, causing an increase in $N_{C1}(t)$. As the PAs leave the Clinic_1 platform to either Doctor_A or Doctor_B platform, this situation causes $N_{C1}(t)$ to decrease. Thus there are forces increasing and decreasing the agent population in the Clinic_1 platform. As time goes by, the number of agents in each platform becomes stable or predictable in the form of an upward straight line. This suggests that the system is adjusting itself to the changes of PA population in each platform. To show that this solution is reasonable, we calculate the number of agents in each platform at the end of 30 minutes. If the patient arrival rate is 1 patient/minute, then there will be 30 PAs created in the 30-minute simulation. The agents will divide themselves up between the Doctor_A and Doctor_B platforms, since there are only two doctors they can choose from. As indicated above, each doctor's examination time is three minutes on average. So, at the end of 30 minutes, Doctor A and Doctor B platform should have five PA left, on average, which is only slightly different from what we observe in Figure 8. Similar calculations for the pharmacy and lab platform indicate that there should be six PAs in Lab_1 and six PAs in Pharmacy_1. It is very similar to what we observe in Figure. 8.

Figure 8a-b. Numerical results of the macroscopic model of ABEHRS (x-axis is time, y-axis is agent population)

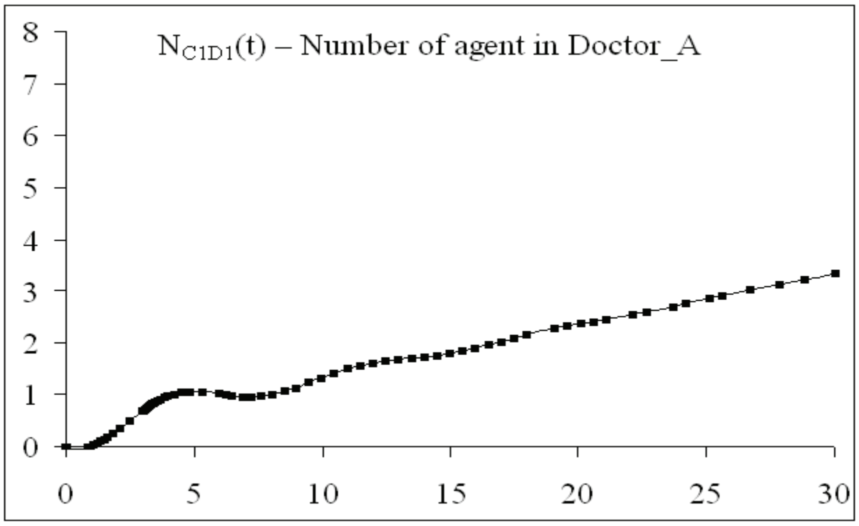


(a) Numerical Solution for NC1(t)

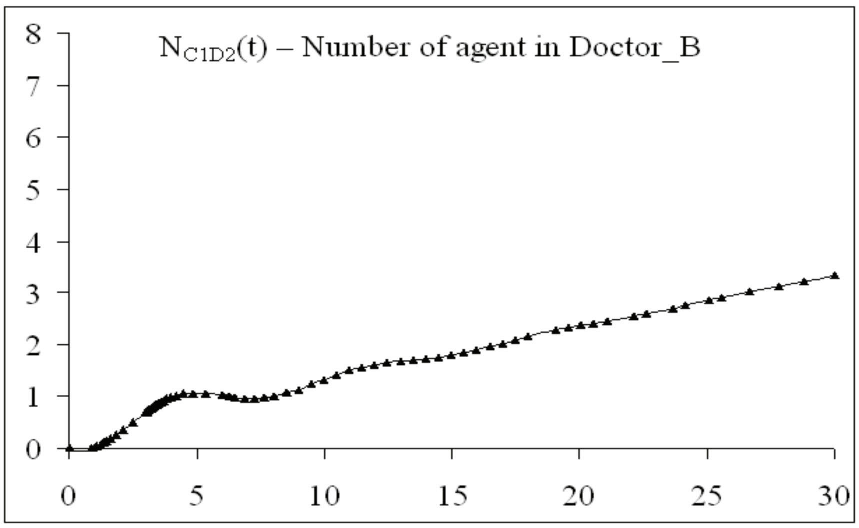


(b) Numerical Solution for NHRC(t)

Figure 8c-d. Numerical results of the macroscopic model of ABEHRS (x-axis is time, y-axis is agent population)

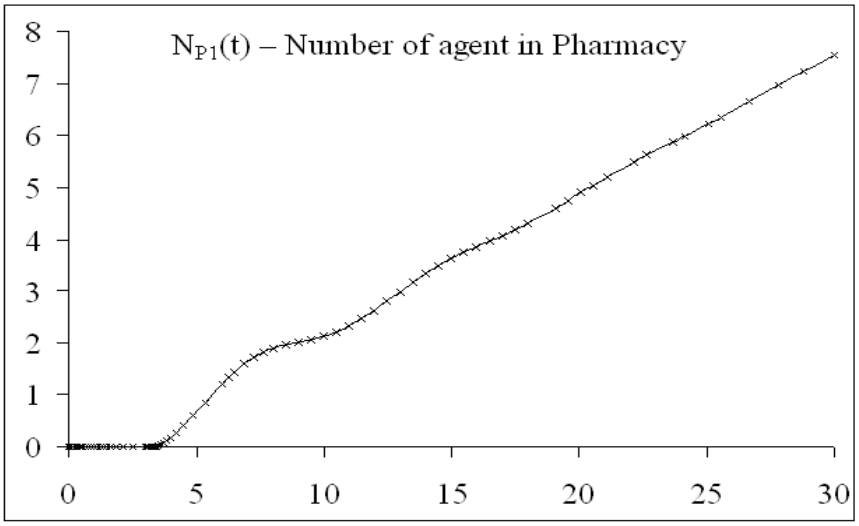


(c) Numerical Solution for NC1D1(t)

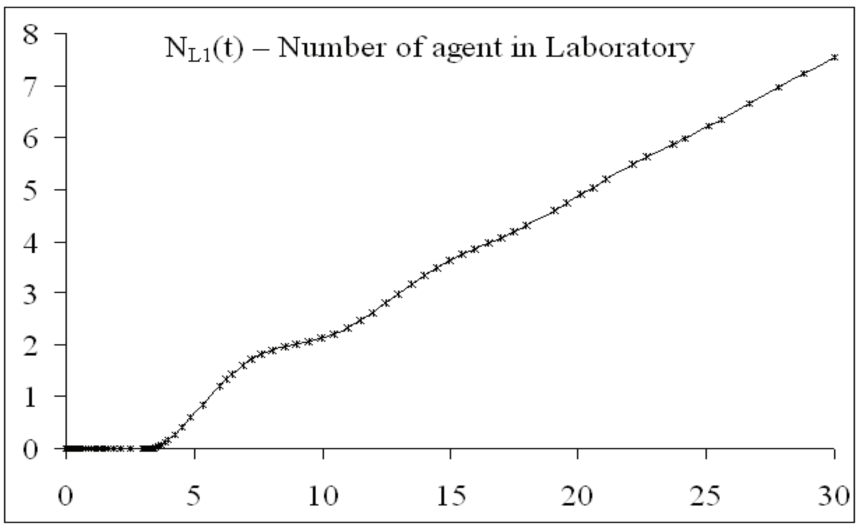


(d) Numerical Solution for NC1D2(t)

Figure 8e-f. Numerical results of the macroscopic model of ABEHRS (x-axis is time, y-axis is agent population)



(e) Numerical Solution for NP1(t)



(f) Numerical Solution for NL1(t)

The above numerical results are subject to our initial assumption of setting all the parameters to some constants that we used in the simulation. Many parameters in the model are free parameters, such as λ, β_A, β_B, β_c, τ_{C1D1}, τ_{C1D2}, τ_{P1}, τ_{L1}. These free parameters are tied to the process of self-organized health records in a sense that they are affecting the time it takes an individual health record to recover from a chaotic state and back to a well-ordered state. For example, if τL1 is large, then the time it takes for the test result to be appended to the HRC database will also be long. Therefore, the values for these parameters are up to the agent designer to choose. Note that all the parameter values used in the model must be positive.

RESULTS COMPARISON BETWEEN MODELS

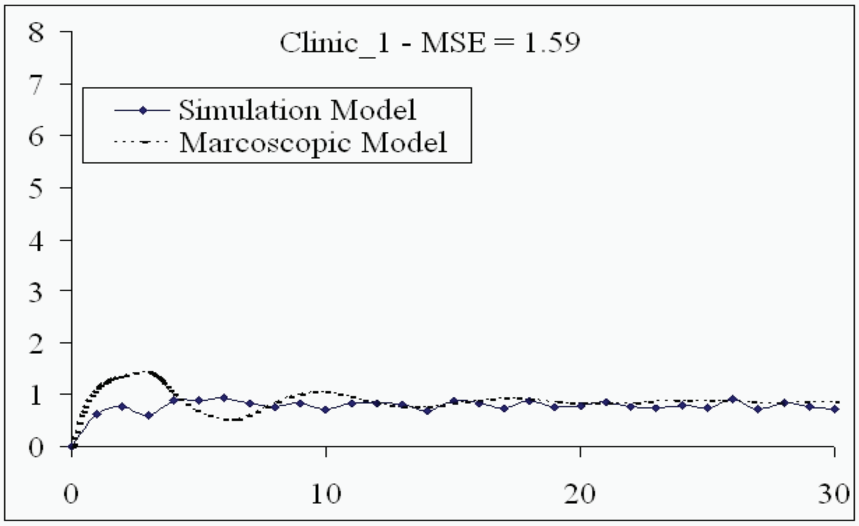
Since the numerical results give us the dynamics of average quantities of agent population at each medical site, the simulation results must also be formulated as an average. Therefore, we executed 30 simulation runs and the average of these simulation results, and then compare them to the numerical results obtained from the macroscopic mathematical model. The mean square error (MSE) is calculated to indicate the difference between the simulation results and the numerical results. We proposed that an acceptable value for the MSE for our study would be 10 (for now); however, this value can be set arbitrarily by agent designers. As shown in Figure 9, we can clearly see that both models for Clinic_1 and HRC platforms have settled to a steady state, and the MSE for Clinic_1 is 1.59, and MSE for HRC is 0.58, and this indicates our parameter values for these sites in our macroscopic model are acceptable. The clusters of PAs (or agent population) in both doctor platforms increase linearly, and in both simulation and numerical results. However, the MSE between simulation and numerical results for Doctor_A is 24.63, and MSE for Doctor B is 20.37. This indicates our parameter values for the both doctor's equations in our macroscopic model need to re-adjust. The cluster of PAs in pharmacy and lab platforms are also exhibited linear aviation. We have only MSE = 6.63 for the pharmacy and MSE = 9.58 for the lab, and this indicates our parameter values for these sites in our macroscopic model are barely acceptable.

FUTURE TRENDS

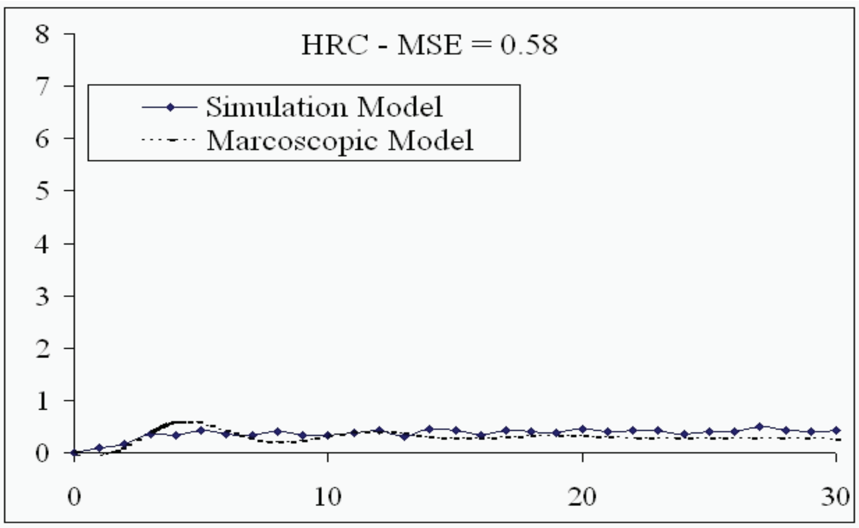
Mobile agent technology has been used to solve most of the problems in health care domain for more than a decade, and its reputation and popularity have only increased. There are numerous studies on applying mobile agent architecture to different fields in the health care domain. For example, medical image retrieval (Liu, Martens, Paranjape, & Benedicenti, 2001), organ and tissue transplant and management (Moreno, 2003), health services scheduling and collaboration (Kirn, 2002), and optimization and management of clinical processes (Heine et al., 2005) all examine the application of mobile agent technology. We believe that in the near future, the mobile agent paradigm will become a dominated force in telemedicine application and in the telemedicine industry. This trend will likely keep increasing until new network computing technology that is even more flexible than mobile agent technology is developed.

There are several challenges, not mentioned in the section of The Electronic Health Record but presented in Moreno (2003), that also need to be addressed before we can expect useful multi-agent systems to the real health care domain. One such challenge is the slow in adoption of new information technology by medical professionals. Many medical professionals feel a loss of control when the information technology becomes too central to the access and retrieval of medical data (Moore, 1996). This is more critical, especially

Figure 9a-b. Results comparison between models (x-axis is time, y-axis is number of agent)

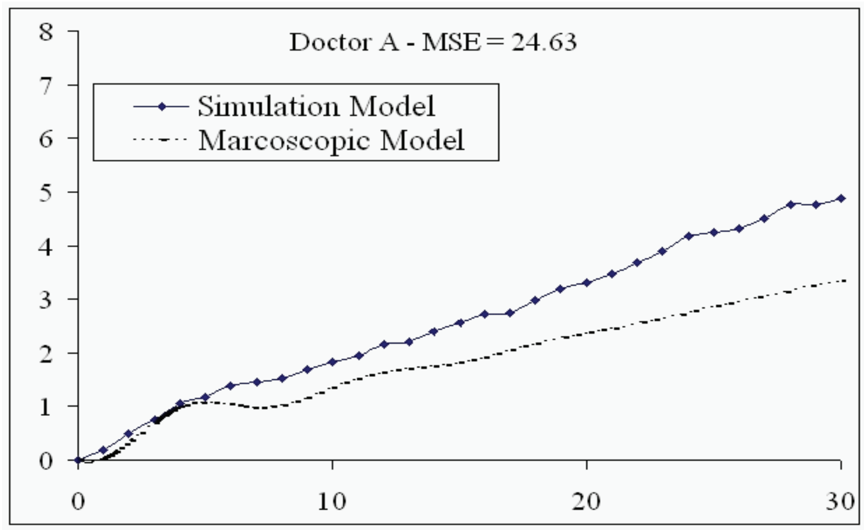


(a) Comparison Between Models for Clinic_1

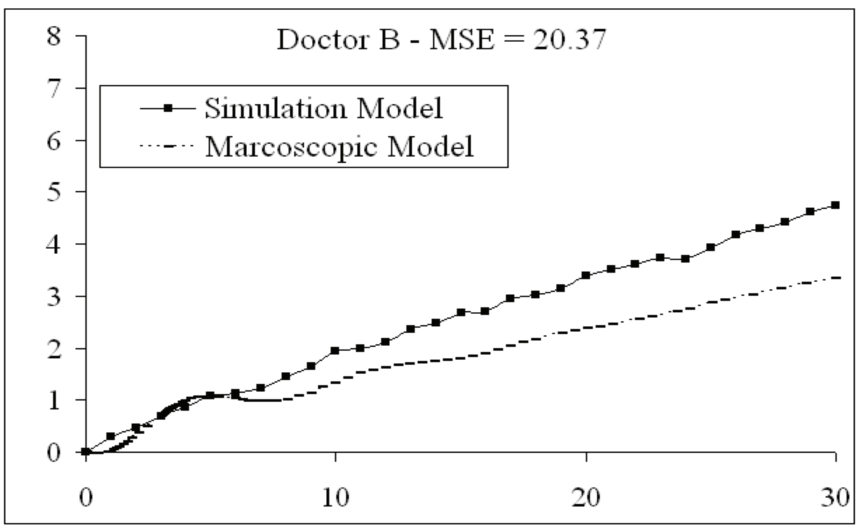


(b) Comparison Between Models for HRC

Figure 9c-d. Results comparison between models (x-axis is time, y-axis is number of agent)

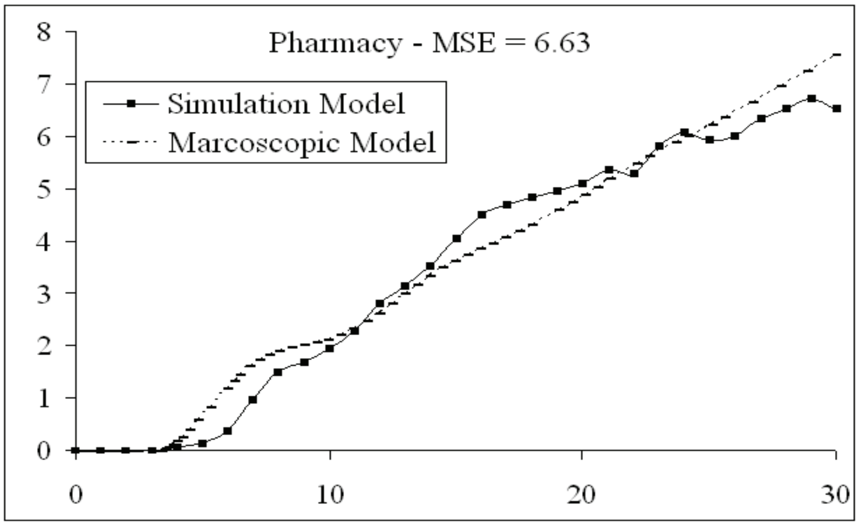


(c) Comparison Between Models for Doctor_A

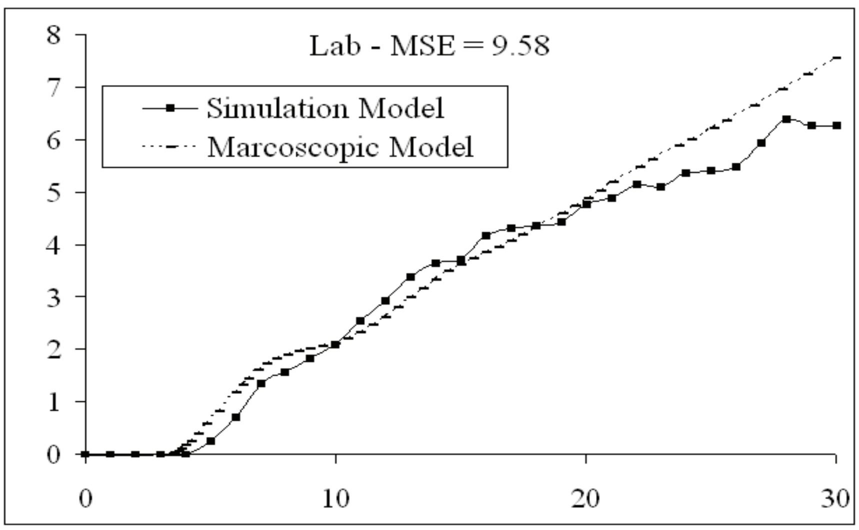


(d) Comparison Between Models for Doctor_B

Figure 9e-f. Results comparison between models (x-axis is time, y-axis is number of agent)



(e) Comparison Between Models for Pharmacy_1



(f) Comparison Between Models for Lab_1

for the technologies that contain a high level of automation, like mobile agents. Before we can see an increase in the rate of adoption by medical professionals, we need to see an acceptance by the general public. Thus, if we see mobile agent enabled cellular phones or PDAs and services (such as agent-controlled appliances, agent-based stock exchange) in our normal day lives, as shown in Kirn (2002), then only will we see acceptance in the medical community.

Intelligent swarms have given the agent designers a new perspective of developing multi-agent system. If an agent system consists of thousands of simple autonomous and distributed agents, then their collective power arising from the interactions between the agents is tremendous. The intrinsic properties of the swarm system can be very significant. We believe that swarm intelligence and multi-agent systems will have a strong impact of solving problems in health care domain.

CONCLUSION

In this chapter, we have presented a general framework and the basic concepts behind ABEHRS. These concepts encapsulated the fundamental principles of self-completing, self-regulating, and self-organizing data sets by using the strength of agent mobility and automation. Based on the principles of the swarm paradigm, we have designed and implemented an ABEHRS model using the TEEMA system. Initial simulation results from modeling ABEHRS have been presented, and from these results it is shown that the fundamental goals of ABEHRS are achieved. It also shown that cluster formations occur in specific sites of the health care system (for example, doctor, pharmacy, and lab platforms). The existence of these clusters, and their growth rate, identifies new challenges and new opportunities for the health care system. These cluster formations can identify valuable collective behaviour in the system, as also seen in Lerman and Galstyan (2001).

In addition to the design and simulation of ABEHRS, we also presented a macroscopic mathematical model of the ABEHRS, which contains a set of differential equations. These equations describe the global dynamics of mobile agents in the ABEHRS. Our numerical analysis of these equations strongly matches our simulation results. However, both the numerical and simulation results are subject to the initial conditions and assumptions, which can significantly alter the behaviour of the system. Further work using different initial conditions and assumptions should be used to help better understand the macroscopic behaviour of the ABEHRS. Also, we may want to look for some parallel to the current health record system to demonstrate that there is a relationship between what we propose in this work and the actual health system in use today.

Finally, our current simulation and numerical work on the ABEHRS does not show any extraordinary or catastrophic collective swarm behaviour. The global simulation dynamics of ABEHRS compares well to the mathematical model. However, our work is only very preliminary, and we have not attempted to push the system to any extremes. This type of testing and analysis is critical in order to ensure that the system will always be stable.

REFERENCES

Agassounon, W., & Martinoli, A. (2002). A macroscopic model of an aggregation simulation using embodied agents in groups of time-varying sizes. *Proceedings of the 2002 Institute of Electrical and Electronics Engineers Systems, Man and Cybernetics Conference* (pp. 250-255). Retrieved April 3, 2005, from http://swis.epfl.ch/backup_caltech/coro/People/alcherio/papers/SMC02.pdf

Bonabeau, E., Dorigo, M., & Theraulaz, G. (1999). *Swarm intelligence: From natural to artificial systems*. New York: Oxford University Press.

Camazine S., Deneubourg, J. L., Franks, N., Sneyd, J., Theraulaz, G., & Bonabeau, E. (2001). *Self-organization in biological systems*. Princeton, NJ: Princeton University Press.

Charles, A. A., Menezes, R., & Tolksdorf, R. (2004). On the implementation of SwarmLinda. *Proceedings of the 42nd Association for Computing Machinery Southeast Regional Conference* (pp. 297-298). Retrieved July 15, 2006, from Association for Computing Machinery Digital Library Database.

Crow, L., & Shadbolt, N. (2001). Extracting focused knowledge from the Semantic Web. *International Journal of Human Computer Studies, 54*(1), 155-184. Retrieved December 27, 2003, from Association for Computing Machinery Digital Library Database.

Gray, R., Kotz, D., Nog, S., Rus, D., & Cybenko, G. (1997). Mobile agents: The next generation in distributed computing. *Proceedings of the Second Aizu International Symposium on Parallel Algorithms and Architectures Synthesis* (pp. 8-24). Retrieved November 15, 2002, from Institute of Electrical and Electronics Engineers Digital Library Database.

Gibbs, C. (2000). *TEEMA reference guide version 1.0*. Regina, Saskatchewan, Canada: TRLabs.

Grimson, J., Grimson, W., & Hasselbring, W. (2000). The SI challenge in health care. *Communications of the Association for Computing Machinery, 43*(6), 49-55. Retrieved May 27, 2003, from Association for Computing Machinery Digital Library Database.

Heine, C., Herrier, R., & Kirn, S. (2005). ADAPT@AGENT.HOSPITAL: Agent-based optimization and management of clinical processes. *International Journal of Intelligent Information Technologies, 1*(1). Retrieved July 10, 2006, from http://wi2.uni-hohenheim.de/team/documents/public/heine/%5BHeHK2005%5D.pdf

Huston T. (2001). Security issues for implementation of e-medical records. *Communications of the Association for Computing Machinery, 44*(9), 89-94. Retrieved January 9, 2005, from Association for Computing Machinery Digital Library Database.

Kadrovach, B. A., & Lamont, G. B. (2002). A particle swarm model for swarm-based networked sensor systems. In *Proceedings of the 2002 ACM Symposium on Applied Computing* (pp. 918-924). Retrieved June 19, 2004, from Association for Computing Machinery Digital Library Database.

Kirn, S. (2002). *Ubiquitous healthcare: The OnkoNet mobile agents architecture.* Paper presented at the 2002 Mobile Computing in Medicine, Heidelberg, Germany. Retrieved September 23, 2004, from http://www.old.netobjectdays.org/pdf/02/papers/node/0278.pdf

Lerman, K., & Galstyan, A. (2001). *A methodology for mathematical analysis of multi-agent systems* (Tech. Rep. No. ISI-TR-529). University of California Information Sciences Institute. Retrieved November 20, 2004, from http://www.isi.edu/%7Elerman/papers/isitr529.pdf

Lerman, K., & Shehory, O. (2000). Coalition formation for large-scale electronic markets. *Proceedings of the Fourth International Conference on Multiagent Systems* (pp. 167-175). Retrieved November 20, 2004, from http://www.isi.edu/%7Elerman/papers/LerShe99.pdf

Liu, A., Martens, R., Paranjape, R., & Benedicenti, L. (2001). Mobile multi-agent system for medical image retrieval. *Proceedings of the 2001 Institute of Electrical and Electronics Engineers Canadian Conference on Electrical and Computer Engineering Conference,* Ontario, Canada. Retrieved October 15, 2003, from Institute of Electrical and Electronics Engineers Digital Library Database.

Martens, R. (2001). *TEEMA TRLabs execution environment for mobile agents*. Regina, Saskatchewan, Canada: TRLabs.

Moore, B. (1996). Acceptance of information technology by health care professionals. *Proceedings of the 1996 Symposium on Computers and the Quality of Life* (pp. 57-60). Retrieved August 9, 2003, from Association for Computing Machinery Digital Library Database.

Moreno, A. (2003). *Medical applications of multi-agent systems.* Paper presented at the 2003 Intelligent and Adaptive Systems in Medicine Workshop, Praha, Czech Republic. Retrieved September 21, 2004, from http://cyber.felk.cvut.cz/EUNITE03-BIO/pdf/Moreno.pdf

Nealon, J., & Moreno, A. (2003). *Agent-based applications in health care.* Paper presented at the 2003 EU-LAT Workshop, Mexico. Retrieved September 21, 2004, from http://www.etse.urv.es/recerca/banzai/toni/MAS/papers.html

Silverman, B., Andonyadis, C., & Morales, A. (1998). Web-based health care agents; The case of reminders and todos, too (R2Do2). *Artificial Intelligence in Medicine, 14*(3), 295-316. Retrieved May 29, 2003, from Science Direct Database.

Smith, K., Paranjape, R., & Benedicenti, L. (2001). Agent behavior and agent models in unregulated markets. *Association for Computing Machinery SIGAPP Applied Computing Review, 9*(3), 2-12. Retrieved May 27, 2003, from Association for Computing Machinery Digital Library Database.

Sugawara, K., & Sano, M. (1997). Cooperative acceleration of task performance: Foraging behavior of interacting multi-robots system. *Physica D, 100*(3-4), 343-354. Retrieved November 27, 2004, from Association for Computing Machinery Digital Library Database.

Tecchia, F., Loscos, C., Conroy, R., & Chrysanthou, Y. (2001). Agent behaviour simulator (ABS): A platform for urban behaviour development. *Proceedings of the First International Game Technology Conference and Idea Expo (GTEC'01) in co-operation with ACM SIGGRAPH and EUROGRAPHICS* (pp. 17-21). Retrieved January 9, 2005, from http://www.equator.ac.uk/var/uploads/TecchiaGTEC2001.pdf

Truszkowski, W., Hinchey, M., Rash, J., & Rouff, C. (2004). NASA's swarm missions: The challenge of building autonomous software. *Institute of Electrical and Electronics Engineers Information Technology Professional, 6*(5), 47-52. Retrieved August 5, 2006, from Institute of Electrical and Electronics Engineers Digital Library Database.

Walczak, S. (2003). A multiagent architecture for developing medical information retrieval agents. *Journal of Medical Systems, 27*(5), 479-498. Retrieved July 29, 2005, from Association for Computing Machinery Digital Library Database.

Wang, Y., Liu, J., & Jin, X. (2003). Modeling agent-based load balancing with time delays. *Proceedings of the 2003 Institute of Electrical and Electronics Engineers/WIC International Conference on Intelligent Agent Technology*, 189+. Retrieved December 15, 2004, from Institute of Electrical and Electronics Engineers Digital Library Database.

Wu, J. (2003). Towards a decentralized search architecture for the Web and P2P systems. *Proceedings of the 2003 Adaptive Hypermedia and Adaptive Web-Based Systems Workshop*. Retrieved August 5, 2006, from http://wwwis.win.tue.nl/ah2003/proceedings/paper18.pdf

This work was previously published in Architectural Design of Multi-Agent Systems: Technologies and Techniques, edited by H. Lin, pp. 305-334, copyright 2007 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 2.20
An Information Technology Architecture for Drug Effectiveness Reporting and Post-Marketing Surveillance

Amar Gupta
The University of Arizona, USA

Ray Woosley
The Critical Path Institute, USA

Igor Crk
The University of Arizona, USA

Surendra Sarnikar
The University of Arizona, USA

ABSTRACT

Adverse drug events impose a large cost on the society in terms of lives and health care costs. In this article, we propose an information technology architecture for enabling the monitoring of adverse drug events in an outpatient setting as a part of the post marketing surveillance program. The proposed system architecture enables the development of a Web-based drug effectiveness reporting and monitoring system that builds on previous studies demonstrating the feasibility of a system in which community pharmacists identify and report adverse drug events. We define the key technical requirements of such a monitoring and reporting system, identify the critical factors that influence the successful implementation and use of the system, and propose information technology solutions that satisfy these requirements.

INTRODUCTION

Adverse drug reactions have been estimated to result in more than 2.1 million injuries and 100,000 deaths each year in the U.S. alone (Lazarou, Pomeranz, & Corey, 1998). The annual economic cost of adverse drug events is estimated to be more than $75 billion (Johnson & Bootman, 1995). Mitigating the impact of adverse drug events requires the implementation of a comprehensive mechanism for monitoring and detecting adverse drug events. Such a mechanism can save lives and reduce health care costs.

Reliable detection adverse drug events is a difficult problem. Although some adverse drug reactions are detected early on during clinical trials, serious adverse drug effects can still go undetected during this phase due to the practical limitations associated with the size and duration of the clinical trials. Recent examples of such cases include Rofecoxib and Cerivastatin (Fontanarosa, Rennie, & DeAngelis, 2004). The FDA monitors for adverse drug events in the post-marketing phase through the MedWatch program (www.fda.gov/MedWatch/report.htm). The MedWatch program, which is mainly a voluntary reporting program, suffers from several limitations, the most critical of which are the under-reporting of adverse events and the lack of a denominator reflecting magnitude of exposure.

In a 1996 article titled "The Clinical Impact of Adverse Event Reporting," the FDA estimated that only 1% of the adverse drug events are reported through the MedWatch program (Food and Drug Administration [FDA], 1996, p. 5). An alternative mechanism for detection of adverse drug events is the use of longitudinal medical records and hospitalization records. However, the availability of data from such records has been limited and obtaining longitudinal medical records is an expensive and time-consuming process. In addition, the extraction of meaningful conclusions from such data is difficult due to data integrity, heterogeneity, and missing data problems.

Several information technology-based solutions have been suggested to help monitor and reduce the adverse drug event problem. Most of the proposed solutions and studies conducted have been limited to inpatient hospital settings. Although a major part of drug dispensing and medications takes place in an outpatient setting, there is limited literature on the detection of adverse drug events in an outpatient setting. In this article, our focus is on methods for the detection of adverse drug events in an outpatient setting and in the post marketing phase using a Web-based reporting system. Specifically, our focus is developing the IT architecture for enabling a large-scale data collection mechanism to support the detection and quantification of previously unrecognized side effects and drug interactions for drugs, especially those newly introduced into the market. We propose an IT architecture for enabling a Web-based reporting and surveillance solution called the Drug Effectiveness Reporting and Monitoring System (DERMS). The DERMS system is based on a community pharmacy-based safety network and proposes the participation of community pharmacies for the collection of clinical response and adverse drug event information from patients. We describe the information technology architecture that forms the supporting infrastructure for the surveillance system and discuss the requirements and success factors necessary for successful implementation of the system.

This article is structured as follows. In the second section, we describe the post marketing effectiveness and safety surveillance program and discuss the limitations of the system in its current form. In the third section, we review previous literature discussing technological solutions to the adverse event detection problem. We briefly describe the Drug Effectiveness Reporting and Monitoring System and propose the enabling IT architecture in the fourth section, and discuss the success factors for its implementation in the fifth section. We discuss the limitations of the system in the sixth section, and make concluding observations in the seventh section.

POST-MARKETING SURVEILLANCE

An effective surveillance process that follows the introduction of a new drug into the market requires the efficient flow of information among the different affected entities including patients, drug companies, the FDA, and healthcare professionals such as doctors, nurses and pharmacists. This should include information on drug usage, patient exposure, interactions, adverse effects, and treatment outcomes. At present, the primary mechanism of disseminating information from the drug companies and FDA are through sales representatives, visiting lecture series, press releases. At present, the primary mechanism of disseminating information from the drug companies and FDA are through press releases, information services, and pharmacy databases that enable timely dissemination of information on drug interactions and labeling information.

While the FDA and drug companies are able to use broadcast mechanisms and information services exist for disseminating information to physicians, pharmacists, and patients, there are currently no widely implemented mechanism for the reverse flow of information on adverse events and medication and treatment outcomes from patients and health care practitioners to the FDA on an ongoing basis. This channel is weak, and is currently confined to MedWatch and other limited mechanisms.

The FDA conducts postapproval monitoring through postmarketing surveillance programs such as the MedWatch system. The MedWatch system relies on voluntary submission of reports by patients and health care providers. In this program, patients and health care providers can submit an adverse event report via several mechanisms including an online report form, fax, phone, and mail. The Medwatch reports, along with adverse event information reported by pharmaceutical companies, are stored in the Adverse Event Reporting System (AERS) database. The database is publicly available for download by clinical reviewers and researchers on a quarterly basis for analyzing drug interactions and monitoring drug safety. (Center for Drug Evaluation and Research [CDER], 2005).

While the MedWatch program has been successful at identifying critical side effects that

Figure 1. Information flows in the post approval phase

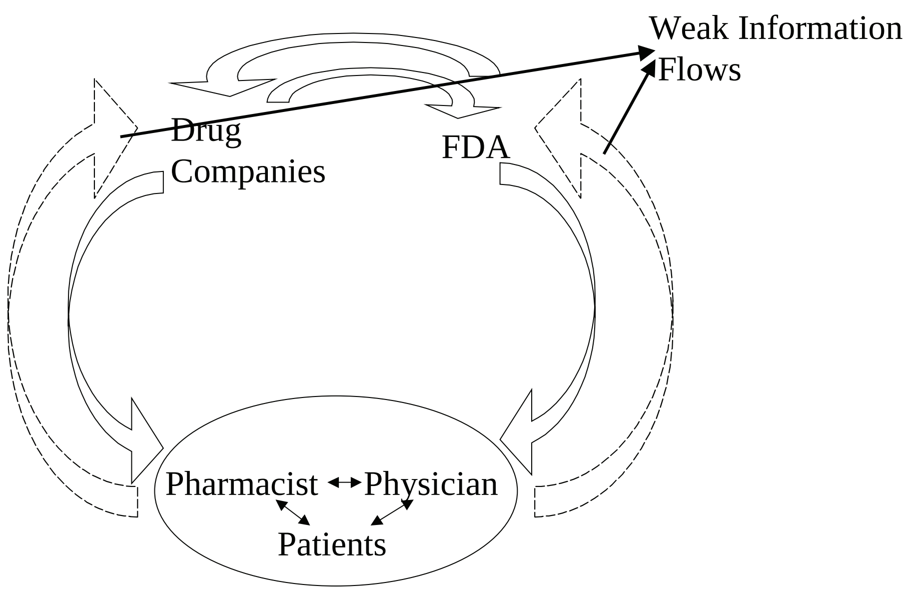


exhibit in the early stages after drug marketing, it suffers from several major limitations that prevent the faster detection of the adverse drug effects. For example, although 19 drugs were withdrawn from the market between 1997 and 2007 based on MedWatch data, it took an average of 6.6 years post introduction ofthe drugs into the market to identify their adverse effects and remove these drugs from the market. In addition, the system is far less effectiveat identifying adverse effects that result from prolonged administration of drugs (Brewer & Colditz, 1999; U.S. Department of Health and Human Services, 1999).

The limitations of MedWatch system include the poor quality of submitted reports, duplicate reporting of events, under-reporting of adverse events, and the absence of the denominator or baseline information required to make meaningful conclusions from the data (Fontanarosa et al., 2004). In addition, the detection of adverse effects resulting from prolonged use of drugs requires the collection longitudinal medical records, which are not readily available to the MedWatch system. Longitudinal medical records are necessary for the detection of adverse effects that manifest late in the chronic administration of a drug, such as in the case of Vioxx, where the increased risk of heart attacks and strokes on prolonged use was not detected by the MedWatch system, but rather through controlled clinical studies.

The MedWatch system is also deficient in providing background data a on number of patients being administered a particular medication. Background rates of information, such as the number of events per number of patients exposed, is essential for the scientific evaluation of adverse event data. Other factors contributing to the noise and biased nature of MedWatch data include increases in adverse event reports in response to media publicity and "dear health care professional" letters. Given the various limitations with current adverse event reporting mechanisms, there is a need for a comprehensive adverse event data collection mechanism that can provide better quality data and serve as an early alert system for newly introduced drugs.

PREVIOUS WORK

Several studies have been conducted analyzing the use of information technology (IT) in managing Adverse Drug Events (ADE). Literature in this area focuses on the detection of adverse events using computer-based mechanisms and the prevention of ADE using IT tools. Computerization and the use of information technology tools for automating health care workflow have resulted in significant improvements in health care delivery and in the prevention of adverse drug events (Bates et al., 1999; Bates et al., 2001; Evans et al., 1992). Computerization and health care information technology systems, such as computerized order entry and clinical decision support systems, have led to a significant reduction in medical errors and improvements in quality of care (Bates et al., 2001).

Surveillance mechanisms for the detection of adverse drug events can be classified as outpatient-based monitoring mechanisms and inpatient-based monitoring mechanisms. Bates, Evans, Murff, Stetson, Pizziferri, and Hripcsak (2003) studied the effectiveness of various information technology tools in detecting adverse events in inpatient and outpatient settings. They determine that information technology tools that analyze administrative data recorded using ICD9 codes are of limited value in identifying adverse drug events, while rule-based detection mechanisms that rely on laboratory test results and occurences of antidote use are able to detect a significantly larger portion of the adverse events. Another finding of the study is the need for natural language processing tools to process free text data for the detection of adverse events. A significant portion of the patient related information such as visit notes, admission notes, progress notes, consultation notes, and nursing notes are stored in the form of free text. Although rule-based mechanisms are able to identify a significant portion of the adverse events, they still underperform chart review-based methods for adverse event detection. This is primarily due to the inability of rule-based mechanisms to identify symptom changes, which are mostly recorded in free text form (Classen, Pestotnik, Evans, & Burke, 1991).

In outpatient care, free text processing tools greatly outperform rule-based mechanisms that rely on ICD-9 codes for the detection of adverse events. Honigman, Lee, Rothschild, Light et al. (2001) reported that code-based mechanism were able to detect only 3% of the adverse events when applied to outpatient data, while free text processing mechanism were able to identify 91% of the adverse events. Anderson, Jay, Anderson, and Hunt (2002) present results from a simulation study designed to analyze the effect of information technology in reducing adverse events. Their primary focus is on the use of information technology tools to reduce prescription errors by automating the prescription workflow using electronic means and the prevention of adverse events by verifying prescriptions against a database of known drug interactions. A detailed review of various methodologies for the detection of adverse events is given in Murff, Patel, Hripcsak, and Bates (2003).

Although several systems have been developed for the detection of adverse drug events given various patient data, the reporting and collection of adverse drug event information itself has not been extensively investigated. Moreover, most of the proposed systems are limited to inpatient settings and single organizations. There is relatively limited literature analyzing the use of information technology for large scale adverse event reporting in an outpatient setting. A study by Gandhi et al. (2000) reports the incidence of adverse drug events in outpatient care to be common and that most such events are not documented in the medical records. A majority of the events is preventable and proper monitoring for symptoms, response, and adequate communication between outpatients and providers can prevent most of the adverse drug events (Tejal et al., 2003).

A series of studies have been conducted over the past few decades to evaluate alternative mechanisms for collection and reporting of adverse drug events in an outpatient setting. Fisher, Bryant, Solovitz, & Kluge (1987) conducted a study to analyze the effectiveness of post-marketing surveillance using outpatient adverse drug event reports. Based on their studies, they conclude that outpatient-based post-marketing surveillance programs that rely on patient initiated reports can be used to complement existing physician-based surveillance systems. Fisher and Bryant (1990) observed that patients are correctly able to differentiate adverse drug events from other adverse clinical events under certain conditions. They observed. They observe that the discrimination between adverse drug events from other adverse clinical events was better when the reporting was initiated by a staff member and the reporting was spontaneous as opposed to an interviewer probed systematic enquiry. Data from patient drug attributions has been observed to be consistent with alternative monitoring methods, such as physician assessments and epidemiological data, and can also be used to improve the discriminatory power of such methods (Fisher, Bryant, Kent, & Davis, 1994). In addition to the Fisher et al. studies, a recent study by Cohen, Kimmel et al. (2005) analyzed the effect of interventions by pharmacists in a community pharmacy setting. The study showed a considerable reduction in adverse events through an audit of discharged patients and a subsequent 9-month follow-up.

A community pharmacist-based outpatient post marketing surveillance system has several potential uses such as early detection of adverse drug reactions, discovery of new therapeutic benefits of the newly introduced drugs (Fisher & Bryant, 1992), and comparison between alternative medications (Fisher, Bryant, & Kent, 1993; Fisher, Kent, & Bryant, 1995). However, the previous studies were limited to short time period and did not explore the use of emerging information technology to leverage the surveillance and monitoring mechanism.

The Critical Path Institute (C-Path, 2005) has proposed a community pharmacy-based surveillance model that is characterized by the following aspects: (1) the data collection is set in an outpatient setting and involves community pharmacies, which are visited by patients more frequently than hospitals, (2) The community pharmacy-based model focuses on pharmacists and pharmacy technicians to collect large-scale data on adverse events and drug effectiveness, and (3) The model is designed to collect baseline information and information on background rates to help conduct rigorous data analysis.

In this article, we propose a Web-based information system, the Drug Effectiveness Reporting and Monitoring System (DERMS). The DERMS system was one of the models developed for consideration toward satisfying the requirements of the Community Pharmacy Safety Network (CPSN) developed by the Critical Path Institute. Although it is currently not a part of the CPSN, the DERMS system can be adapted to serve as a general pharmacy-based patient safety system.

DRUG EFFECTIVENESS REPORTING AND MONITORING SYSTEM

In this section, we give a brief overview of the Drug Effectiveness Reporting and Monitoring System, previously described in Gupta el al. (2007), describe the key processes of the DERMS system and propose a system architecture for supporting the key processes to be implemented in the DERMS system.

Overview

The key requirements of the DERMS system are derived from the community-pharmacy based model for post-marketing surveillance proposed by the Critcal Path Institute. The community pharmacy based model includes a large-scale data collection mechanism that involves pharmacists and pharmacy technicians in identifying and collecting of adverse event information. After medical evaluation, pharmacists and pharmacy technicians usually constitute the first point of contact with patients as outpatients. Hence, they can potentially collect and maintain evolving historical information on the patient's medication history. Such history would include comprehensive information on the various types of medications taken by the patient, along with the corresponding duration of use for each medication. Such records can serve as an alternative source of information for evaluating the long-term effects of clinical medicines. The perceived direct and indirect benefits of such as system include the following: (1) the creation of longitudinal medical records by integrating patient medication history with baseline and periodically collected follow-up information on the patient's medical condition, and (2) Faster detection of adverse events using a systematic monitoring procedure implemented at the point of medication dispensing and (3) assessment of comparative effectiveness.

Key Processes

The drug effectiveness reporting and monitoring system is characterized by three key processes that include the data collection process, surveillance and monitoring process, and surveillance administration process. We describe each of these processes in detail below.

Surveillance Administration Process

The surveillance administration process basically captures the key tasks of the agency responsible for administering the surveillance mechanisms and the infrastructure. The surveillance administration process involves the identification of newly introduced drugs that need to be monitored. It also includes the identification of appropriate data items that need to be captured and the design and development of questionnaires for eliciting and capturing adverse event information. The questionnaires developed in this process are used in the data collection process, which we describe next.

Data Collection Process

The data collection process is illustrated in Figure 2. The process is initiated when a patient visits a pharmacy to fill a prescription. If the prescribed drug has been selected by a surveillance administrator for surveillance, the pharmacist proceeds to collect further information about the patient with the patient's consent. For patients who are not already in the system, a basic patient information questionnaire is used to collect information on patient demographics.

A baseline information questionnaire is administered at the start of a medication to collect basic information about the patient's health status before begining the new medication. At the time of each refill, a follow-up questionnaire is administered to the patient to record the patient's health status and query for any adverse drug effects. In the case of severe adverse drug effects, a MedWatch report is filed by the pharmacist. Each of the questionnaires administered to the patient was designed during the surveillance administration process.

The questionnaires vary based on the type of the drug being monitored. In order to design these questionnaires, the research team studied previous examples and the work of others. This included examination of post-marketing surveillance programs of FDA, as well as of the allied research and monitoring endeavors. Further, the research team studied questionnaires designed by researchers of the Arizona Center for Research Therapeutics (AzCERT). Based on different needs, five types of forms were delineated. These were: the basic patient information form; the baseline information questionnaire; the routine follow-up questionnaire; the special follow-up questionnaire; and the adverse event reporting form. The special follow-up questionnaire is used for medications that are known to have potential harmful side effects usually occurring after a certain period has elapsed. A screenshot of the data collection process within the DERMS system is shown in Figure 3.

Surveillance and Monitoring Process

While the data collection process is executed by the pharmacist, the surveillance and monitoring process is primarily executed by research and quality improvement organizations. An overview of the process is given in Figure 4. In this pro-

Figure 2. Data collection process

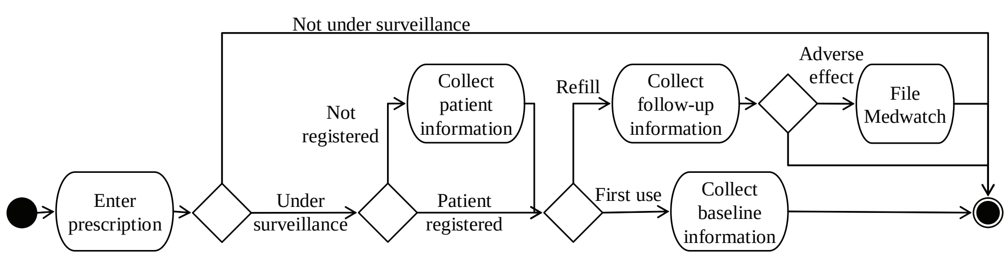
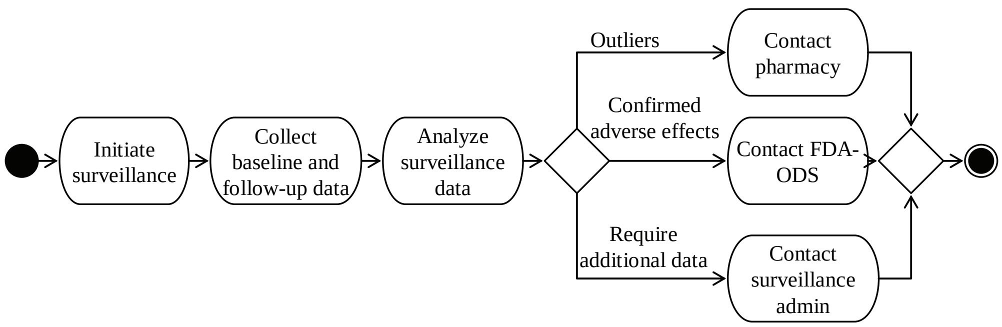


Figure 3. Data collection forms

Figure 4. Surveillance process



cess, the data collected during the data collection process is analyzed by researchers to identify possible drug interactions and signals suggesting serious side effects.

Following the analysis, three possible actions are supported by the DERMS system. In the case of suspect data points or outliers, a researcher can contact the concerned pharmacy for follow-up information. In the case of confirmed adverse effects, a report can be sent to the FDA's office of drug safety. If the researcher requires the collection of additional information through follow-up questionnaires, the surveillance administration can be contacted for modifications or the design of specialized follow-up questionnaires to be administered during the surveillance process. A screenshot from the surveillance process within the DERMS system is shown in Figure 5.

Figure 5. Surveillance visualization screenshot

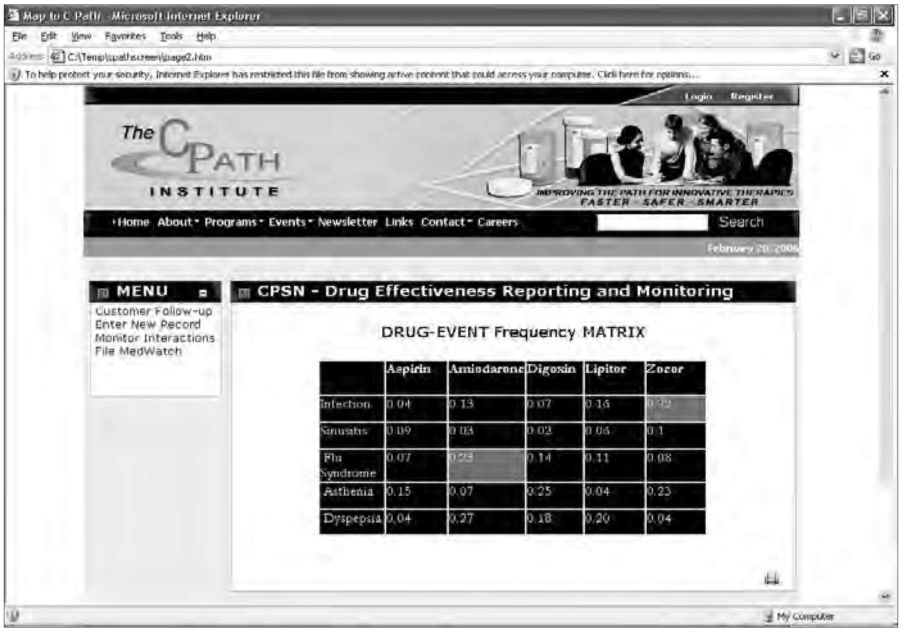
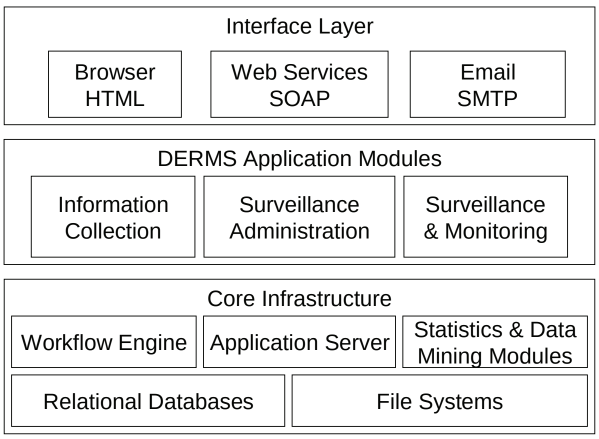


Figure 6. System architecture



System Architecture

In order to support the above mentioned key processes, we propose a three layer system architecture, as illustrated in Figure 6. It consists of a core infrastructure layer, an application layer, and an interface layer. The core IT infrastructure supporting the drug effectiveness reporting and monitoring system consists of a centralized relational database and file system for storing the surveillance data and associated documents. The core infrastructure also includes an application server, a workflow engine, and a statistics and data-mining module that help execute the business logic implemented in the DERMS modules. Three application modules corresponding to the key processes supported by the DERMS system are included in the application layer. They implement the business logic and processes that support the data collection, administration, and surveillance and monitoring mechanisms.

The interface layer consists of a HTML interface accessible through a Web browser, a Web service interface, and an e-mail interface. The HTML interface is the primary Web-based interface used by the pharmacists to execute the data collection process. The Web service interface can be used to interface with pharmacy information systems to directly retrieve data from pharmacy systems. The Web service interface can also be by researchers to interface with statistical and analysis software. The e-mail interface is used for communication between various entities involved in the data collection, administration, and surveillance and monitoring processes.

Critical Success Factors

The proposed large-scale post marketing surveillance system involves the participation of multiple stakeholders and is influenced by several factors that determine its adoption and successful implementation. We reviewed literature in the area of event reporting systems (Barach & Small, 2000), post-marketing surveillance methods (Fisher et al., 1987), pharmaco-epidemiological studies involving the participation of community pharmacists (Farris, Kumbera, Halterman, & Fang, 2002; Oh, McCombs, Cheng, & Johnson, 2002; Schommer, Pederson, Doucette, Gaither, & Mott, 2002; Weinberger, 2002), and information technology adoption (Menachemi, Burke, & Brooks, 2004) to determine the key factors that influence the successful implementation of such a system.

Barach and Small (2000) draw lessons from an analysis of various nonmedical critical event reporting systems to prescribe a set of guidelines for the design of medical event reporting systems. They identify six different factors as being critical to the successful adoption and high quality of a medical event reporting system: "immunity (as far as practical); confidentiality or data de-identification (making data untraceable to caregivers, patients, institutions, time); independent outsourcing of report collection and analysis by peer experts; rapid meaningful feedback to reporters and all interested parties; ease of reporting; and sustained leadership support." We analyze each of these factors and identify the critical elements of the community pharmacy-based surveillance model to derive the technological requirements of the DERMS model.

Immunity: A key barrier to adoption of an incident reporting system is the fear of reprisal or a lack of trust for an individual and fear of litigation for organizations. As such, immunity to the greatest extent possible is important for successful adoption of an incident reporting system. In the context of the community pharmacy-based adverse event reporting system, this translates into immunity for pharmacists and the community pharmacy participating in the reporting program. From a technical requirements point of view, enabling immunity requires the use of mechanism that provide confidentiality of users, anonymity to the pharmacists and pharmacies, and mechanisms that prevent the traceability of actions by unauthorized users.

Confidentiality: Confidentiality of data is an important element for a successful reporting system. For health care applications, confidentiality implies protecting the privacy rights of the patients by de-identifying patient data. However, de-identification of data sometimes leads to duplication of records. Therefore, data management mechanisms that enable de-duplication of records and the identification of unique individual records while preserving patient privacy need to be developed. Access control mechanisms and data encryption technologies need to be provided in order to ensure the security of data and prevent unauthorized use of the data.

Evaluation: Barach and Small (2000) report that an independent collection of reports and analysis by peer evaluation is an important factor influencing the quality of an incident report-

Table 1. Critical success factors

Success Factor	Implementation	Possible IT Solutions
Immunity	Anonymity of reporters, participants	Data Encryption and de-identification. Access control.
Independent Reporting and Evaluation	Involvement of Community Pharmacies, Quality Improvement Organizations and administrative entity.	Workflow systems, Mediators and Web Services for heterogeneous data and application integration. Access control and relational views.
Confidentiality	Patient Privacy, Data de-identification	Data element identification and Probabilistic de-duplication algorithms.
Feedback	Summary reports and information	Reporting modules and data validation and verification algorithms
Ease of Reporting	Minimal reporting time, Capture of key and minimal data elements	Assistive technologies and intuitive user interfaces.
Leadership	Surveillance Administration, Dashboards, Real-time monitoring of key metrics, problem detection and communication capability.	Communication modules, Monitoring and reporting modules.
Workload Minimization	Financial Incentives, Workload distribution between Pharmacists and Pharmacy Technicians	Integration with pharmacy systems to minimize computer entry.

ing system. In the community pharmacy-based approach, this is achieved by outsourcing the data evaluation to regional quality improvement organizations, data collection to community pharmacies, and monitoring and overview to an independent administrative entity. Providing the above features would require a scalable and flexible mechanism that would enable multiple diverse entities to seamlessly exchange data by integrating heterogeneous applications and data sources and at the same time provide privacy and data security, and prevent unauthorized access. Recent developments in information technology, such as workflow system, Web services, service-oriented architecture (SOA), and grid computing can provide a successful implementation to support the independent collection and evaluation features of the community pharmacy-based system.

Feedback: Feedback to incident reporters and all participating stakeholders is necessary for successful adoption and implementation of an incident reporting system. The reporting and data analysis modules, along with workflow and communication tools, can be used to provide meaningful feedback to the interested participating users.

Reporting: Two major factors need to be considered when designing the data collection process: the ease of reporting the data and the quality of the data being collected. Previous studies (Bates et al., 2003) have shown that typical hospital incident reports and ICD-9 based reporting mechanisms are inadequate for detecting adverse drug events. Reporting mechanisms need to be customized for each drug and drug combinations to collect relevant symptomatic information. While the data fields determine the quality of the data being collected, the design of the report affects the adoption of the reporting system. Complexity and amount of time spent reporting is a major barrier to large scale adoption of a reporting system. The reporting system needs to be designed such that it leverages the users familiarity with other computer-based systems, and minimizes the amount of data to be manually entered and the overall reporting time.

Leadership: Continued leadership is necessary to maintain and manage an incident reporting system and to effectively respond to changing needs. Monitoring and communication capability are key to enabling effective leadership. In the DERMS system, an administrative module is provided to initiate and monitor the surveillance process. Graphical tools, integrated e-mail, and messaging mechanisms can be used to provide this functionality.

Workload Minimization: A key barrier to adoption of the system is the addition to workload because of increased reporting responsibilities. Time and motion studies indicate that pharmacist spend around 6-7% of their time in computer entry activities (Murray, Loos, Tu, Eckert, Zhou, & Tierney, 1998). A study by Oh et al. (2002) estimates that pharmacists need to spend an additional 3 minutes of time for patient consultation and adverse drug effect monitoring. As long as the additional computer order entry time is minimal, resistance to adoption should be minimal.

The introduction of additional workload is a key problem, especially in locations with shortages in pharmacists. However, the following mechanisms can be considered to alleviate the problem. First, the additional workload can be distributed between a pharmacist and a pharmacy technician, such that the computer entry activities are handled by a pharmacy technician while the activities related to the elicitation and identification of adverse events are delegated to a pharmacist. Second, depending on resource constraints, the surveillance mechanism can be limited to patients who are prescribed certain newly introduced drugs, thereby lowering the additional workload.

Incentives: Previous studies have indicated that the provision of financial incentives has had a positive effect on patient counseling activities of pharmacists, resulting in reduced adverse events (Farris et al., 2002). As such, financial incentives can serve an additional factor in promoting the adoption of a community pharmacy-based system. In addition to financial incentives, job satisfaction is a key driver in increased pharmacist involvement in patient counseling and drug therapy reviews. A study by Schommer (2002) concludes that pharmacists prefer to spend more time on patient consultation and drug use management instead of medication dispensing and business management.

Factors Influencing Pharmacy Participation

The successful adoption and continued use of the proposed system by community pharmacies and pharmacists is dependent on several factors. In order to promote the successful adoption of the new surveillance system, the features of the system also need to be aligned with the interests of the community pharmacy and the professional interests of the pharmacists. Based on previous studies on IT adoption, we hypothesize that while organizational buy-in is necessary for initial adoption of a new system, its continued use is dependent on the perceived usefulness of the system and its alignment toward the skill and professional interests of the pharmacists.

Several operational factors also need to be considered for the successful implementation of the proposed surveillance system. For example, an article describing randomized control trials conducted to evaluate the effectiveness of pharmaceutical care programs in Indianapolis (Weinberger, 2002) highlights several operational difficulties that occur in programs involving community pharmacies. The study, initiated at a Revco pharmacy chain, analyzed new pharmaceutical care programs aimed at giving the pharmacists a greater role in providing the patients with better health care. The data were initially transmitted from Revco to the Indianapolis Network for Patient Care (INPC) for purposes of consolidation and analysis. As this experimental study progressed, CVS acquired Revco in a corporate acquisition. Apart from the problems caused by differences in computer systems of the two organizations, there were problems created by major differences in their management policies. For example, CVS required the patients to give their categorical affirmative response before patient data could be utilized in any manner. In order to address this new requirement, a decision was made to offer a sum of $60 as incentive to patients who were willing to let their data be used for purposes of this experimental study. With this incentive, 21% of the patients responded to CVS, with five-sevenths of them agreeing to let the data be used, and the balance two-sevenths declining the offer. In order to increase the response rate, CVS personnel initiated follow-up efforts. Finally, one-fifths of the persons originally contacted agreed to accept the offer of $60 in lieu of data to be utilized for the experimental study.

The above experience emphasizes four critical success factors for achieving progress in this area. First, senior management must accept the need for such studies and be prepared to explicitly support the endeavor through its entire lifestyle; without such close involvement, the effort will fail. Second, pharmacists should view this function as an integral part of their job of dispensing drugs and interacting with patients on issues related to drugs; in order to make this scenario feasible, financial incentives may need to be provided to pharmacists. Third, the policies of major pharmacy chains vary significantly from each other; and discussions need to occur among them in order to generate consensus on this critical issue that impacts human lives. Fourth, new mechanisms need to be developed to share relevant chains across otherwise competing entities in a manner that meets applicable guidelines for information, security, and safety, while simultaneously ensuring that the risks to human lives is minimized.

Limitations

At the beginning of this article, we had identified some of the weaknesses of the current MedWatch system. The concept demonstration prototype described in this article mitigates some of the problems, but several of them still remain and need further research and attention. The areas requiring further attention are discussed in the following paragraphs.

First, the concept demonstration prototype system studied the relevant issues in one city only (Tucson, Arizona). The automated assimilation of the information from diverse information systems, each characterized by its own design and significantly different from others, will require use of advanced concepts from the realm of integration of heterogeneous information systems. Similarly, state-of-the-art ideas related to data mining and knowledge discovery will need to be employed. The scalability of the concept demonstration prototype needs to be examined in detail in order to evaluate the potential feasibility of utilizing the proposed approach over an extended geographic area.

Second, the prototype system tracks drugs that are provided to customers across the counters at retail outlets of major pharmaceutical chains. If this concept is extended to smaller chains and individual shops, it will need to deal with still greater variety of legacy hardware and software. The problem is further complicated by the fact that patients now acquire drugs by mail, both from outlets in the U.S. and abroad, using Web-based and telephone-based mechanisms to place the concerned purchase orders for drugs. No effective mechanism currently exists for tracking the purchase of such drugs.

Third, our approach lacks the ability to track samples of drugs that have been provided by the medical physician to the patient. Such dispensing of drugs by the physicians may need to be monitored, especially for drugs that have been introduced in the recent past. The same Web-based interface could be used to enter the requisite information by personnel in the physician's office. Also, some pharmaceutical companies now provide magnetic cards that can be redeemed at pharmacies for samples of drugs. If this new concept is used for all samples, this limitation will be overcome in terms of monitoring of drug samples.

Fourth, our proposed system relies heavily on pharmacists in terms of their expertise and goodwill in terms of talking with the patients, eliciting requisite information from them, and entering the same using the Web-based interface.

Fifth, patients currently obtain drugs from multiple pharmacy outlets that may belong to different chains. Without a common identifier, it is difficult to track that the medicines were indeed purchased for use by the same patient. The most obvious identifier would be the social security number in the U.S. However, current regulations and concerns for patient privacy prevent such usage. New options need to be explored.

CONCLUSION

In this article, we propose an IT architecture that forms the enabling infrastructure for a new post-marketing surveillance tool for newly introduced drugs called the Drug Effectiveness Reporting and Monitoring System. We briefly define the key characteristics of the DERMS system and propose a system architecture for supporting the key processes implemented in the DERMS system. We then analyze the critical success factors for the DERMS-based post marketing surveillance mechanism and identify supporting IT solutions.

Future work in this area involves further investigation of the critical success factors and development of instruments to validate the hypothesized critical success factors. In addition, large-scale implementation of such as system requires further investigation of the privacy and security issues related to data collection, storage, and sharing processes.

REFERENCES

Anderson, J.G., Jay, S.J., Anderson, M., & Hunt, T.J. (2002). Evaluating the capability of information technology to prevent adverse drug events: A computer simulation approach. *Journal of American Medical Informatics Association*, *9*, 479-490.

Barach, P., & Small, S. (2000). Reporting and preventing medical mishaps: Lessons from non-medical near miss reporting systems. *British Medical Journal*, *320*, 759-763.

Bates, D.W., Cohen, M., Leape, L.L. et al. (2001). Reducing the frequency of errors in medicine using information technology. *Journal of American Medical Informatics Association, 8*, 299-308.

Bates, D.W., Evans, R.S., Murff, H., Stetson, P.D., Pizziferri, L., & Hripcsak, G. (2003). Detecting adverse events using information technology. *Journal of American Medical Informatics Association, 10*, 115-28.

Brewer, T., & Colditz, G.A. (1999). Postmarketing surveillance and adverse drug reactions: Current perspectives and future needs. *Journal of the American Medical Association*, *281*, 824-829.

Center for Drug Evaluation and Research. Adverse event reporting system. Retrieved March 15, 2007, from http://www.fda.gov/cder/aers/default.htm.

Classen, D.C., Pestotnik, S.L., Evans, R.S., & Burke, J.P. (1991). Computerized surveillance of adverse drug events in hospital patients. *Journal American Medical Association*, *266*, 2847-2851.

Cohen, M.M., Kimmel, N.L., et al. (2005). Medication safety program reduces adverse drug events in a community hospital. *Qual Saf Health Care, 14*, 169-174.

The Critical Path Institute. (2005, Spring). C-Path Newsletter.

Evans, R., Pestotnik, S., & Classen, D. et al. (1992). Prevention of adverse drug events through computerized surveillance. In *Proceedings of the Annu Symposium Comput Appl Med Care* (pp. 437-41).

Farris, K.B., Kumbera, P., Halterman, T., & Fang, G. (2002). Outcomes-based pharmacist reimbursement: Reimbursing pharmacists for cognitive services. *J Managed Care Pharm*., *8*(5), 383-93.

Fisher, S., & Bryant, S.G. (1990). Postmarketing surveillance: Accuracy of patient drug attribution judgments. *Clin Pharmacol Ther, 48*, 102-107.

Fisher, S., & Bryant, S.G. (1992). Postmarketing surveillance of adverse drug reactions: Patient self-monitoring. *J Am Board Fam Pract, 5*, 17-25.

Fisher, S., Bryant, S.G., & Kent, T.A. (1993). Postmarketing surveillance by patient self-monitoring: Trazodone versus fluoxetine. *J Clin Psychopharmacol, 13*, 235-242.

Fisher, S., Bryant, S.G., Kent, T.A., & Davis, J.E. (1994). Patient drug attributions and post-marketing surveillance. *Pharmacotherapy, 14*, 202-209.

Fisher, S., Bryant, S.G., Solovitz, B.L., & Kluge, R.M. (1987). Patient-initiated postmarketing surveillance: A validation study. *J Clin Pharmacol, 27*, 843-854.

Fisher, S., Kent, T.A., & Bryant, S.G. (1995). Postmarketing surveillance by patient self-monitoring: Preliminary data for sertraline versus fluoxetine. *J Clin Psychiatry, 56*, 288-296.

Fontanarosa, P., Rennie, D., & DeAngelis, C. (2004). Postmarketing surveillance: Lack of vigilance, lack of trust. *Journal of American Medical Association*, *292*(21).

Food and Drug Administration. (1996, October). The clinical impact of adverse event reporting.

Gandhi, T.K., Burstin, H.R., Cook, E.F., Puopolo, A.L., Haas, J.S., Brennan, T.A., & Bates, D.W. (2000). Drug complications in outpatients. *Journal of General Internal Medicine*, *15*(3), 207-208.

Gandhi, T.K., Weingart, S.N., Borus, J., Seger, A.C., Peterson, J., Burdick, E. et al. (2003). Adverse drug events in ambulatory care. *New England Journal of Medicine*, *348*(16),1556-64.

Honigman, B., Lee, J., Rothschild, J., Light, P. et al. (2001). Using computerized data to identify adverse drug events in outpatients. *Journal of American Medical Informatics Association, 8*, 254-266.

Johnson, J.A., & Bootman, J.L. (1995). Drug-related morbidity and mortality. A cost-of-illness model. *Archives of Internal Medicine*, *155*(18), 1949-1956.

Lazarou, J., Pomeranz, B., & Corey, P. (1998). Incidence of adverse drug reactions in hospitalized patients. *Journal of American Medical Association*, *279*, 1200-1205.

Menachemi, N., Burke, D., & Brooks, R.G. (2004). Adoption factors associated with patient safety-related information technology. *J Healthc Qual*., *26*(6), 39-44.

Murff, H.J., Patel, V.L., Hripcsak, G., & Bates, D.W. (2003). Detecting adverse events for patient safety research: A review of current methodologies. *Journal of Biomedical Informatics*, *36*(1-2), 131-43.

Murray, M.D., Loos, B., Tu, W., Eckert, G.J., Zhou, X.H., & Tierney, W.M. (1998). Effects of computer-based prescribing on pharmacist work patterns. *Journal of American Medical Informatics Association*, *5*(6), 585-586.

Oh, Y., McCombs, J.S., Cheng, R.A., & Johnson, K.A. (2002). Pharmacist time required for counseling in an outpatient pharmacy. *American Journal of Health-System Pharmacy*, *59*(23), 2346-2355.

Schommer, J.C., Pedersen, C.A., Doucette, W.R., Gaither, C.A., & Mott, D.A. (2002). Community pharmacists' work activities in the United States during 2000. *Journal of American Pharmacists Association*, *42*(3), 399-406.

U.S. Department of Health and Human Services. (1999, May). Managing the risk from medical product use, creating a risk management framework. Taskforce on Risk Management Report.

Weinberger, M., Murray, M.D., Marrero, D.G., Brewer, N., Lykens, M., Harris, L.E. et al. (2002, August). Issues in conducting randomized controlled trials of health services research interventions in nonacademic practice settings: The case of retail pharmacies. *Health Services Research*.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 2, Issue 3, edited by J. Tan, pp. 65-80, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 2.21
Information Retrieval by Semantic Similarity

Angelos Hliaoutakis
Technical University of Crete (TUC), Greece

Giannis Varelas
Technical University of Crete (TUC), Greece

Epimenidis Voutsakis
Technical University of Crete (TUC), Greece

Euripides G. M. Petrakis
Technical University of Crete (TUC), Greece

Evangelos Milios
Dalhousie University, Canada

ABSTRACT

Semantic Similarity relates to computing the similarity between conceptually similar but not necessarily lexically similar terms. Typically, semantic similarity is computed by mapping terms to an ontology and by examining their relationships in that ontology. We investigate approaches to computing the semantic similarity between natural language terms (using WordNet as the underlying reference ontology) and between medical terms (using the MeSH ontology of medical and biomedical terms). The most popular semantic similarity methods are implemented and evaluated using WordNet and MeSH. Building upon semantic similarity, we propose the Semantic Similarity based Retrieval Model (SSRM), a novel information retrieval method capable for discovering similarities between documents containing conceptually similar terms. The most effective semantic similarity method is implemented into SSRM. SSRM has been applied in retrieval on OHSUMED (a standard TREC collection available on the Web). The experimental results demonstrated promising performance improvements over classic information retrieval

methods utilizing plain lexical matching (e.g., Vector Space Model) and also over state-of-the-art semantic similarity retrieval methods utilizing ontologies.

INTRODUCTION

Semantic Similarity relates to computing the similarity between concepts, which are not necessarily lexically similar. Semantic similarity aims at providing robust tools for standardizing the content and delivery of information across communicating information sources. This has long been recognized as a central problem in Semantic Web where related sources need to be linked and communicate information to each other. Semantic Web will also enable users to retrieve information in a more natural and intuitive way (as in a "query-answering" interaction).

In the existing Web, information is acquired from several disparate sources in several formats (mostly text) using different language terminologies. Interpreting the meaning of this information is left to the users. This task can be highly subjective and time consuming. To relate concepts or entities between different sources (the same as for answering user queries involving such concepts or entities), the concepts extracted from each source must be compared in terms of their meaning (i.e., semantically). Semantic similarity offers the means by which this goal can be realized.

This article deals with a certain aspect of Semantic Web and semantics, that of semantic text association, and text semantics respectively. We demonstrate that it is possible to approximate algorithmically the human notion of similarity using semantic similarity and to develop methods capable of detecting similarities between conceptually similar documents even when they don't contain lexically similar terms. The lack of common terms in two documents does not necessarily mean that the documents are not related. Computing text similarity by classical information retrieval models (e.g., Vector Space, Probabilistic, Boolean (Yates & Neto, 1999)) is based on lexical term matching. However, two terms can be semantically similar (e.g., can be synonyms or have similar meaning) although they are lexically different. Therefore, classical retrieval methods will fail to associate documents with semantically similar but lexically different terms.

In the context of the multimedia semantic Web, this article permits informal textual descriptions of multimedia content to be effectively used in retrieval, and obviates the need for generating structured metadata. Informal descriptions require significantly less human labor than structured descriptions.

In the first part of this article, we present a critical evaluation of several semantic similarity approaches for computing the semantic similarity between terms using two well-known taxonomic hierarchies namely WordNet[1] and MeSH[2]. WordNet is a controlled vocabulary and thesaurus offering a taxonomic hierarchy of natural language terms developed at Princeton University. MeSH (Medical Subject Heading) is a controlled vocabulary and a thesaurus developed by the U.S. National Library of Medicine (NLM)[3] offering a hierarchical categorization of medical terms. Similar results for MeSH haven't been reported before in the literature. All methods are implemented and integrated into a semantic similarity system, which is accessible on the Web.

In the second part of this article, we propose the "Semantic Similarity Retrieval Model" (*SSRM*). *SSRM* suggests discovering semantically similar terms in documents (e.g., between documents and queries) using general or application specific term taxonomies (e.g., WordNet or MeSH) and by associating such terms using semantic similarity methods. Initially, *SSRM* computes *tf·idf* weights to term representations of documents. These representations are then augmented by semantically similar terms (which are discovered from WordNet or MeSH by applying a range query in the neighborhood of each term in the taxonomy) and by re-computing weights to all new and pre-existing terms. Finally, document similarity is computed by associating semantically similar terms in the documents and in the queries respectively and by accumulating their similarities.

SSRM together with the term-based Vector Space Model (Salton, 1989) (the classic document retrieval method utilizing plain lexical similarity) as well as the most popular semantic information retrieval methods in the literature (Richardson & Smeaton, 1995; Salton, 1989; Voorhees, 1994) are all implemented and evaluated on OHSUMED (Hersh, Buckley, Leone, & Hickam, 1994), a standard TREC collection with 293,856 medical articles, and on a crawl of the Web with more than 1.5 million Web pages with images. *SSRM* demonstrated promising performance achieving better precision and recall than its competitors.

RELATED WORK

Query expansion with potentially related (e.g., similar) terms has long been considered a means for resolving term ambiguities and for revealing the hidden meaning in user queries. A recent contribution by Collins-Thomson and Callan (2005) proposed a framework for combining multiple knowledge sources for revealing term associations and for determining promising terms for query expansion. Given a query, a term network is constructed representing the relationships between query and potentially related terms obtained by multiple knowledge sources such as synonym dictionaries, general word association scores, co-occurrence relationships in corpus or in retrieved documents. In the case of query expansion, the source terms are the query terms and the target terms are potential expansion terms connected with the query terms by labels representing probabilities of relevance. The likelihood of relevance between such terms is computed using random walks and by estimating the probability of the various aspects of the query that can be inferred from potential expansion terms. *SSRM* is complementary to this approach: It shows how to handle more relationship types (e.g., hyponyms, hypernyms in an ontology) and how to compute good relevance weights given the *tf·idf* weights of the initial query terms. *SSRM* focuses on semantic relationships; a specific aspect of term relationships not considered in Collins-Thomson et al. (2005) and demonstrates that it is possible to enhance the performance of retrievals using this information alone.

SSRM is also complementary to Voorhees (1994) as well as to Richardson and Smeaton (1995). Voorhees proposed expanding query terms with synonyms, hyponyms, and hypernyms in WordNet but did not propose an analytic method for setting the weights of these terms. Voorhees reported some improvement for short queries, but little or no improvement for long queries. Richardson and Smeaton proposed taking the summation of the semantic similarities between all possible combinations of document and query terms. They ignored the relative significance of terms (as captured by *tf·idf* weights) and they considered neither term expansion nor re-weighting. Our proposed method takes term weights into account, introduces an analytic and intuitive term expansion and re-weighting method, and suggests a document similarity formula that takes the previous information into account. Similarly to *SSRM*, the text retrieval method (Mihalcea, Corley, & Strapparava, 2006) works by associating only the most semantically similar terms in two documents and by summing up their semantic similarities (weighted by the inverse document frequency *idf*). Query terms are neither expanded nor re-weighted as in *SSRM*. Notice that *SSRM* associates all terms in the two documents and accumulates their semantic similarities.

The methods previously referred to allow for ordering the retrieved documents by decreasing similarity to the query taking into account that two documents may match only partially (i.e., a retrieved document need not contain all query terms). Similarly, to classic retrieval models like VSM, *SSRM* allows for non-binary weights in queries and in documents (initial weights are computed using the standard *tf idf* formula). The experimental results in this article demonstrate that *SSRM* performs better (achieving better precision and recall) than its competitors like Salton (1989) and ontology-based methods (Richardson & Smeaton, 1995; Voorhees, 1994).

Query expansion and the term re-weighting in *SSRM* resemble also earlier approaches, which attempt to improve the query with terms obtained from a similarity thesaurus (e.g., based on term to term relationships (Mandala, Takenobu, & Hozumi, 1998; Qiu & Frei, 1993)). This thesaurus is usually computed by automatic or semi-automatic corpus analysis (global analysis) and would not only add new terms to *SSRM* but also reveal new relationships not existing in a taxonomy of terms. Along the same lines, Possas, Ziviani, Meira, and Neto (2005) exploit the intuition that co-occurring terms occur close to each other and propose a method for extracting patterns of co-occurring terms and their weights by data mining. These approaches depend on the corpus.

SSRM is independent of the corpus and works by discovering term associations based on their conceptual similarity in a lexical ontology specific to the application domain at hand (i.e., WordNet or MeSH in this article). The proposed query expansion scheme is complementary to methods, which expand the query with co-occurring terms (e.g., "railway," "station") in retrieved documents (Attar & Fraenkel, 1977) (local analysis). Expansion with co-occurring terms (the same as a thesaurus like expansion) can be introduced as additional expansion step in the method. Along the same lines, *SSRM* needs to be extended to work with phrases (Liu, Liu, Yu, & Meng, 2004).

SEMANTIC SIMILARITY

Issues related to semantic similarity algorithms along with issues related to computing semantic similarity on WordNet and MeSH are discussed next.

WordNet

WordNet[4] is an online lexical reference system developed at Princeton University. WordNet attempts to model the lexical knowledge of a native speaker of English. WordNet can also be seen as ontology for natural language terms. It contains around 100,000 terms, organized into taxonomic hierarchies. Nouns, verbs, adjectives, and adverbs are grouped into synonym sets (synsets). The synsets are also organized into senses (i.e., corresponding to different meanings of the same term or concept). The synsets (or concepts) are related to other synsets higher or lower in the hierarchy defined by different types of relationships. The most common relationships are the *Hyponym/Hypernym* (i.e., Is-A relationships), and the *Meronym/Holonym* (i.e., Part-Of relationships). There are nine noun and several verb Is-A hierarchies (adjectives and adverbs are not organized into Is-A hierarchies). Figure 1 illustrates a fragment of the WordNet Is-A hierarchy.

MeSH

MeSH[5] (Medical Subject Headings) is a taxonomic hierarchy (ontology) of medical and biological terms (or concepts) suggested by the U.S. National Library of Medicine (NLM). MeSH terms are organized in Is-A taxonomies with more general terms (e.g., "chemicals and drugs") higher in a taxonomy than more specific terms (e.g., "aspirin"). There are 15 taxonomies with more than 22,000 terms. A term may appear in more than one taxonomy. Each MeSH term is described by several properties the most important of them being the *MeSH Heading* (MH) (i.e., term name

Figure 1. A fragment of the WordNet Is-A hierarchy

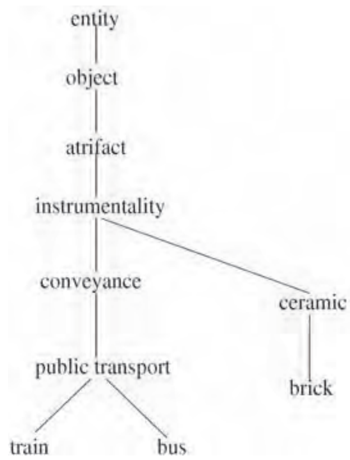
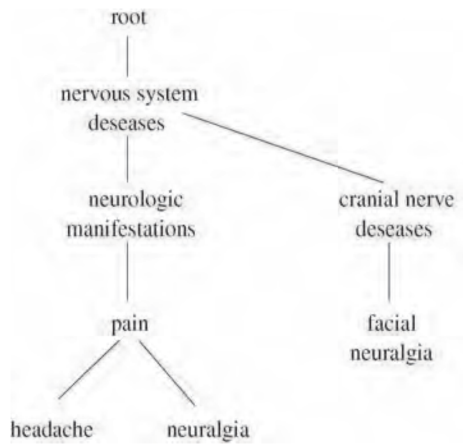


Figure 2. A fragment of the MeSH Is-A hierarchy



or identifier), *Scope Note* (i.e., a text description of the term) and *Entry Terms* (i.e., mostly synonym terms to the MH). Entry terms also include stemmed MH terms and are sometimes referred to as quasi-synonyms (they are not always exactly synonyms). Each MeSH terms is also characterized by its MeSH tree number (or code name) indicating the exact position of the term in the MeSH tree taxonomy (e.g., "D01,029" is the code name of term "Chemical and drugs"). Figure 2 illustrates a fragment of the MeSH Is-A hierarchy.

Semantic Similarity Methods

Several methods for determining semantic similarity between terms have been proposed in the literature and most of them have been tested on WordNet[6]. Similar results on MeSH haven't been reported in the literature.

Semantic similarity methods are classified into four main categories:

1. Edge Counting Methods: Measure the similarity between two terms (concepts) as a function of the length of the path linking the terms and on the position of the terms in the taxonomy (Leacok & Chodorov, 1998; Li, Bandar, & McLean, 2003; Rada, Mili, Bicknell, & Blettner, 1989; Richardson, Smeaton, & Murphy, 1994; Wu & Palmer, 1994).
2. Information Content Methods: Measure the difference in information content of the two terms as a function of their probability of occurrence in a corpus (Lin, 1993; Lord, Stevens, Brass, & Goble, 2003; Jiang & Conrath, 1998; Resnik, 1999). In this article, information content is computed according to Seco, Veale, and Hayes (2004). The taxonomy (WordNet or MeSH in this article) is used as a statistical resource for computing the probabilities of occurrence of terms. More general concepts (higher in the hierarchy) with many hyponyms convey less information content than more specific terms (lower in the hierarchy) with less hyponyms. This approach is independent of the corpus and also guarantees that the information content of each term is less than the information content of its subsumed terms. This constraint is common to all methods of this category. Computing information content from a corpus does not always guarantee this requirement. The same method is also applied for computing the information content of terms.

3. Feature Based Methods: Measure the similarity between two terms as a function of their properties (e.g., their definitions or "glosses" in WordNet or "scope notes" in MeSH) or based on their relationships to other similar terms in the taxonomy. Common features tend to increase the similarity and (conversely) non-common features tend to diminish the similarity of two concepts (Tversky, 1977).
4. Hybrid methods combine the previous ideas (Rodriguez & Egenhofer, 2003). Term similarity is computed by matching synonyms, term neighborhoods, and term features. Term features are further distinguished into parts, functions, and attributes and are matched similarly to (Tversky, 1977).

Semantic similarity methods can also be distinguished between:

1. Single Ontology similarity methods, which assume that the terms which are compared are from the same ontology (e.g., MeSH).
2. Cross Ontology similarity methods, which compare terms from two different ontologies (e.g., WordNet and MeSH).

An important observation and a desirable property of most semantic similarity methods is that they assign higher similarity to terms which are close together (in terms of path length) and lower in the hierarchy (more specific terms), than to terms which are equally close together but higher in the hierarchy (more general terms).

Edge counting and information content methods work by exploiting structure information (i.e., position of terms) and information content of terms in a hierarchy and are best suited for comparing terms from the same ontology. Because the structure and information content of different ontologies are not directly comparable, cross ontology similarity methods usually call for hybrid or feature based approaches. The focus of this work is on single ontology methods. For details on the methods used in the work, please refer to Varelas (2005).

Additional properties of the similarity methods previously referred to are summarized in Table 1. It shows method type, whether similarity affected by the common characteristics of the concepts which are compared, whether it decreases with their differences, whether the similarity is a symmetric property, whether its value is normalized in [0,1] and, finally, whether it is affected by the position of the terms in the taxonomy.

Semantic Similarity System

All methods previously stated are implemented and integrated into a semantic similarity system which is accessible on the Web[7]. Figure 3 illustrates the architecture of this system. The system communicates with WordNet and MeSH. Each term is represented by its tree hierarchy (corresponding to an XML file, which is stored in the XML repository. The tree hierarchy of a term represents the relationships of the term with its hyponyms and hypernyms. These XML files are created by the XML generator using the WordNet XML Web-Service[8]. The purpose of this structure is to facilitate access to terms stored in the XML repository by indexing the terms by their name of identifier (otherwise accessing a term would require exhaustive searching through the entire WordNet or MeSH files). The information content of all terms is also computed in advance and stored separately in the information content database. The user is provided with several options at the user interface (e.g., sense selection, method selection).

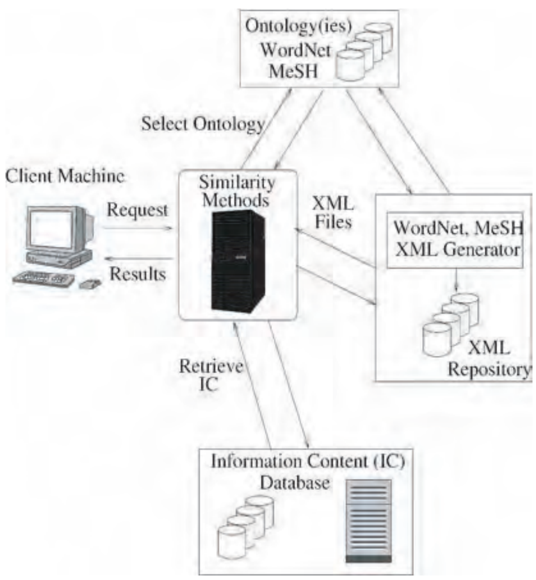
Evaluation of Semantic Similarity Methods

In the following, we present a comparative evaluation of the similarity methods referred to previously.

Table 1. Summary of semantic similarity methods

Method	Method Type	Increases with Commonality	Decreases with Difference	Symmetric Property	Normalized in [0,1]	Position in Hierarchy
(Rada et al., 1989)	Edge Counting	yes	yes	yes	yes	no
(Wu & Palmer, 1994)	Edge Counting	yes	yes	yes	yes	yes
(Li et al., 2003)	Edge Counting	yes	yes	yes	yes	yes
(Leacok & Chodorov, 1998)	Edge Counting	no	yes	yes	no	yes
(Richardson et al., 1994)	Edge Counting	yes	yes	yes	yes	yes
(Resnik, 1999)	Info Content	yes	no	yes	no	yes
(Lin,1993)	Info Content	yes	yes	yes	yes	yes
(Lord et al., 2003)	Info Content	yes	no	yes	yes	yes
(Jiang & Conrath, 1998)	Info Content	yes	yes	yes	no	yes
(Tversky, 1977)	Feature	yes	yes	no	yes	no
(Rodriguez & Egenhofer, 2003)	Hybrid	yes	yes	no	yes	no

Figure 3. Semantic similarity system



Semantic Similarity on WordNet

In accordance with previous research (Resnik, 1999), we evaluated the results obtained by applying the semantic similarity methods presented in this work to the same pairs used in the experiment by Miller and Charles (1991): 38 undergraduate students were given 30 pairs of nouns and were asked to rate the similarity of each pair on a scale from 0 (not similar) through 4 (perfect synonymy). The average rating of each pair represents a good estimate of how similar the two words are.

We compared the computed similarity scores for the same terms as in Miller and Charles with the human relevance results reported there. The similarity values obtained by all competitive computational methods (all senses of the first term are compared with all senses of the second term) are correlated with the average scores obtained by the humans in Miller and Charles (1991). The higher the correlation of a method the better the method is (i.e., the more it approaches the results of human judgments).

Table 2 shows the correlation obtained by each method. Jiang et al. (1998) suggested removing one of the pairs from the evaluation. This increased the correlation of their method to 0.87. The method by Li et al. (2003) is among the best and it is also the fastest. These results lead to the following observations:

1. Information Content methods perform very well and close to the upper bound suggested by (Resnik, 1999).
2. Methods that consider the positions of the terms in the hierarchy (e.g., Li et al., 2003) perform better than plain path length methods (e.g., Rada et al., 1989).

Table 2. Evaluation of semantic similarity methods on WordNet

Method	Method Type	Correlation
(Rada et al., 1989)	Edge Counting	0.59
(Wu & Palmer, 1994)	Edge Counting	0.74
(Li et al., 2003)	Edge Counting	0.82
(Leacok & Chodorov, 1998)	Edge Counting	0.82
(Richardson et al., 1994)	Edge Counting	0.63
(Resnik, 1999)	Info Content	0.79
(Lin, 1993)	Info Content	0.82
(Lord et al., 2003)	Info Content	0.79
(Jiang & Conrath, 1998)	Info Content	0.83
(Tversky, 1977)	Feature	0.73
(Rodriguez & Egenhofer, 2003)	Hybrid	0.71

3. Methods exploiting the properties (i.e., structure and information content) of the underlying hierarchy perform better than Hybrid and Feature based methods, which do not fully exploit this information. However, Hybrid and feature based methods (e.g., Rodriguez et al., 2003), are mainly targeted towards cross ontology similarity applications where edge counting and information content methods do not apply.

Semantic Similarity on MeSH

An evaluation of Semantic Similarity methods on MeSH haven't been reported in the literature before. For the evaluation, we designed an experiment similar to that by Miller and Charles (1991) for WordNet: We asked a medical expert to compile a set of MeSH term pairs. A set of 49 pairs was proposed, together with an estimate of similarity between 0 (not similar) and 4 (perfect similarity) for each pair. To reduce the subjectivity of similarity estimates, we created a form-based interface with all pairs on the Web[9] and we invited other medical experts to enter their evaluation (the interface is still accepting results by experts worldwide). So far, we received estimates from 12 experts.

The analysis of the results revealed that: (a) Some medical terms are more involved or ambiguous leading to ambiguous evaluation by many users. For each pair, the standard deviation of their similarity (over all users) was computed. Pairs with standard deviation higher than a user defined threshold t=0.8 were excluded from the evaluation. (b) Medical experts were not at the same level of expertise and (in some cases) gave unreliable results. For each user we computed the standard deviation of their evaluation (over all pairs). We excluded users who gave significantly different results from the majority of other users. Overall, 13 out of the 49 pairs and 4 out of the 12 users were excluded from the evaluation.

Following the same procedure as in the WordNet experiments, the similarity values obtained by each method (all senses of the first term are compared with all senses of the second term) are correlated with the average scores obtained by the humans.

The correlation results are summarized in Table 3. These results lead to similar observations with the previous experiment:

1. Edge counting and information content methods perform about equally well. However, methods that consider the positions of the

Table 3. Evaluation of semantic similarity methods on MeSH

Method	Method Type	Correlation
(Rada et al., 1989)	Edge Counting	0.50
(Wu & Palmer, 1994)	Edge Counting	0.67
(Li, et al., 2003)	Edge Counting	0.70
(Leacok & Chodorov, 1998)	Edge Counting	0.74
(Richardson, et al., 1994)	Edge Counting	0.64
(Resnik, 1999)	Info Content	0.71
(Lin, 1993)	Info Content	0.72
(Lord et al., 2003)	Info Content	0.70
(Jiang & Conrath, 1998)	Info Content	0.71
(Tversky, 1977)	Feature	0.67
(Rodriguez & Egenhofer, 2003)	Hybrid	0.71

terms (lower or higher) in the hierarchy (e.g., Li et al., 2003) perform better than plain path length methods (e.g., Rada et al., 1989).

2. Hybrid and feature based methods exploiting properties of terms (e.g., scope notes, entry terms) perform at least as well as information content and edge counting methods (exploiting information relating to the structure and information content of the underlying taxonomy), implying that term annotations in MeSH represent significant information by themselves and that it is possible to design even more effective methods by combining information from all the sources listed previously (term annotations, structure information and information content).

SEMANTIC SIMILARITY RETRIEVAL MODEL (*SSRM*)

Traditionally, the similarity between two documents (e.g., a query q and a document d) is computed according to the Vector Space Model (VSM) (Salton, 1989) as the cosine of the inner product between their document vectors:

$$Sim(q,d) = \frac{\sum_i q_i d_i}{\sqrt{\sum_i q_i^2 \sum_i d_i^2}} \qquad (1)$$

where q_i and d_i are the weights in the two vector representations. Given a query, all documents are ranked according to their similarity with the query. This model is also known as the “bag of words model” for document retrieval.

The lack of common terms in two documents does not necessarily mean that the documents are unrelated. Semantically similar concepts may be expressed in different words in the documents and the queries, and direct comparison by word-based VSM is not effective. For example, VSM will not recognize synonyms or semantically similar terms (e.g., “car,” “automobile”).

SSRM suggests discovering semantically similar terms using term taxonomies like WordNet or MeSH. Query expansion is also applied as a means for capturing similarities between terms of different degrees of generality in documents and queries (e.g., “human,” “man”). Queries are augmented with conceptually similar terms, which are retrieved by applying a range query in the neighborhood of each term in an ontology. Each query term is expanded by synonyms, hyponyms, and hypernyms. The degree of expansion is controlled by the user (i.e., so that each query term may introduce new terms more than one level higher or lower in an ontology). *SSRM* can work with any general or application specific ontology.

The selection of ontology depends on the application domain (e.g., WordNet for image retrieval on the Web (Varelas, Voutsakis, Raftopoulou, Petrakis, & Milios, 2005), MeSH for retrieval in medical document collections (Hliaoutakis, Varelas, Petrakis, & Milios, 2006)).

Query expansion by *SSRM* resembles the idea by Voorhees (1994). However, Voorhees did not show how to compute good weights for the new terms introduced into the query after expansion nor it showed how to control the degree of expansion. Notice that, high degree of expansion results in topic drift. *SSRM* solves this problem and implements an intuitive and analytic method for setting the weights of the new query terms.

Voorhees relied on the Vector Space Model (VSM) and therefore on lexical term matching for computing document similarity. Therefore, it is not possible for this method to retrieve documents with conceptually similar but lexically different terms. *SSRM* solves this problem by taking all possible term associations between two documents into account and by accumulating their similarities.

Similarly to VSM, queries and documents are first syntactically analyzed and reduced into term vectors. Very infrequent or very frequent terms are eliminated. Each term in this vector is represented by its weight. The weight of a term is computed as a function of its frequency of occurrence in the document collection and can be defined in many different ways. The term frequency-inverse document frequency model (Salton, 1989) is used for computing the weight: The weight d_i of a term *i* in a document is computed as $d_i = tf_i \cdot idf_i$, where tf_i is the frequency of term *i* in the document and idf_i is the inverse document frequency of *i* in the whole document collection.

Then SSRM works in three steps:

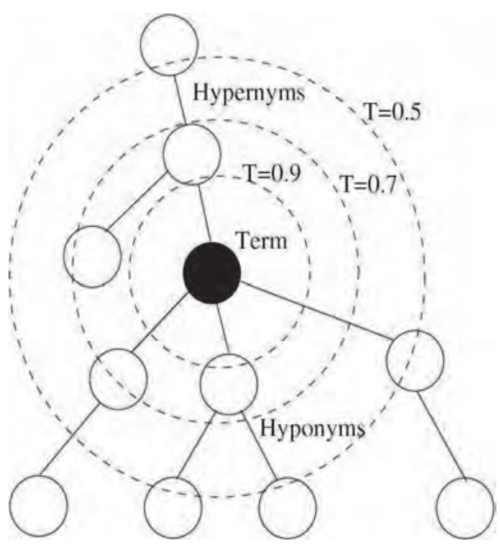
Query Re-Weighting: The weight q_i of each query term *i* is adjusted based on its relationships with other semantically similar terms *j* within the same vector:

$$q_i' = q_i + \sum_{sim(i,j) \geq t}^{j \neq i} q_j sim(i,j) \qquad (2)$$

where *t* is a user defined threshold (*t*=0.8 in this article). Multiple related terms in the same query reinforce each other (e.g., "railway," "train," and "metro"). The weights of non-similar terms remain unchanged (e.g., "train," "house"). For short queries specifying only a few terms the weights are initialized to 1 and are adjusted according to the formula in Equation 2.

Query Expansion: First, the query is augmented by synonym terms, using the most common sense of each query term. Then, the query is augmented by terms higher or lower in the tree hierarchy (i.e., hypernyms and hyponyms) which are semantically similar to terms already in the query. Figure 4 illustrates this process: Each query term is represented by its tree hierarchy. The neighborhood of the term is examined and all terms with similarity greater than threshold *T* are also included in the query vector. This expansion may include terms more than one level higher or lower than the original term. Then, each query

Figure 4. Term expansion



term i is assigned a weight as follows:

$$q_i' = \begin{cases} q_i + \sum\limits_{sim(i,j) \geq T}^{i \neq j} \frac{1}{n} q_j \, sim(i,j), & i \text{ had weight } q_i \\ \sum\limits_{sim(i,j) \geq T}^{i \neq j} \frac{1}{n} q_j \, sim(i,j), & i \text{ is a new term} \end{cases} \quad (3)$$

where n is the number of hyponyms of each expanded term j. For hypernyms $n=1$. The summation is taken over all terms j introducing terms to the query. It is possible for a term to introduce terms that already existed in the query. It is also possible that the same term is introduced by more than one other terms. Equation 2, suggests taking the weights of the original query terms into account and that the contribution of each term in assigning weights to query terms is normalized by the number n of its hyponyms. After expansion and re-weighting, the query vector is normalized by document length, like each document vector.

Document Similarity: The similarity between an expanded and re-weighted query q and a document d is computed as:

$$Sim(q,d) = \frac{\sum_i \sum_j q_i q_j \, sim(i,j)}{\sum_i \sum_j q_i q_j} \quad (4)$$

where i and j are terms in the query and the document respectively. Query terms are expanded and re-weighted according to the previous steps while document terms d_j are computed as *tf idf* terms (they are neither expanded nor re-weighted). The similarity measure in Equation 4 is normalized in the range [0,1]. Figure 5 presents a summary of *SSRM*.

Discussion

SSRM relaxes the requirement of classical retrieval models that conceptually similar terms be mutually independent (known also as "synonymy problem"). It takes into account dependencies between terms during its expansion and re-weighting steps. Their dependence is expressed quantitatively by virtue of their semantic similarity and this information is taken explicitly into account in the computation of document similarity. Notice however the quadratic time complexity of *SSRM* due to Equation 3 as opposed to the linear time complexity of Equation 1 of VSM. To speed up similarity computations, the semantic similarities

Figure 5. SSRM algorithm

Input: Query q, Document d, Semantic Similarity function *sim()*, Thresholds t, T, Ontology.
Output: Document similarity value $Sim(d,q)$.

1. **Compute query term vector**: $q=(q_1,q_2,...)$ using *tf idf* weighting scheme.
2. **Compute document term vector**: $d=(d_1,d_2,...)$ using *tf idf* weighting scheme.
3. **Query re-weighting**: For all term i in q computer new weight based on other semantically similar terms j in q as $q_i' = q_i + \sum\limits_{sim(i,j) \geq t}^{j \neq i} q_j \, sim(i,j)$.
4. **Query expansion**: For all terms j in q retrieve terms i from ontology satisfying $sim(i,j) \geq T$.
5. **Term re-weighting**: For all terms i in q compute new weight as
$$q_i' = \begin{cases} q_i + \sum\limits_{sim(i,j) \geq T}^{i \neq j} \frac{1}{n} q_j \, sim(i,j), & i \text{ had weight } q_i \\ \sum\limits_{sim(i,j) \geq T}^{i \neq j} \frac{1}{n} q_j \, sim(i,j), & i \text{ is a new term} \end{cases}$$
6. **Query Normalization**: Normalize query by length.
7. **Compute document similarity**: $Sim(q,d) = \frac{\sum_i \sum_j q_i q_j \, sim(i,j)}{\sum_i \sum_j q_i q_j}$.

between pairs of MeSH or WordNet terms are stored in a hash table. To reduce space only pairs with similarity greater than 0.3 are stored.

SSRM approximates VSM in the case of non-semantically similar terms: If $sim(i,j)=0$ for all $i \neq j$ then Equation 3 is reduced to Equation 1. In this case, the similarity between two documents is computed as a function of weight similarities between identical terms (as in VSM).

Expanding and re-weighting is fast for queries, which are typically short, consisting of only a few terms, but not for documents with many terms. The method suggests expansion of the query only. However, the similarity function will take into account the relationships between all semantically similar terms between the document and the query (something that VSM cannot do).

The expansion step attempts to automate the manual or semi-automatic query re-formulation process based on feedback information from the user (Rochio, 1971). Expanding the query with a threshold T will introduce new terms depending also on the position of the terms in the taxonomy: More specific terms (lower in the taxonomy) are more likely to expand than more general terms (higher in the taxonomy). Notice that expansion with low threshold values T (e.g., T=0.5) is likely to introduce many new terms and diffuse the topic of the query (topic drift). The specification of threshold T may also depend on query scope or user uncertainty. A low value of T might be desirable for broad scope queries or for initially resolving uncertainty as to what the user is really looking for. The query is then repeated with higher threshold. High values of threshold are desirable for very specific queries: Users with high degree of certainty might prefer to expand with a high threshold or not to expand at all.

The specification of T in Equation 2 requires further investigation. Appropriate threshold values can be learned by training or relevance feedback (Rui, Huang, Ortega, & Mechrota, 1998). Word sense disambiguation (Patwardhan, Banerjee, & Petersen, 2003) can also be applied to detect the correct sense to expand rather than expanding the most common sense of each term. *SSRM* also makes use of a second threshold t for expressing the desired similarity between terms within the query (Equation 1). Our experiments with several values of t revealed that the method is rather insensitive to the selection of this threshold. Throughout this article we set t = 0.8.

Evaluation of *SSRM*

SSRM has been tested on two different applications and two data sets respectively. The first application is retrieval of medical documents using MeSH and the second application is image retrieval on the Web using WordNet.

The experimental results illustrate that it is possible to enhance the quality of classic information retrieval methods by incorporating semantic similarity within the retrieval method.

SSRM outperforms classic and state-of-the-art semantic information retrieval methods (Richardson et al., 1995; Salton, 1989; Voorhees, 1994). The retrieval system is built upon Lucene[10], a full-featured text search engine library written in Java. All retrieval methods are implemented on top of Lucene.

The following methods are implemented and evaluated:

1. Semantic Similarity Retrieval Model (SSRM): Queries are expanded with semantically similar terms in the neighborhood of each term. The results that follow correspond to two different thresholds T=0.9 (i.e. the query is expanded only with very similar terms) and T=0.5 (i.e., the query is expanded with terms which are not necessarily conceptually similar). In WordNet, each query term is also expanded with synonyms. Because no synonymy relation is defined in MeSH we did not apply expansion to Mesh terms in the query with Entry Terms. Semantic similarity in SSRM is computed by Li et al., 2003).

2. Vector Space Model (VSM) (Salton, 1989): Text queries can also be augmented by synonyms.
3. Term expansion (Voorhees, 1994): The query terms are expanded always with hyponyms one level higher or lower in the taxonomy and synonyms. The method did not propose an analytic method for computing the weights of these terms.
4. Semantic similarity accumulation (Richardson et al., 1995): Accumulates the semantic similarities between all pairs of document and query terms. It ignores the relative significance of terms (as it is captured by *tf idf*). Query terms are not expanded nor re-weighted as in SSRM.

In the experiments that follow, each method is represented by a *precision/recall* curve. For each query, the best 50 answers were retrieved (the *precision/recall* plot of each method contains exactly 50 points). Precision and recall values are computed from each answer set and therefore, each plot contains exactly 50 points. The top-left point of a precision/recall curve corresponds to the *precision/recall* values for the best answer or best match (which has rank 1) while the bottom right point corresponds to the *precision/recall* values for the entire answer set. A method is better than another if it achieves better precision and better recall. As we shall see in the experiments, it is possible for two precision-recall curves to crossover. This means that one of the two methods performs better for small answer sets (containing less answers than the number of points up to the cross-section), while the other performs better for larger answer sets. The method achieving higher precision and recall for the first few answers is considered to be the better method (based on the assumption that typical users focus their attention on the first few answers).

Information Retrieval on OHSUMED

SSRM has been tested on OHSUMED[11] (a standard TREC collection with 293,856 medical articles from Medline published between 1988-1991) using MeSH as the underlying ontology. All OHSUMED documents are indexed by title, abstract and MeSH terms (MeSH Headings). These descriptions are syntactically analyzed and reduced into separate vectors of MeSH terms which are matched against the queries according to Equation 3 (as similarity between expanded and re-weighted vectors). The weights of all MeSH terms are initialized to one while the weights of titles and abstracts are initialized by *tf idf*. The similarity between a query and a document is computed as:

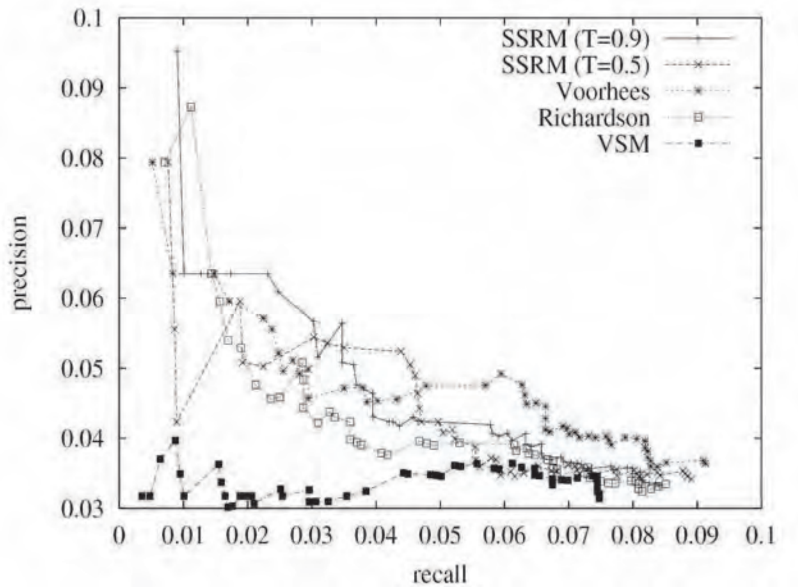
$$Sim(q,d) = Sim(q,d_{MeSH-terms}) + Sim(q,d_{title}) + Sim(q,d_{abstract}) \qquad (5)$$

where $d_{MeSH\text{-}terms}$, d_{title} and $d_{abstract}$ are the representations of the document MeSH terms, title and abstract respectively. This formula suggests that a document is similar to a query if its components are similar to the query. Each similarity component can be computed either by VSM or by *SSRM*.

For the evaluations, we applied the subset of 63 queries of the original query set developed by Hersh et al. (1994). The correct answers to these queries were compiled by the editors of OHSUMED and are also available on the Web along with the queries. A document is considered similar to a query if the query terms are included in the document. OHSUMED provides the means for comparing the performance of different methods. However, it is not particularly well suited for semantic information retrieval with *SSRM*. A better criterion would be to judge whether a document is on the topic of the query (even if it contains lexically different terms).

The results in Figure 6 demonstrate that *SSRM* with expansion with very similar terms T=0.9 and for small answer sets (i.e., with less than eight an-

Figure 6. Precision-recall diagram for retrievals on OHSUMED using MeSH



swers) outperforms all other methods (Richardson et al., 1995; Salton, 1989; Voorhees, 1994). For larger answer sets, Voorhees (1994) is the best method. For answer sets with 50 documents all methods (except VSM) perform about the same. *SSRM* with expansion threshold *T*=0.5 performed worse than *SSRM* with *T*=0.9. An explanation may be that it introduced many new terms and not all of them are conceptually similar with the original query terms.

Image Retrieval on the Web

Searching for effective methods to retrieve information from the Web has been in the center of many research efforts during the last few years. The relevant technology evolved rapidly thanks to advances in Web systems technology (Arasu, Cho, Garcia-Molina, Paepke, & Raghavan, 2002) and information retrieval research (Yates et al., 1999). Image retrieval on the Web, in particular, is a very important problem in itself (Kherfi, Ziou, & Bernardi, 2004). The relevant technology has also evolved significantly propelled by advances in image database research (Smeulders, Worring, Santini, Gupta, & Jain, 2000).

Image retrieval on the Web requires that content descriptions be extracted from Web pages and used to determine which Web pages contain images that satisfy the query selection criteria. Several approaches to the problem of content-based image retrieval on the Web have been proposed and some have been implemented on research prototypes, for example, ImageRover (Taycher, Cascia, & Sclaroff, 1997), WebSEEK (Smith & Chang, 1997), Diogenis (Aslandongan & Yu, 2000), and commercial systems, Google Image Search[12], Yahoo[13], and Altavista[14]. Because, methods for extracting reliable and meaningful image content from Web pages by automated image analysis are not yet available images on the Web are typically described by text or attributes associated with images in html tags (e.g., filename, caption, alternate text etc.). These are automatically extracted from the Web pages and are used in retrievals. Google, Yahoo, and AltaVista are example systems of this category.

We choose the problem of image retrieval based on surrounding text as a case study for this evaluation. *SSRM* has been evaluated through *IntelliSearch*[15], a prototype Web retrieval system for Web pages and images in Web pages. An earlier system we built supported retrievals using only VSM (Voutsakis, Petrakis, & Milios, 2005). In this article, the system has been extended to support retrievals using *SSRM* with WordNet as the underlying reference ontology. The retrieval system of *IntelliSearch* is built upon Lucene and the database stores more than 1.5 million Web pages with images.

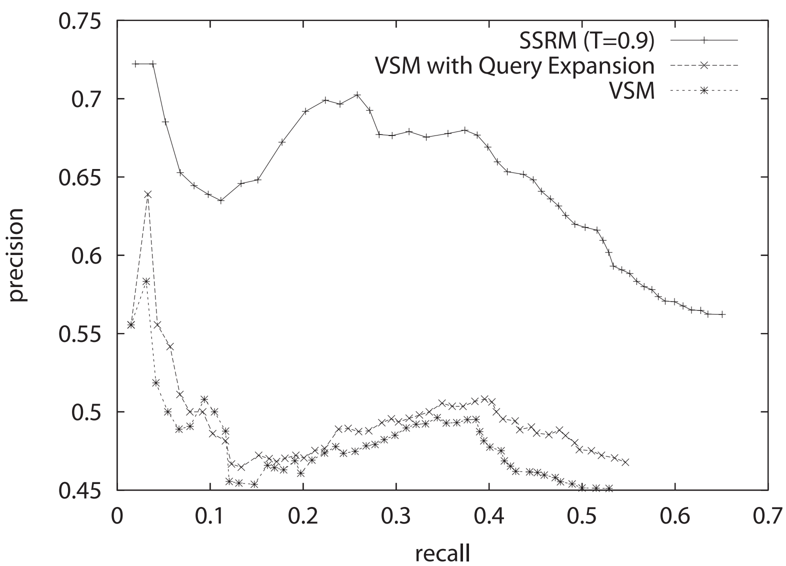
As it is typical in the literature (Petrakis, Kontis, Voutakis, & Milios, 2005; Shen, Ooi, & Tan, 2000; Voutsakis et al., 2005), the problem of image retrieval on the Web is treated as one of text retrieval as follows: Images are described by the text surrounding them in the Web pages (i.e., captions, alternate text, image file names, page title). These descriptions are syntactically analyzed and reduced into term vectors, which are matched against the queries. Similarly, to the previous experiment, the similarity between a query and a document (image) is computed as:

$$Sim(q,d) = Sim(q,d_{image-file-name}) + Sim(q,d_{caption}) + Sim(q,d_{page-title}) + Sim(q,d_{alternate-text}) \quad (6)$$

For the evaluations, 20 queries were selected from the list of the most frequent Google image queries. These are short queries containing between 1 and 4 terms. The evaluation is based on human relevance judgments by five human referees. Each referee evaluated a subset of four queries for both methods.

Figure 7 indicates that *SSRM* is far more effective than VSM achieving up to 30% better precision and up to 20% better recall. A closer look into the results reveals that the efficiency of *SSRM* is mostly due to the contribution of non-identical but semantically similar terms. VSM (like most classical retrieval models relying on lexical term matching) ignore this information. In VSM, query terms may also be expanded with synonyms. Experiments with and without expansion by synonyms are presented. Notice that VSM with query expansion by synonyms improved the results of plain VSM only marginally, indicating that the performance gain of *SSRM* is not due to

Figure 7. Precision-recall diagram for retrievals on the Web using WordNet



the expansion by synonyms but rather due to the contribution of semantically similar terms.

CONCLUSION

This article makes two contributions. The first contribution is to experiment with several semantic similarity methods for computing the conceptual similarity between natural language terms using WordNet and MeSH. To our knowledge, similar experiments with MeSH have not been reported elsewhere. The experimental results indicate that it is possible for these methods to approximate algorithmically the human notion of similarity reaching correlation (with human judgment of similarity) up to 83% for WordNet and up to 74% for MeSH. The second contribution is SSRM, information retrieval method that takes advantage of this result. *SSRM* outperforms VSM, the classic information retrieval method and demonstrates promising performance improvements over other semantic information retrieval methods in retrieval on OHSUMED, a standard TREC collection with medical documents, which is available on the Web. Additional experiments have demonstrated the utility of SSRM in Web image retrieval based on text image descriptions extracted automatically. *SSRM* has been also tested on Medline[16], the premier bibliographic database of the U.S. National Library of Medicine (NLM) (Hliaoutakis et al., 2006). All experiments confirmed the promise of SSRM over classic retrieval models. SSRM can work in conjunction with any taxonomic ontology like MeSH or WordNet and any associated document corpus. Current research is directed towards extending *SSRM* to work with compound terms (phrases), and more term relationships (in addition to the Is-A relationships).

ACKNOWLEDGMENT

Dr Qiufen Qi of Dalhousie University for prepared the MeSH terms and the queries for the experiments with MeSH and evaluated the results of retrievals on Medline. We thank Nikos Hurdakis, and Paraskevi Raftopoulou for valuable contributions into this article. The U.S. National Library of Medicine provided us with the complete data sets of MeSH and Medline. This article was funded by project MedSearch/BIOPATTERN (Fp6, Project No 508803) of the European Union (EU), the Natural Sciences and Engineering Research Council of Canada, and IT Interactive Services Inc.

REFERENCES

Arasu, A., Cho, J., Garcia-Molina, H., Paepke, A., & Raghavan, S. (2002). Searching the Web. *ACM Transactions on Internet Technology, 1*(1), 2-43.

Aslandongan, Y. A., & Yu, C. T. (2000). Evaluating strategies and systems for content-based indexing of person images on the Web. In *Proceedings of the International Conference on Multimedia* (pp. 313-321).

Attar, R., & Fraenkel, A. S. (1977). Local feedback in full text retrieval systems. *Journal of the ACM, 23*(3), 397-417.

Collins-Thomson, K., & Callan, J. (2005). Query expansion using random walk models. In *Proceedings of CIKM* (pp. 704-711).

Hersh, W. R., Buckley, C., Leone, T. J., & Hickam, D. H. (1994). OHSUMED: An interactive retrieval evaluation and new large. In *Proceedings of ACM SIGIR* (pp. 192-201).

Hliaoutakis, A., Varelas, G., Petrakis, E. G. M., & Milios, E. (2006). MedSearch: A retrieval system for medical information based on semantic similarity. In *Proceedings of ECDL* (pp. 512-515).

Jiang, J. J., & Conrath, D. W. (1998). Semantic similarity based on corpus statistics and lexical taxonomy. In *Proceedings of the International Conference on Research in Computational Linguistics.*

Kherfi, M. L., Ziou, D., & Bernardi, A. (2004). Image retrieval from the World Wide Web: Issues, techniques, and systems. *ACM Computing Surveys, 36*(1), 35-67.

Leacok, C., & Chodorov, M. (1998). Combining local context and WordNet similarity for word sense identification. In C. Fellbaum (Ed.), *WordNet: An electronic lexical database and some of its applications* (chap. 11). Boston: MIT Press.

Li, Y., Bandar, Z. A., & McLean, D. (2003). An approach for measuring semantic similarity between words using multiple information sources. *IEEE Transactions on Knowledge and Data Engineering, 15*(4), 871-882.

Lin, D. (1993). Principle-based parsing without overgeneration. In *ACL* (pp. 112-120).

Liu, S., Liu, F., Yu, C., & Meng, M. (2004). An effective approach to document retrieval via utilizing wordnet and recognizing phrases. In *Proceedings of ACM SIGIR* (pp. 266-272).

Lord, P. W., Stevens, R. D., Brass, A., & Goble, C. A. (2003). Investigating semantic similarity measures across the gene ontology: The relationship between sequence and annotation. *Bioninformatics, 19*(10), 1275-1283.

Mandala, R., Takenobu, T., & Hozumi, T. (1998). The use of WordNet in information retrieval. In *COLING/ACL* (pp. 469-477).

Mihalcea, R., Corley, C., & Strapparava, C. (2006, July). Corpus-based and knowledge-based measures of text semantic similarity. In *Proceedings of the American Association for Artificial Intelligence (AAAI 2006).* Boston.

Miller, G., & Charles, W. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes, 6*(1), 1-28.

Patwardhan, S., Banerjee, S., & Petersen, T. (2003, February). Using measures of semantic relatedness for word sense disambiguation. In *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics*, Mexico City, Mexico (pp. 241-257).

Petrakis, E., Kontis, K., Voutakis, E., & Milios, E. (2005). Relevance feedback methods for logo and trademark image retrieval on the Web. In *Proceedings of ACM SAC, IAR* (pp. 23-27).

Possas, B., Ziviani, N., Meira, W., & Neto, B. R. (2005). Set-based vector model: An efficient approach for correlation-based ranking. *ACM Transactions on Information Systems, 23*(4), 397-429.

Qiu, Y., & Frei, H. P. (1993). Concept-based query expansion. In *Proceedings of SIGIR* (pp. 160-169).

Rada, R., Mili, E., Bicknell, E., & Blettner, M. (1989). Development and application of a metric on semantic nets. *IEEE Transaction on Systems, Man, and Cybernetics, 19*(1), 17-30.

Resnik, O. (1999). Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity and natural language. *Journal of Artificial Intelligence Research, 11*, 95-130.

Richardson, R., & Smeaton, A. (1995). *Using WordNet in a knowledge-based approach to information retrieval* (Working Paper: CA-0395). Dublin, Ireland: School of Computer Applications.

Richardson, R., Smeaton, A., & Murphy, J. (1994). *Using WordNet as a knowledge base for measuring semantic similarity between words* (Working paper CA-1294). Dublin, Ireland: School of Computer Applications, Dublin City University.

Rochio, J. J. (1971). Relevance feedback in information retrieval. In G. Salton (Ed.), *The SMART retrieval system—Experiments in automatic document processing* (pp. 313-323). Englewood Cliffs, NJ: Prentice Hall.

Rodriguez, M. A., & Egenhofer, M. J. (2003). Determining semantic similarity among entity classes from different ontologies. *IEEE Trans. on Knowledge and Data Engineering, 15*(2), 442-456.

Rui, Y., Huang , T. S., Ortega, M., & Mechrota, S. (1998). Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Transaction on Circ. and Syst. for Video Technology, 8*(5), 644-655.

Salton, G. (1989). *Automatic text processing: the transformation analysis and retrieval of information by computer*. Reading, MA: Addison-Wesley.

Seco, N., Veale, T., & Hayes, J. (2004). *An intrinsic information content metric for semantic similarity in WordNet*. Ireland: Dept. of Computer Science, University College Dublin.

Shen, H. T., Ooi, B. C., & Tan, K. L. (2000). Giving meanings to WWW images. In *Proceedings of the International Conference on Multimedia* (pp. 39-47).

Smeulders, A. W. M, Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1349-1380.

Smith, J. R., & Chang, S. F. (1997). Visually searching the Web for content. *IEEE Multimedia, 4*(3), 12-20.

Taycher, L., Cascia , M. L. & Sclaroff, S. (1997). Image digestion and relevance feedback in the ImageRover WWW search engine. In *Proceedings of the International Conference on Visual Information Systems* (pp. 85-94).

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*(4), 327-352.

Varelas, G. (2005). *Semantic similarity methods in WordNet and their application to information retrieval on the Web* (TR-TUC-ISL-01-2005). Retrieved from http://www.intelligence.tuc.gr/publications/Varelas.pdf

Varelas, G., Voutsakis, E., Raftopoulou, P., Petrakis, E. G. M., & Milios, E. (2005). Semantic similarity methods in WordNet and their application to information retrieval on the Web. In *WIDM* (pp. 10-16).

Voorhees, E. M. (1994). Query expansion using lexical-semantic relations. In *Proceedings of the ACM SIGIR* (pp. 61-69).

Voutsakis, E., Petrakis, E., & Milios, E. (2005). Weighted link analysis for logo and trademark image retrieval on the Web. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI)* (pp. 581-585).

Wu, Z., & Palmer, M. (1994). Verb semantics and lexical selection. In *ACL* (pp. 133-138).

Yates, R. B., & Neto, B. R. (1999). *Modern information retrieval*. Boston: Addison Wesley Longman.

END NOTES

1 http://wordnet.princeton.edu
2 http://www.nlm.nih.gov/mesh
3 http://www.nlm.nih.gov
4 http://wordnet.princeton.edu
5 http://www.nlm.nih.gov/mesh
6 http://marimba.d.umn.edu/cgi-bin/similarity/similarity.cgi
7 http://www.intelligence.tuc.gr/similarity
8 http://wnws.sourceforge.net
9 http://www.intelligence.tuc.gr/mesh
10 http://lucene.apache.org

[11] http://trec.nist.gov/data/t9_filtering.html
[12] http://images.google.com
[13] http://images.search.yahoo.com
[14] http://www.altavista.com/image
[15] http://www.intelligence.tuc.gr/intellisearch
[16] http://www.nlm.nih.gov/databases/databases_medline.html

This work was previously published in International Journal on Semantic Web & Information Systems, Vol. 2, Issue 3, edited by A. Sheth and M. Lytras, pp. 55-73, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 2.22
A Distributed Patient Identification Protocol Based on Control Numbers with Semantic Annotation

Marco Eichelberg
OFFIS, Germany

Thomas Aden
OFFIS, Germany

Wilfried Thoben
OFFIS, Germany

ABSTRACT

One important problem of information systems in health care is the localisation and access to electronic patient records across health care institute boundaries, especially in an international setting. The complexity of the problem is increased by the absence of a globally accepted standard for electronic health care records, the absence of unique patient identifiers in most countries, and the strict data protection requirements that apply to clinical documents. This article describes a protocol that allows the identification of locations of patient records for a given patient and provides access to these records, if granted, under consideration of the legal and technical requirements. The protocol combines cryptographic techniques with semantic annotation and mediation and presents a simple Web-service-based access to clinical documents.

INTRODUCTION

Information technology used in the health care sector is most often characterised by heterogeneity of systems, longevity of data and devices, and high availability requirements. While the heterogeneity and longevity of systems is a consequence of the fact that expensive special-purpose devices such as MRI (magnetic resonance imaging) scanners are produced by only a few vendors and need to be integrated with the existing IT infrastructure, the longevity and availability requirements of medical data such as images and diagnostic reports are related directly to the care process. Since patient treatment needs to continue, even if some part of the IT infrastructure is off-line, a distributed data storage with loose message-based coupling between devices is used most often. This implies that data inconsistencies between systems are not unusual and need to be accounted for. These properties of IT systems in health care explain the pressing need for interface standardisation and interoperability in this field, reflected by comprehensive interface standards like DICOM (NEMA, 2004) and HL7 (2003). According to the CEN/ISSS eHealth Standardization Focus Group (2004), a study currently being performed at the request of the European Commission, the key strategic aims for applications of information and communication technology to health services include improving access to clinical records and enabling patient mobility and cross-border access to health care. While much work has been devoted to developing standard system interfaces for applications within a single health care enterprise (i.e., hospital or private practice), the digital cross-enterprise exchange of clinical records is certainly the exception rather than the norm today, particularly in the case of cross-border communication. A key issue in this field is the absence of a unique identifier that could be used to unambiguously identify records pertaining to a particular patient. In this article, we discuss requirements for locating and accessing clinical records across enterprise and country borders under consideration of data protection and propose a protocol based on control numbers and semantic annotation that addresses these requirements. This protocol is being developed within the framework of the ARTEMIS project, which is introduced in the following section.

THE ARTEMIS PROJECT

The ARTEMIS project (ARTEMIS Consortium, 2004; Dogac et al., in press), funded by the European Union, aims to improve the interoperability of clinical information systems among different organisations, based on Semantic Web Services and suitable domain ontologies. A health care organisation can join the ARTEMIS peer-to-peer (P2P) network and advertise electronic services, such as the provision of access to a patient's electronic health care record (given suitable authorisation) and access to different subsystems (e.g., patient admission or laboratory information systems). Within the ARTEMIS network, further services might be invoked dynamically, for example, in order to translate and map among different representations of health care information. In ARTEMIS, all participating health care organisations (peers) are coupled loosely via the ARTEMIS P2P network. Groups of participating organisations are coupled via so-called Super Peers, which are connected among each other. The project is carried out with partners from Turkey, Germany, Greece, and the United Kingdom.

THE PATIENT IDENTIFICATION PROTOCOL

While a number of projects currently are attempting to establish central electronic health record (EHR) archives for certain regions or countries, most clinical records are still kept and maintained at the place of their creation. This means that, given a patient with a disease requiring long-term treatment (such as diabetes), related clinical records may be located at one or more family doctors' practices, several specialists, labs, and a number of hospitals. In particular, the patient may not even be aware of all the locations where records relevant to a particular medical problem may be kept. Any protocol that attempts to make relevant clinical documents available in digital form needs to take this distributed nature of document storage into account. The advent of wide area networks such as the Internet, along with various VPN (virtual private network) technologies provides a solution to the underlying problem of a digital transport connection between document requestor and document provider, but it does not solve the problem of how to *locate* the relevant records.

The task of locating relevant medical records is complicated by the fact that there is no unique patient identifier that could be broadcast as a query in order to locate information pertaining to one patient. While countries such as Turkey, Norway, and Sweden maintain a national person identifier that commonly is used as the index key for medical records, no such unique identifier is available in most other countries either for historic reasons or due to data protection regulations. This means that a query applicable to cross-border health care delivery only can be based on the patient demographics that are commonly available, including the patient's name, date and place of birth, sex, nationality, and postal address.

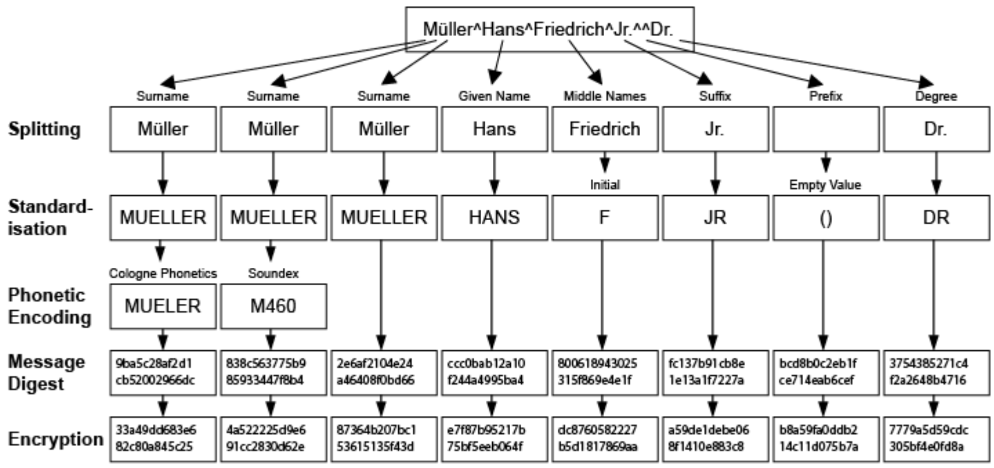
It should be noted that the set of demographics available may depend on the location (i.e., a national patient identifier certainly would be included in any query *within* a country in which it is valid) and on the patient's health condition (i.e., whether the patient is able to provide the doctor with additional information not contained in the passport or driver's license, which may be the only source of information available for an emergency patient). It also should be noted that spelling errors in medical record archives are not uncommon and may need to be accounted for, using, for example, phonetic encoding techniques. An additional challenge for cross-border application is the different character sets used in different European countries. For example, names of patients of Turkish origin are certainly common in German hospital information systems, since this group accounts for more than 2% of the population. However, Turkish names may contain characters not present in the Latin-1 alphabet (ISO, 1999b) commonly used by information systems in Germany, since the Turkish language requires the Latin-9 alphabet (ISO, 1999a), which means that the spelling of one name may differ, depending on the character set supported by the information system.

Since medical records generally are considered to be sensitive personal information, it would neither be appropriate (or legal) for a health care enterprise to allow third parties to browse through the demographics of the local record archive, nor would it be appropriate if a request of the form *Hospital X is looking for prior psychiatry records for patient Hans Friedrich Müller, born 12-24-1960 in Hamburg* were made available widely (i.e., broadcast) to all health care institutes in the network that might possibly have such records available. Clearly, the query already communicates information to the recipient that needs to be protected under the applicable data protection rules; namely, the facts that Mr.Müller currently receives treatment at Hospital X and may have had a prior psychiatric treatment. We propose the use of control numbers along with semantic annotation and a probabilistic record linkage as a way of addressing the possible fuzziness of demographic data and at the same time preventing an inappropriate (i.e., premature and unlimited) communication of personal data.

Control Numbers and Record Linkage

Control numbers is a concept that is used in the epidemiological cancer registries in Germany to allow record linkage of anonymised records that describe cancer cases and are collected independently of multiple sources, as described by Thoben, Appelrath, and Sauer (1994). The patient identification protocol makes use of this concept in a modified form. The generation of a set of control numbers from a set of demographic values is performed through the following series of five processing steps, as shown in Figure 1:

Figure 1. Creation of control numbers



- **Splitting.** This first step initialises the process. The available demographics are split into fields that later are converted into different control numbers. For example, the date of birth typically would be split into different components for year, month, and day.
- **Standardisation.** This step addresses character set issues. Each component is standardised according to a set of rules that needs to be known to all parties participating in the protocol. Standardisation typically would involve conversion of names to upper-case ASCII characters, zero-padding of numbers such as day and month of birth, and the initialisation of unknown and empty fields with well known constants representing the concept of an unknown or empty value. The standardisation process certainly could be extended to Unicode to also cover multi-byte character sets such as Chinese or Japanese Kanji, for which a conversion of names to ASCII may not be appropriate, but this topic is not discussed further within the scope of this article.
- **Phonetic encoding.** Optionally, name components may be encoded with a phonetic encoding, such as the Soundex coding system used by the United States Census Bureau, the Metaphone algorithm or the Cologne phonetics for the German language (Postel, 1969). It should be noted that phonetic encoding is highly language-specific, and therefore, different encodings are likely to be used in different countries. For this reason, a name component always would be converted to at least two different control numbers, one with and one without phonetic encoding.
- **Message digest.** Each standardised and possibly phonetically encoded field in the set of demographics is subjected to a cryptographic message digest algorithm such as MD5, SHA-1, or RIPEMD-160 in this step. Due to the cryptographic properties of this class of algorithms, it is not possible to efficiently construct a matching input string to a given message digest; that is, the digest function is not reversible. It should be noted, however, that the set of possible input values for a patient's sex, year, month and day of birth, and so forth is rather limited. A dictionary-based attack, therefore, would quickly re-identify the plain-text source for these fields, along with most of the given names and family names that also could

be determined by means of a dictionary attack.

- **Encryption.** In the final step, each message digest is encrypted with a secret encryption key that needs to be known by all parties in the protocol that generate control numbers to be compared with each other. Each encrypted message digest is called a control number, and the complete set of control numbers describes the patient demographics in a form that does not allow for the reproduction of any plain-text field by any party that does not have access to the encryption key, but it allows for a comparison of two different sets of control numbers without the need for access to the encryption key.

Given a set of control numbers describing a query and a larger number of sets of control numbers describing all patients in a record repository, matches in the repository can be identified using *record linkage*, defined by Winkler (1999) as "the methodology of bringing together corresponding records from two or more files or finding duplicates within files." We propose the use of a probabilistic rather than a deterministic record linkage algorithm, as described by Jaro (1989) as well as Blakely and Salmond (2002). This class of algorithms not only compares each pair of control numbers for equality, but it also considers the significance of each control number, based on an estimation of the true positive rate and the false positive rate; that is, the probabilities of a pair of identical control numbers of representing the same or different patients. The probabilistic record linkage also allows one to compensate to a certain degree for missing data (control numbers that are not available at the record repository), typing errors, and so forth. Basically, the algorithm allows a person to identify the most promising matches from a larger set. The final choice of records that would be treated as definitive matches depends on a threshold and possibly on human choice. Further issues in record linkage, such as array comparison, are discussed in Aden, Eichelberg, and Thoben (2004).

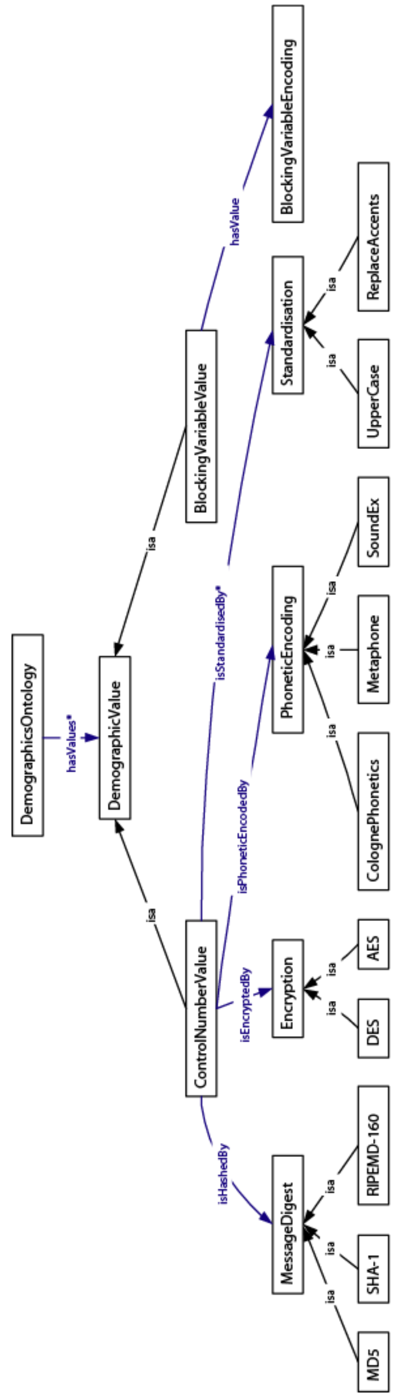
Blocking Variables

The protocol as previously described would require every record repository to create a set of control numbers for each patient in the repository and to repeat this for each incoming query because of the encryption key, which would change with each request. This, of course, is not practical for a large number of patients. Therefore, the protocol uses the concept of so-called *blocking variables* to reduce the number of possible candidate records for which control numbers have to be computed and evaluated. Blocking variables are simply plain-text demographics that are transmitted in unencrypted form (a transport level encryption can be used, of course, but the blocking variables are available to the record repository in plain-text form). Blocking variables must be chosen carefully to make sure that a patient cannot be identified from the blocking variables alone. As an example, the day of birth (not including month and year) certainly would not be sufficient to identify any specific patient but would reduce the number of candidates for which control numbers would have to be computed by a factor of up to 30. Similarly, a postal code of the place of residence or the place of birth would not be sufficient to identify a patient, even when combined with the day of birth, but would reduce significantly the number of candidates in most record repositories (except for hospitals actually located at the given place, where a large number of patients may have the same place of birth or residence). The set of blocking variables needs to be chosen carefully in order to reduce the number of candidate records as much as possible while avoiding the risk of exposing the patient's identity.

Semantic Annotation

As previously described, we can expect different countries to use different, though certainly overlapping, sets of control numbers accounting for country- or region-specific aspects such as phonetic encoding or national unique patient identifiers. Since control numbers only can be compared for binary equality but not evaluated in any other way, it is of prime importance for all parties participating in the protocol to understand exactly what each control number (and each blocking variable) means and which control number is supported by which party. The use of ontology-based semantic annotation allows one to introduce the amount of flexibility into the protocol that is needed to make it work in an international setting, where different sets of control numbers might be supported by different actors. Each request or response dataset consisting of a list of control numbers and blocking variables is encoded as a Web ontology language (OWL) (Dean & Schreiber, 2004) instance that describes the demographics from which each entry was generated, as well as the processing steps that were applied to the entry. Figure 2 shows the core demographics ontology defined for this purpose. A *DemographicsOntology* used by a health care enterprise to describe the supported set of control numbers and blocking variables consists of a set of *DemographicsValue* entities that are either control numbers or blocking variables (the asterisk in the *hasValues* relationship type denotes a one-to-many relationship). For control numbers, the processing steps are described through properties; for blocking variables, the *BlockVariable Encoding* describes the standardisation process or value range. Since different countries typically will use different sets of control numbers, more than one *DemographicsOntology* might exist. The country-, region-, or hospital-specific demographics ontologies (called local ontologies) are defined as extensions to the core ontology. Figure 3 shows a subset of such a local ontology, describing a

Figure 2. Demographics ontology describing control numbers and blocking variables

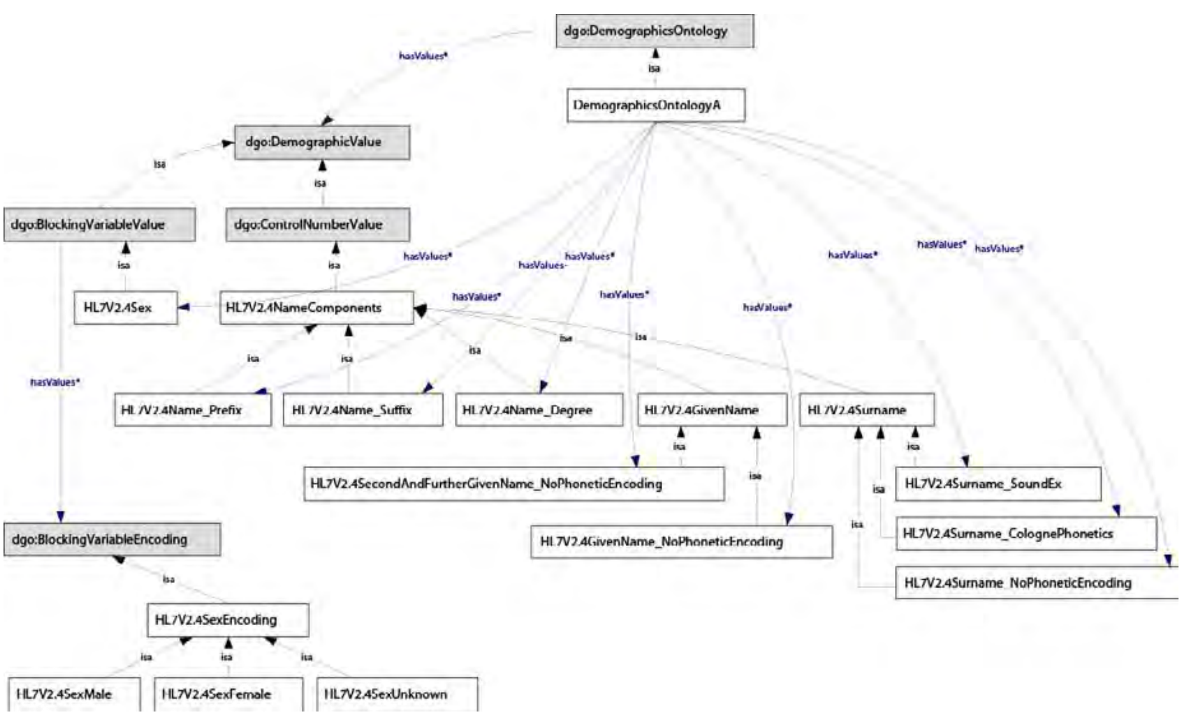


single blocking variable (the patient's sex) and a set of control numbers for the name components as defined in HL7. All ontology classes shown in grey with the prefix *dgo:* in their name refer to the core ontology. OWL restrictions on properties that are not shown in this figure define the algorithms used in the local ontology for standardisation, phonetic encoding, message digest, and encryption. Also not shown is a binary property *isUnique* defined for each subclass of *DemographicValue* that denotes control numbers that guarantee a unique identification of the patient for a match. These properties are used to initialise the weights of the record linkage process accordingly. If the requestor and a record repository in the protocol use different local ontologies, then a direct or indirect mapping between these ontologies is used, which is available both to the record repository (for the purpose of interpreting the blocking variables) and the record linkage server, which is introduced next. A mapping would be defined as an ontology describing *owl:equivalentClass* relationships between ontology classes in the different ontologies, allowing for a reconciliation between the ontology instances using rather simple inference rules. Instance nodes (i.e., individual control numbers or blocking variables) for which no equivalence in the other local ontology can be found simply are omitted from the record linkage process.

Knowledge Distribution

As described, the security of the control numbers is based on the combination of a message digest algorithm followed by an encryption with a session key. The session key, which is randomly generated for each query, guarantees that responses from different queries cannot be compared in order to derive information from the evaluation of a large

Figure 3. Local ontology describing control numbers and blocking variables for one country or institute



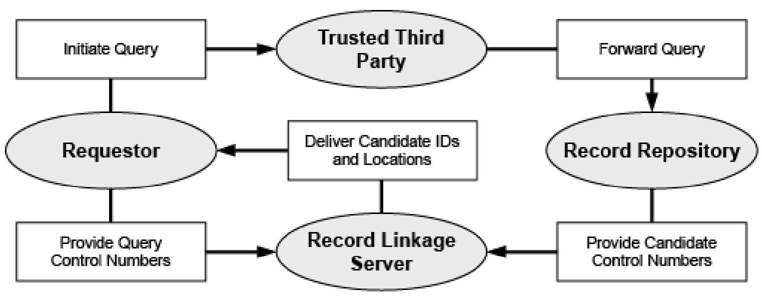
set of queries. However, the problem of dictionary attacks still remains: Any entity that has access to both a set of control numbers and the session key could attempt to decrypt the control numbers and use a dictionary of names, dates, and so forth to re-identify the patient described by the control numbers. In order to prevent this, the following simple policy needs to be established:

Any entity in the network that has access to the session key must not be granted access to any set of control numbers except the control numbers generated by this entity itself.

This means that the initiator of a query for a specific patient (called the *requestor*) cannot receive the control numbers generated by the record repositories responding to the query. The record repositories, on the other hand, cannot be given access to the control numbers in the query, because they need to access the session key in order to generate their own sets of control numbers for all patients matching the blocking variables. Fortunately, the record linkage process does not require access to the session key — only to the control numbers in the query and the responses that were pre-filtered by the record repositories using the blocking variables. Therefore, the protocol needs to introduce additional actors between requestor and record repositories in order to guarantee that no patient-related information is leaked through the protocol before explicit authorisation has been granted. The actors and their transactions in a slightly simplified form are shown in Figure 4. The complete protocol is detailed in Eichelberg, Aden, and Riesmeier (2004). Figure 4 shows the actors involved in the protocols as ovals and the transactions (i.e., message exchange between them) as boxes with an arrow indicating the direction of the information flow. Only a single record repository is shown in the diagram; however, the protocol is designed to work with a potentially large number of record repositories and possibly with more than one instance of the other actors involved. The two additional actors involved are as follows:

- **Record Linkage Server (RLS).** An entity in the network that receives a set of control numbers from the requestor describing the patient to be identified and a possibly large number of sets of control numbers generated by the record repositories, based on the blocking variables and the session key. The RLS performs the probabilistic record linkage and identifies a number of candidates; that is, patient records in the record repositories that have a high probability of referring to the same patient described by the query control numbers. Once the RLS has completed its work and authorisation has been given by the record repositories (see discussion that follows), the set of best matches (based on a threshold value) is reported back to the requestor completing the patient identification process. Since the RLS

Figure 4. Actors and transactions in the Patient Identification Protocol (simplified)



does not have access to the session key, it cannot re-identify the patient demographics from the control numbers it receives. The RLS must be a trusted component in the network in the sense that both requestor and record repositories have to rely on the RLS to never give away any control numbers it receives and never trying to obtain a session key in order to perform a re-identification of patient demographics.

- **Trusted Third Party (TTP).** An entity in the network that enables distribution of the query containing the session key and blocking variables among the record repositories in the network, preventing any direct communication between requestor and record repositories at this stage. The TTP also can act as a gatekeeper to the network in the sense that only requestors accepted by the TTP are able to initiate patient identification queries, and only repositories accepted by the TTP are able to provide information to the RLS. Each match retrieved by the requestor from the RLS consists of a resource identifier (e.g., URL), under which the record repository that has provided this matching record can be contacted, a set of flags indicating for each of the control numbers in the query whether or not they matched or were missing in the response, and a candidate identifier, a simple numeric identifier generated by the record repository as an alias to the local patient ID that identifies a specific patient in the context of the current query transaction and can be used to request medical records for this specific patient.

Medical Record Access

Once a requestor has identified one or more record repositories that have clinical documents pertaining to the current patient available, the appropriate documents have to be selected, requested, retrieved and displayed. This comprises the second phase of the protocol described in this article. Since no global standard for electronic health care records exists at this time that could be used to provide record access in the cross-border scenario previously described, we suggest using the Retrieve Information for Display (RID) (IHE, 2004) protocol defined by the Integrating the Healthcare Enterprise (IHE) initiative (Eichelberg, Poiseau, Wein & Riesmeier, 2003) for this purpose. IHE is a non-profit initiative addressing the issue of system integration in health care with strong participation from industry, science, and medical professional societies in North America, Europe and Japan. While IHE does not develop standards as such, it selects and recommends appropriate standards for specific use cases and develops restrictions (application profiles) for these standards that allow for a simplified system integration. The result of this technical work is published as the *IHE Technical Framework* and revised annually. The *IHE Technical Framework for IT Infrastructure* (IHE, 2004) gives the following short overview about the purpose of the RID integration profile:

The Retrieve Information for Display Integration Profile (RID) provides simple and rapid read-only access to patient-centric clinical information that is located outside the user's current application but is important for better patient care (for example, access to lab reports from radiology department). It supports access to existing persistent documents in well-known presentation formats such as HL7 CDA (Level 1), PDF, JPEG, etc. It also supports access to specific key patient-centric information such as allergies, current medications, summary of reports, etc. for presentation to a clinician.

IHE interface definitions are based on actors (the IT systems involved in the protocol) and transactions (interfaces between actors). In the case of RID, the actors are called *information source* and *display*. The information source, which corresponds to the record repository in the previous discussion, is a system that maintains a database of persistent clinical documents and specific key patient-centric information such as allergies, current medications, summary of reports, and so forth. The display, which corresponds to the requestor, is a system that accesses the information source, retrieves patient-centric information or persistent documents, and displays them to a human observer. The focus of the integration is visual presentation, not a complete integration of the structured databases on which the actors might be based. Documents are exchanged in well-known presentation formats. Communication between display and information source is always initiated by the display and implemented as a Web Service using the Web Services description language (WSDL) with a binding to HTTP GET.

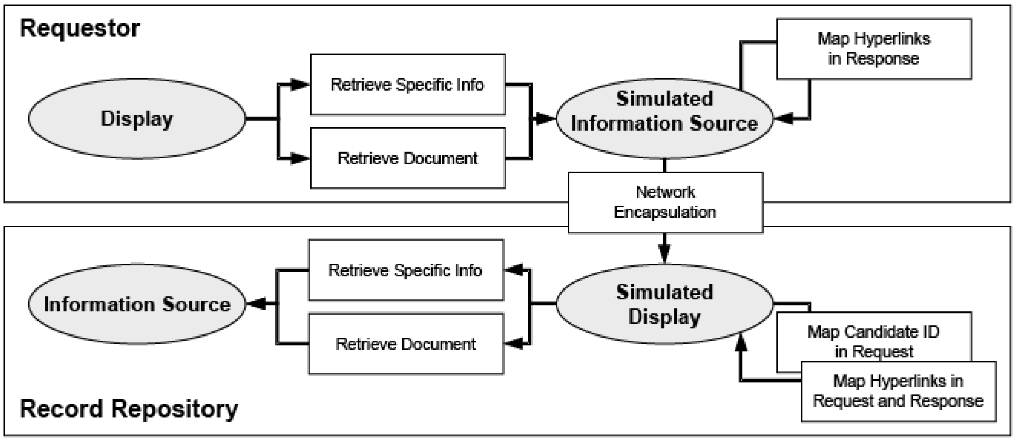
The RID protocol is ideal for integration with the patient identification process described before, because the only *a priori* information that the display needs when accessing a record repository is the patient identifier. This identifier is provided by the patient identification protocol in the form of the candidate identifier, which describes a single patient and is only valid in the context of a single query transaction. Figure 5 shows how an existing RID information source at the record repository site and an existing RID display system at the requestor site can be combined with an additional layer of middleware to provide medical record access in combination with the patient identification protocol.

On the repository site, a simulated display receives the incoming requests from the requestor, replaces the candidate ID generated during the patient identification process by the real patient ID locally used in the hospital, and forwards the request to the existing information source, to which this actor looks like any local display actor. Before forwarding the response to the requestor, all hyperlinks in the response have to be replaced with links pointing to the simulated display, because these hyperlinks typically will refer to resources that cannot be reached directly from the outside world (due to firewall protection and network address translation). Subsequent requests for these mapped hyperlinks simply would be redirected from the simulated display to the original resource. In a similar manner, a hyperlink lookup scheme can be implemented on the requestor site inside the simulated information source to address firewall and network address translation issues. As a side effect, the two simulated actors can provide a secure tunneling of the request and response message, which is usually transmitted in clear text form within a hospital, using any available transport level security protocol. Figure 6 shows the combined sequence of transactions for the patient identification protocol and the following medical record access. In this figure, the PID requestor from Figure 4 and the RID display from Figure 5 have been grouped into a single system providing the user interface to the end user. This is the typical use case and also reflects the ARTEMIS implementation. The scalability aspect of the protocol (i.e., involvement of multiple TTPs and repositories) is not shown for reasons of clarity. The grouped requestor/display actor and the simulated information source are part of the IT infrastructure of the user's site. The TTP and the RLS are central components provided by the network, and all other actors are located at the repository site.

Data Protection and Patient Consent

The EU data protection directive (Council of the European Communities, 1995) and the recommendation of the Council of Europe on the protection of medical data (Council of Europe, 1997) determine that neither person identifying data nor medical records are allowed to be exchanged between different organisations unless either the patient agrees on the exchange of the identifying data and medical records for a specific purpose or the vital interests of the patient are touched.

Figure 5. Access to medical records based on IHE retrieve information for display



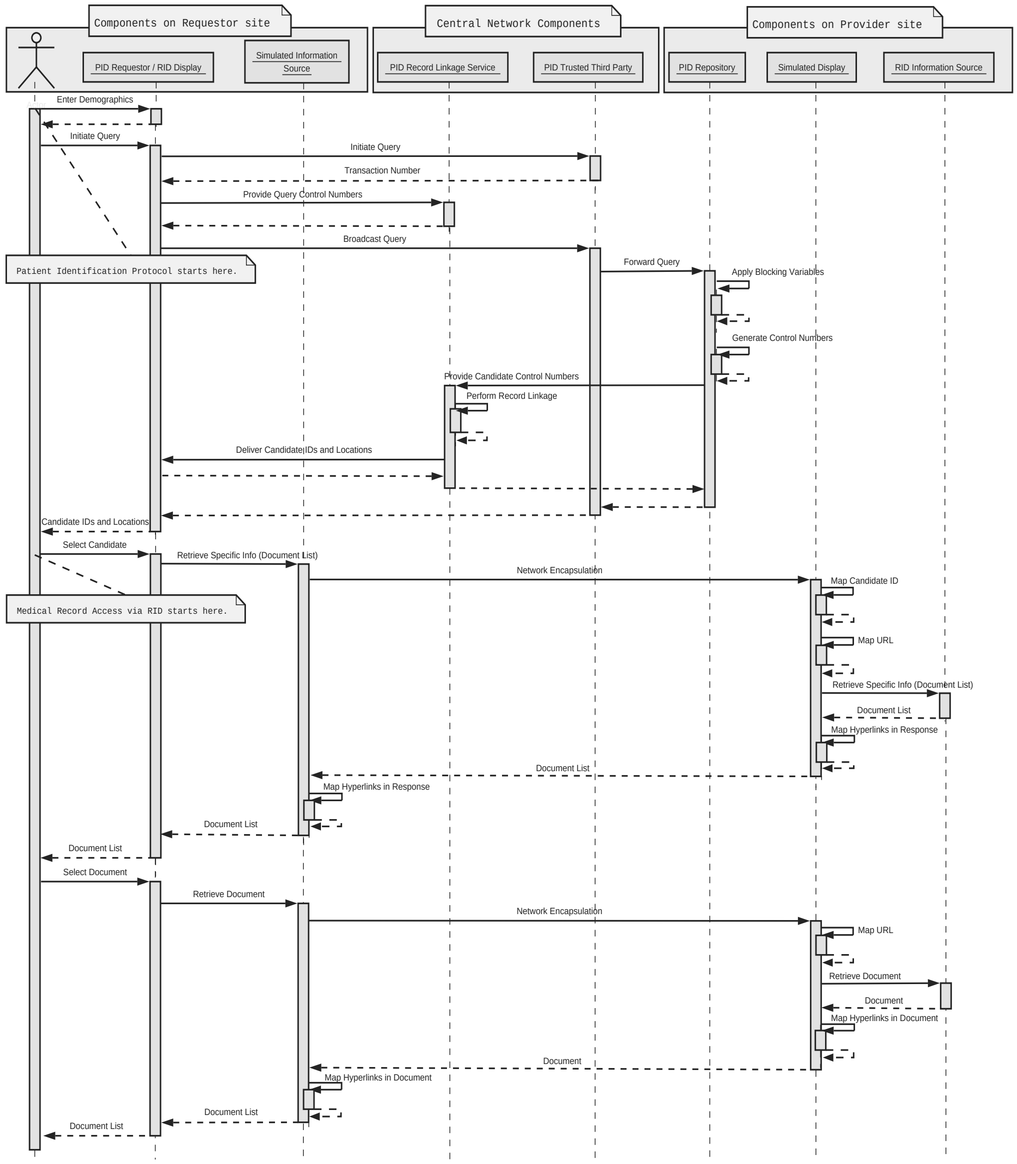
Therefore, in most cases, an explicit patient consent will be a prerequisite to an exchange of medical records.

Since the patient is most likely to be at the health care institute searching for and requesting to access clinical records (i.e., the requestor), the question arises how the patient consent can be demonstrated to the record repository; in particular, if the exact location of the repository is unknown. It should be noted that no information that would allow one to identify a patient or derive any information about a patient is exchanged in the patient identification protocol, until the RLS provides a list of matches to the requestor that is in the last step of the first phase of the protocol depicted in Figure 4. This list of matches enables the requestor to identify resources (i.e., health care institutes) that have clinical documents for the given patient available with a certain probability. The second more significant exchange of information takes place during the second phase of the protocol when the requestor uses the modified Retrieve Information for Display protocol to request and access clinical documents for a specific patient. Correspondingly, there are two places at which additional barriers can be installed in the protocol, barriers that will be opened automatically or manually once the patient consent has been demonstrated to the record repository:

- The requestor may deposit a proof of the patient consent at the RLS when initiating the query. The RLS would make this document available only to those record repositories that have been identified as matches in the record linkage and request permission to report them back to the requestor. Only repositories granting this permission would ever be reported to the requestor.
- The RID access to clinical documents for a specific candidate may require additional authorisation at the repository site requiring a prior communication of the proof of patient consent. It also would be possible to restrict access within the RID protocol to specific documents or document types (e.g., radiology reports).

Two different approaches for demonstrating patient consent are conceivable — the use of a public-key infrastructure (PKI) for patients or a conventional out-of-band communication using possibly digitised letters, telephone, mail, or fax conversation. A PKI for patients would allow for a solution of the issue of patient consent in a fully digital manner by providing a digital signature of the patient authorising the request. This, however, would require that certificates and security

Figure 6. Sequence of transactions for patient identification and medical record access



tokens for the private key, such as smartcards, be issued for all patients. While a few countries, including Belgium, Taiwan, and Germany, have plans for a PKI for the health care sector (Dietzel & Riepe, 2004), there are no such plans in many other countries, and there is no sign of a global or at least a European-wide harmonisation of these PKIs (the European health insurance card does not include PKI functionality).

Since a PKI-based solution will not be available in the foreseeable future for cross-border communication, a conventional approach needs to be supported. As an example, the patient may be asked to sign a letter of consent at the requesting health care institute, and a digitised copy of this letter could be deposited with the RLS and made available to the record repositories for which matches were found during the record linkage process. The data protection officer at each health care institute operating a record repository would check the letter of consent and, if appropriate, give permission to the RLS to report the location of the record repository back to the requestor and grant access to certain clinical documents. In cases where an original handwritten signature is legally required, the digitised document could be accompanied by a contact address at the requesting health care institute where a delivery of an original document (e.g., by express service) could be requested. While it is not realistic to assume a working PKI for patients as a prerequisite to a digital cross-border exchange of clinical records, a working PKI for the health care institutes connected to the network could be implemented rather easily, again utilising the gatekeeper role of the TTP entity in the network. This would assure that only authenticated entities could participate in the protocol either as a requestor or as a record repository and, therefore, would provide a minimum level of quality control of the data communicated within the network.

RESULTS

The patient identification protocol establishes a concept that, to our knowledge, has not been available in the health care sector before — an undirected search for patient records (i.e., a search that does not require *a priori* information about the location of the records) that does not violate data protection requirements. The concept matches well with the peer-to-peer network structure established in the context of the ARTEMIS project, where many health care institutes may form a network of services in a very dynamic manner. The use of semantic annotation allows one to cope with the fact that different institutes identify patients with different demographics and that no globally unique identifier for patients will be available for the foreseeable future. The possibility of including national extensions allows one to maximise the specificity of the search algorithm in environments that can make use of such extensions, while the inclusion of phonetic encoding allows one to improve sensitivity by gracefully handling spelling errors. The integration of the Retrieve Information for Display protocol for the second phase (i.e., the request and transmission of clinical documents once their location has been identified) provides a simple read-only access to clinical documents that easily can be integrated with the existing legacy systems.

An implementation of the patient identification protocol is currently under development in the context of the ARTEMIS project. The implementation is based on the ARTEMIS middleware, which provides for the peer-to-peer infrastructure required by the protocol; that is, message transmission and broadcast facilities based on the JXTA peer-to-peer communication platform (Gong, 2002), a policy-based security architecture for a secure communication over unsecured networks and a pre-selection of record repositories that participate in a query based on geographical location (e.g., "we are looking for records located in Germany only"). Pilot applications that will evaluate the protocol in a clinical setting are planned in the UK and in Turkey. From these prototypes, we expect practical experience about use patterns, performance, and scalability of the protocol. Krieg, Hense, Lehnert, & Mattauch (2001) report about an evaluation of the probabilistic record linkage based on control numbers, which is used in the Münster Cancer Registry in Germany. They use a set of 19 control numbers computed from the patient's name (including name of birth, prior name, and phonetic encoding of the name components) and the day of birth. The clear-text blocking variables comprise the month and year of birth, the patient's sex, and place of residence. They evaluated 27,262 record linkage processes performed in the year 1998 for a database of 101,880 patients. In this case, the record linkage provided a false positive rate (i.e., patients that were incorrectly selected as a match) of 0.36% and a false negative rate (i.e., matches in the database that were not found) of 1.81%. Krieg et al. (2001) emphasise that these figures depend significantly on the size of the database. Nevertheless, the figures indicate that the patient identification protocol provides good results even with a rather limited set of demographics available. It should be noted that in countries that have a national patient identifier available, the record linkage process would be able to provide 0% false positive and false negative rate in repositories that support the national patient ID while still providing results in the order of magnitude shown above for other repositories (i.e., other countries). Additional work will be needed to improve the internationalisation of the splitting and standardisation algorithms for control number generation, which currently are based mainly on German experiences with epidemiological cancer registries. Fine-tuning may be needed for the parameters of the record linkage process in order to maximise specificity and sensitivity of the record linkage process.

CONCLUSION

The patient identification protocol described in this article provides a solution for a common problem in the health care sector that is likely to become very important with the increasing mobility of the workforce in Europe—locating and accessing prior clinical records for the continuity of care. The solution combines techniques from different domains—control numbers, blocking variables, and record linkage procedures as used in epidemiological registries, knowledge distribution, and TTP services from cryptographic communication protocols, semantic annotation, and ontology-based mediation, core technologies of the Semantic Web. The protocol reflects the way clinical records typically are stored, indexed, and accessed in today's health care information systems. While the advent of a universally accepted standard for electronic health care records possibly could render the patient identification protocol obsolete in the future, there is no indication at the moment that such a standard is going to be established in the foreseeable future, and even then, the issue of historical records would remain for a number of years. Therefore, the authors believe that the patient identification protocol can be a significant help in improving access to clinical information while safeguarding data protection rules and protecting the patient's right of self-determination regarding his or her clinical records.

ACKNOWLEDGMENT

This work is supported by the European Commission through the IST-1-002103-STP ARTEMIS R&D project.

REFERENCES

Aden, T., Eichelberg, M., & Thoben, W. (2004). A fault-tolerant cryptographic protocol for patient record requests. In *Proceedings of EuroPACS-MIR 2004 in the Enlarged Europe* (pp. 103-106). Trieste, Italy: E.U.T. Edizioni Università di Trieste.

ARTEMIS Consortium. (2004). The ARTEMIS Project. Retrieved from http://www.srdc.metu.edu.tr/webpage/projects/artemis/

Blakely, T., & Salmond, C. (2002). Probabilistic record linkage and a method to calculate the positive predictive value. *International Journal of Epidemiology, 31*, 1246-1252.

CEN/ISSS eHealth Standardization Focus Group. (2004). *Current and future standardization issues in the e-Health domain: Achieving interoperability* (Draft 4.1). Geneva, Switzerland: European Committee for Standardization/Information Society Standardization System.

Council of Europe. (1997). *Recommendation No. R(97)5 of the Committee of Ministers to Member States on the protection of medical data* (Tech. Rep.). Strasbourg, France: Council of Europe Publishing.

Council of the European Communities. (1995). Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal of the European Communities, L281*, 31-39.

Dean, M., & Schreiber, G. (2004). *OWL Web ontology language reference* (W3C Recommendation). Cambridge, MA: World Wide Web Consortium.

Dietzel, G., & Riepe, C. (2004). Modernizing health care in Germany by introducing the eHealthcard. *Swiss Medical Informatics,* 52, 18-22.

Dogac, A., et al. (in press a). Exploiting ebXML registry semantic constructs for handling archetype metadata in healthcare informatics. *International Journal of Metadata, Semantics and Ontologies*.

Dogac, A., et al. (in press b). Artemis: Deploying semantically enriched Web Services in the healthcare domain. *Information Systems Journal*.

Eichelberg, M., Aden, T., & Riesmeier, J. (2004). *Relevant electronic healthcare record standards and protocols for accessing medical information* (ARTEMIS deliverable D5.1.1). Oldenburg, Germany: ARTEMIS Consortium.

Eichelberg, M., Poiseau, E., Wein, B. B., & Riesmeier, J. (2003). IHE Europe: Extending the IHE initiative to the European health care market. In H. K. Huang, & O. M. Ratib (Eds.), *Medical imaging 2003: PACS and integrated medical information systems: Design and evaluation* (Vol. 5033) (pp. 1-10).

Gong, L. (2002). *Project JXTA: A technology overview* (Tech. Rep.). Palo Alto, CA.: Sun Microsystems, Inc.

HL7. (2003). *Application protocol for electronic data exchange in health care environments* (Version 2.5) (ANSI Standard). Ann Arbor, MI: Health Level Seven.

IHE. (2004). *IT infrastructure technical framework revision 1.1* (Tech. Rep.). Oak Brook, IL: Integrating the Healthcare Enterprise.

ISO. (1999a). *Information processing: 8-bit single-byte coded graphic character sets. Part 13: Latin alphabet No. 9* (International Standard No. 8859-13). Geneva, Switzerland: International Organization for Standardization.

ISO. (1999b). *Information processing: 8-bit single-byte coded graphic character sets. Part 1: Latin alphabet No. 1* (International Standard No. 8859-1). Geneva, Switzerland: International Organization for Standardization.

Jaro, M. A. (1989). Advances in record linkage methodology as applied to matching the 1985 census of Tampa, Florida. *Journal of the American Statistical Association*, *84*(406), 414-420.

Krieg, V., Hense, H.-W., Lehnert, M., & Mattauch, V. (2001). Record linage mit kryptographierten Identitätsdaten in einem bevölkerungsbezogenen Krebsregister. *Gesundheitswesen, 63*, 376-382.

NEMA. (2004). *Digital imaging and communications in medicine (DICOM)* (NEMA Standards Publication No. PS 3.x–2004). Rosslyn, VA: National Electrical Manufacturers Association.

Postel, H.-J. (1969). Die Kölner Phonetik — Ein Verfahren zur Identifizierung von Personennamen auf der Grundlage der Gestaltanalyse. *IBM-Nachrichten, 19*, 925-931.

Thoben, W., Appelrath, H.-J., & Sauer, S. (1994). Record linkage of anonymous data by control numbers. In W. Gaul, & D. Pfeifer (Eds.), *From* *data to knowledge: Theoretical and practical aspects of classification, data analysis and knowledge organisation, Oldenburg* (pp. 412-419). Heidelberg, Germany: Springer Verlag.

Winkler, W. E. (1999). *The state of record linkage and current research problems* (Tech. Note). Washington, DC: U. S. Bureau of the Census.

This work was previously published in International Journal on Semantic Web & Information Systems, Vol. 1, No. 4, edited by A. Sheth & M.D. Lytras, pp. 24-43, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 2.23
Image Mining for the Construction of Semantic-Inference Rules and for the Development of Automatic Image Diagnosis Systems

Petra Perner
Institute of Computer Vision and Applied Computer Sciences (IBal), Germany

ABSTRACT

This chapter introduces image mining as a method to discover implicit, previously unknown and potentially useful information from digital image and video repositories. It argues that image mining is a special discipline because of the special type of data and therefore, image-mining methods that consider the special data representation and the different aspects of image mining have to be developed. Furthermore, a bridge has to be established between image mining and image processing, feature extraction and image understanding since the later topics are concerned with the development of methods for the automatic extraction of higher-level image representations. We introduce our methodology, the developed methods and the system for image mining which we successfully applied to several medical image-diagnostic tasks.

INTRODUCTION

The increasing number of digital-image and video repositories have made image mining an important task. Image mining means a process of nontrivial extraction of implicit, previously unknown and potentially useful information from image databases. The application of image mining will help to get some additional knowledge about specific features of different classes and the way in which they are expressed in the image. This method can elicit nonformalized expert knowledge; it can automatically create effective models for decision-making, and can help to find some inherent non-evident links between classes and their imaging in the picture. It can help to get some nontrivial conclusions and predictions on the basis of image analysis. The new knowledge obtained as a result of data analysis in the database can enhance the professional knowledge of the expert or the user of the image-database. This knowledge can also be used for teaching novices or can support image analysis and diagnosis by the expert. It can be used for semantic annotation of digital visual content to enable sophisticated semantic querying of the media in terms familiar to the user's domain, whilst also ensuring that the information and knowledge have a much greater chance of being discovered and exploited by services, agents and applications on the Web.

An additional advantage of image-mining application in decision-making of medical or other tasks is on the long-run the opportunity of creating fully automatic image-diagnosis systems that could be very important and useful in the case of lacking knowledge for decision-making.

In this chapter we present our methods and methodology for performing image mining. We describe the recent state of the art in image mining, the questions that can be answered by applying image-mining methods to image databases, and the problems concerned with image mining. A design of image-mining tools is considered followed by a presentation of our methods and the developed tool for image mining. A methodology for image mining that was created and tested in the task of Hep-2 cell analysis is described. Finally, we summarize our experience in applications of image-mining methodology in different medical tasks, such as pre-clinical diagnosis of peripheral lung cancer on the basis of lung tomograms, lymph-node diagnosis and investigation of breast diseases in MRI and the inspection of microscopic images of cell-based assays. Conclusions and plans for future work are given at the end of this chapter.

BACKGROUND

As in data mining, we can classify image mining into two main problem types: prediction and knowledge discovery. While prediction is the strongest goal, knowledge discovery is the weaker approach and usually occurs prior to prediction. In prediction we want to discover a model that allows us, based on the model, to predict new data in the respective classes. In knowledge discovery we want to discover similar groups of database entries, frequent patterns, and deviations from a normal status or just relations among the database entries.

Image mining differs from data mining in respect of the data and the nature of the data. The raw image is of 2-dimensional or 3-dimensional numerical data type. Videos are temporal sequences of numerous 2-d images. The image information can be represented by the 2-d or 3-d image matrix itself, by low-level features such as edges, blobs and regions, or by high-level features that allow a human to semantically understand the image content. The different data types in which an image or an image sequence can be represented and the resulting need for special methods and techniques offer a data-type dimension to data mining and make image mining a specific field.

Consequently, image mining can be applied to all the different data representations of an image. Which of the data representations is used, usually depends on the question under study. If we mine our images for knowledge that can be used to construct a fully automatic image-interpretation system, we have different questions to answer: We have to mine the images for regions-of-interest and separate them from the background of the image. Once we have found these regions we need to mine them for distinguishing features and later on we are interested in discovering rules that allow us to classify these regions into different patterns. Based on all this information discovered we might be able to build an automatic image-interpretation system.

In video mining it is of interest to detect scenes, group them into similar groups or detect events. These three main tasks require a specific image-mining procedure applicable to videos.

Unlike in other fields we can only do this automatically if we have proper automatic information-extracting methods. This brings us to the topics of image processing, feature extraction and image understanding. Image mining is closely related to these topics and will only succeed if a bridge can be established between image mining and these topics.

All the above mentioned questions can inherently suffice in other applications. The problem of searching the regions of special visual attention or interesting patterns in a large set of images has been studied for medical CT and MRI image sets in Megalooikonomou, Davatzki, and Herskovits (1999) and Eklund, You, and Deer (2000) or in satellite images (Burl & Lucchetti, 2000). Usually, experienced experts have discovered this information. However, the amount of images, which is being created by modern sensors, makes the development of methods necessary that can tackle this task for the expert. Therefore, standard primitive features able to describe the visual changes in the image background are being extracted from the images, and the significance of these features is being tested by a sound statistical test (Burl & Lucchetti, 2000; Megalooikonomou, Davatzki, & Herskovits, 1999).

Clustering is applied in order to explore the images seeking similar groups of spatial connected components (Zaiane & Han, 2000) or similar groups of objects (Eklund, You, & Deer, 2000). Association rules are used for finding significant patterns in the images (Burl & Lucchetti, 2000). A method for obtaining the principal objects, characters and scenes in a video by measuring the reoccurrence of spatial configurations of viewpoint invariant features is described in Sivic and Zisserman (2004). How to discover the editing patterns of videos from different editors with data mining and to use the discovered rules for editing new video material is described in Matsuo, Amano, and Uehara (2002). Multilevel sequential association mining is applied to explore associations among the audio and visual cues in Zhu, Wu, Elmagardmid, Feng, and Wu (2005). The associations are classified by assigning each of them with a class label. Their appearances in the video is recognized and used to construct video indices. A method for unsupervised classification of events in multi-camera indoors surveillance video by applying self-organizing map (SOM) is described in Petrushin (2005).

The measurement of image features in these regions or patterns provides the basis for pattern recognition and image classification. Computer-vision researches are conducted to create proper models of objects and scenes, to obtain image features and to develop decision rules that allow one to analyze and interpret the observed images. Computer-assisted diagnosis methods of image processing, segmentation, and feature measurements are successfully used for this purpose (Kehoe & Parker, 1991; Perner, 1998; Schröder, Niemann, & Sagerer, 1988).

The basic research of image mining and machine learning in pattern recognition deals with a wide spectrum of different problems that need to be solved for image mining. The developed methods reach from learning local structural features by a multiscale relevance function (Palenichka & Volgin, 1999) to the construction of artificial neural nets for estimating the local mean grey value for image processing (Jahn, 1999) or for model selection and fitting if the models are linear manifolds and data points distribute to the union of a finite number of linear data points (Imiya & Oatani, 2001). Multiple classifier systems and case-based reasoning (Perner, 1998, 2005; Schmidt & Gierl, 2001) are further approaches for prediction models. Case-based reasoning methods for image segmentation (Perner, 1999) as well as for the recognition of similar video scenes (Zhu, Wu, Elmagardmid, Feng, & Wu, 2005) have been developed. Multiple classifier systems can be designed by applying unsupervised learning to ensure that the multiple classifiers works efficiently by making independent errors (Giacinto & Roli, 1999). First-order rule induction (Malerba, Esposito, Lanza, & Lisi, 2001) is applied for the recognition of morphological patterns based on symbolic terms in topographic maps. Methods for clustering with relevance feedback (Bhanu & Dong, 2001) have been developed to overcome the gap between low-level visual features and human high-level concepts and rule-based ensembles are applied to solve regression problems (Indurkhya & Weiss, 2001). The incremental appearance of most media is now considered by incremental clustering techniques (Bougila & Ziou, 2005; Jänichen & Perner, 2005). Sampling schedules for face recognition and finger print recognition based on neural nets that are adaptive to the data and allow using only a portion of data from a huge dataset for the development of recognition models are developed in Satyanarayana and Davidson (2005). These few examples show that image mining is still a basic research topic with a lot of different subtopics, but on the other hand systems for image mining are needed that can be used in practice.

The mining process is often done bottom-up. As many numerical features as possible are extracted from the images, in order to achieve the final goal—the classification of the objects (Fischer & Bunke, 2001; Perner, Zscherpel, & Jacobsen, 2001). However, such a numerical approach usually does not allow the user to understand the way in which the reasoning process has been achieved.

The second approach to pattern recognition and image classification is an approach based on the symbolic description of images made by the expert (Perner, 2000). This approach can present to the expert in an explicit form the way in which the image has been interpreted. The experts having the domain knowledge usually prefer the second approach.

Normally simple numerical features are not able to give a description of complex objects and scenes. They can be described by an expert with the help of non-formalized symbolic descriptions, which reflect some gestalt in the expert domain knowledge. This task of semantic tagging becomes more popular since such descriptions enable sophisticated semantic querying of the media in terms familiar to the user's domain, whilst also ensuring that the information and knowledge have a much greater chance of being discovered and exploited by services, agents and applications on the Web. In contrast to that it is also the basis for the development of automatic image interpretation systems.

One problem is how to find out the relevant descriptions of the object (or the scene) for its interpretation and how to construct a proper procedure for the extraction of these features. This top-down approach is the more practical approach for most applications. However, the symbolic description of images and feature estimation face numerous difficulties:

1. A skilled expert knows how to interpret the image, but often has no well-defined vocabulary to describe the objects, visual patterns and gestalt variances, which are standing behind the expert's diagnostic decisions. When the expert is asked to make this knowledge explicit, the expert usually cannot specify and verbalize it.
2. Although numerous efforts are going on to develop such a vocabulary for specific medical tasks (for example, the ACR-BIRADS-code has been constructed for image analysis in mammography) and MPEG-7 standard (Martinez, 2001) is used for ontology-based annotations of natural images, the problem of the difference between "displaying and naming" still exists.

3. A developed description language will differ for example from medical school to medical school and as a result the obtained symbolical description of image features by a human will be expert-dependent and subjective.
4. Besides this, the developed vocabulary usually consists of a large number of different symbolic features (image attributes) and feature values. It is not clear a-priori whether all the attributes included into the vocabulary are necessary for the diagnostic reasoning process. To select the necessary and relevant features would make the reasoning process more effective.

We propose a methodology of image mining that allows one to learn a compact vocabulary for the description of objects and to understand how this vocabulary is used for diagnostic reasoning or in semantic-inference rules of semantic tagging. Besides that we also use basic image features that have been calculated from the image. This methodology can be used for a wide range of image-diagnostic tasks and semantic annotations.

Developed methodology takes into account the recent status of the art in image analysis and feature extraction and combines it with new methods of data mining. It allows us to extract quantitative information from the image and when possible combine it with subjectively determined diagnostic features, and then to mine this information for the relevant diagnostic knowledge acquisition by objective methods such as data mining.

Our methodology should help to solve some cognitive, theoretical and practical problems:

1. It will reproduce and display a decision model of an expert for specific task solutions.
2. It will show the pathway of human reasoning and classification. Image features will be discovered which are basic for correct decision-making by the expert.
3. A developed model will be used as a tool to support the decision-making of a human who is not an expert in a specific field of knowledge. It can be used for teaching novices, as well as for semantic image retrieval.

The application of data mining will help to get some additional knowledge about specific features of different classes and the way in which they are expressed in the image. It could help to find some inherent non-evident links between classes and their imaging in the picture that could be used to make some nontrivial conclusions and predictions on the basis of elicited knowledge.

DESIGN CONSIDERATIONS

Besides the development of new methods for image mining, there is a need for systems that can support a user at all steps of an image-mining task. We developed a tool for data mining which was to meet several requirements:

1. The tool had to be applicable for a wide range of image-diagnostic tasks and image modalities that occur for example in medical practice. It was to be used for semantic annotation of large image contents for image retrieval.
2. It should allow the users to develop their own symbolic descriptions of images in the terms which are appropriate to the specific diagnostic task.
3. Users should have a possibility for updating or adding features according to new images or a diagnostic problem.
4. It should support the user in the analysis and interpretation of images; for example, in the evaluation of new imaging devices and radiographic materials.
5. It should assist the user in learning a proper prediction model based on different methods that are applicable for different data-type characteristics.

6. It should support the user in finding groups of features, objects or relations by proper clustering methods.

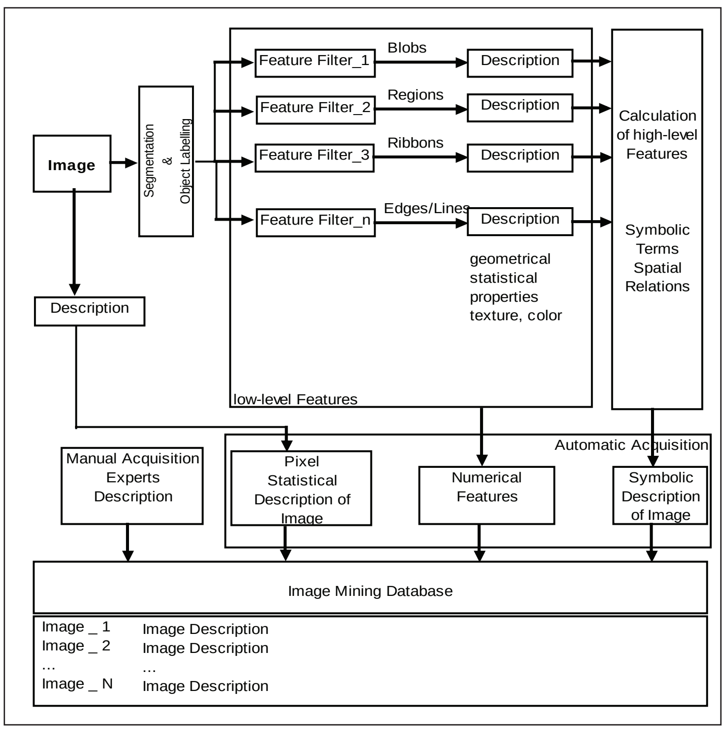
Taking into account these criteria and the recent state-of-the-art in image analysis, we provided an opportunity for semiautomatic image processing and analysis to enhance imaging of diagnostically important details in the image and measure some image features directly in the image and by this way to support the user by the analysis of images. The user has to have the possibility to interact with the system in order to adapt the results of image processing.

This image-processing unit should provide extraction of such low-level features as blobs, regions, ribbons, lines and edges. On the basis of these low-level features we are able to calculate then some high-level features to describe the image. Besides that, the image-processing unit should allow evaluation of some statistical image properties, which might give valuable information for the image description.

However, some diagnostically important features, such as "irregular structure inside the nodule," "tumor" are not so-called low-level features. They present some gestalts of expert domain knowledge. Development of an algorithm for extraction of such image features can be a complex or even unsolvable problem. So we identify different ways of representing the contents of an image that belongs to different abstraction levels (see Figure 1). We can describe an image:

- By statistical properties (i.e., that is, the lowest abstraction level)
- By low-level features and their statistical properties such as regions, blobs, ribbons, edges and lines (this is the next higher abstraction level)
- By high-level or symbolic features that can be obtained from the low-level features

Figure 1. Overview on image descriptions based on different abstraction levels



- By an expert's symbolic description, which is the highest abstraction level

The image-processing unit combined with the data-evaluation unit should allow a user to learn the relevant diagnostic features and effective models for the image interpretation. Therefore, the system as a whole should meet the following criteria:

1. Support the medical person as much as possible by the extraction of the necessary image details (region of interest).
2. Fulfill measurement of the feature values directly in the image when possible.
3. Display the interesting image details to the expert.
4. Store in a database the measured feature values as well as the subjective description of images by the expert.
5. Import these data from the database into the data-mining unit.

SYSTEM DESCRIPTION

Overall System Architecture

Figure 2 shows a scheme of the tool *ImageMiner Version 2.1*. There are two main parts in the tool:

- The online part that is comprised of the image analysis (Figure 3) and the image interpretation part
- The off-line part that is comprised of the database and the data-mining and knowledge discovery part (Figure 4)

The tool is written in C++ and runs under Windows NT. These two units communicate over a database of image descriptions which is created in the frame of the image-processing unit. This database is the basis for the image-mining unit.

The online part can automatically detect the objects, extract image features from the objects and classify the recognized objects into the respective classes based on the previously stored decision rules. The interface between the offline and the online part is the database where images and calculated image features are stored. The off-line part can mine the images for a prediction

Figure 2. Architecture of an image mining tool

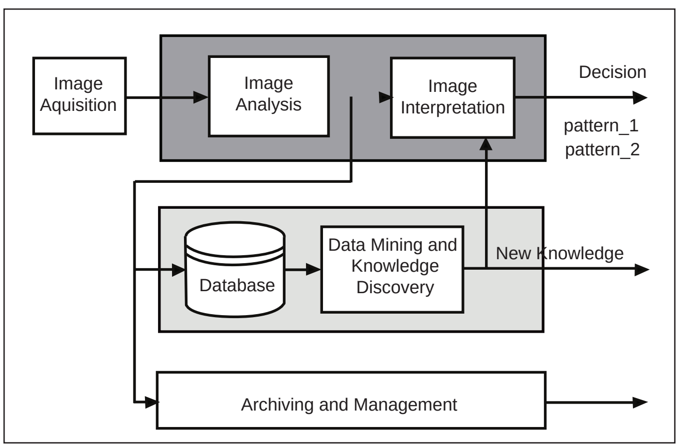


Figure 3. (a)-(f) Image analysis unit for the extraction of the image descriptions

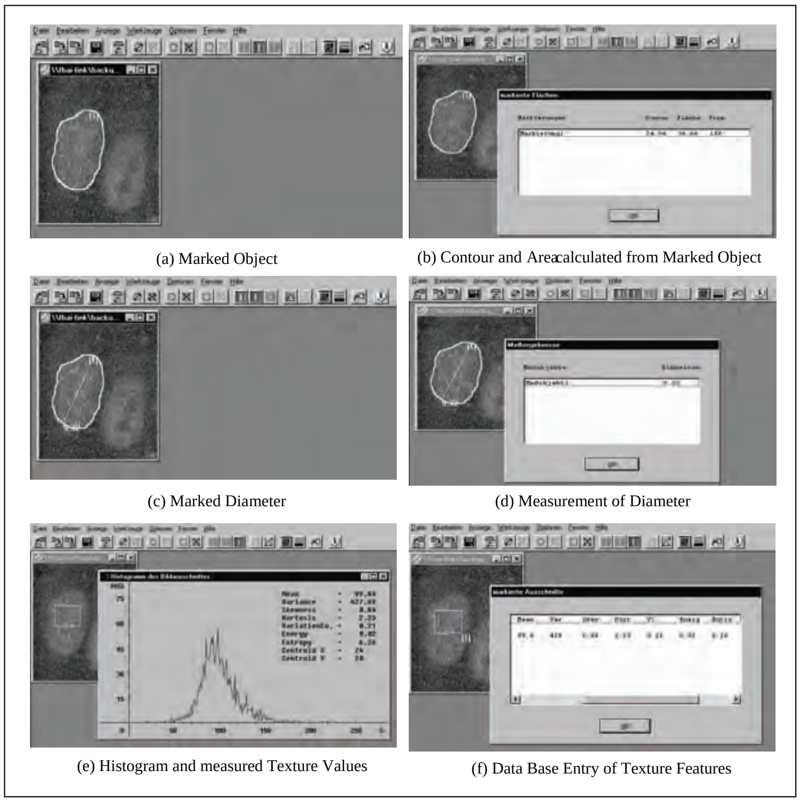
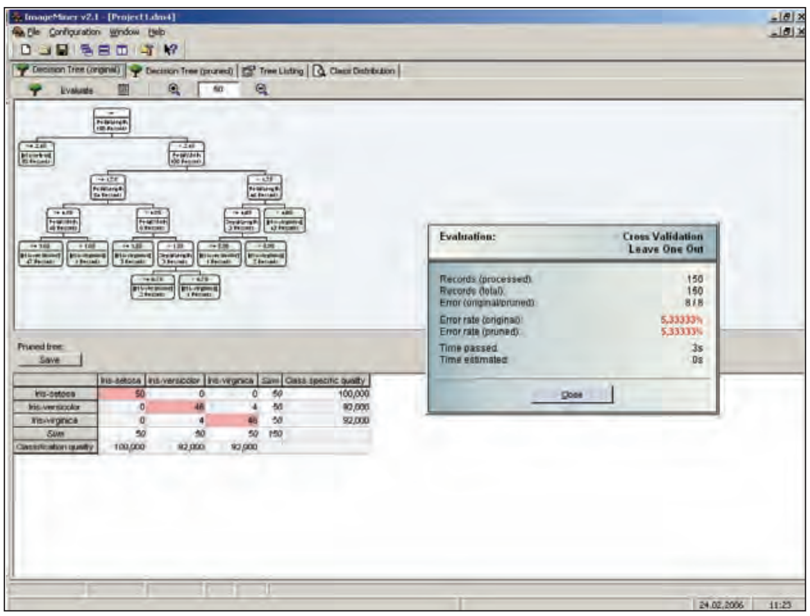


Figure 4. Tool ImageMiner Version 2.1



model or discover new groups of objects, features or relations. These similar groups can be used for learning the classification model or just for understanding the domain. In the latter case the discovered information is displayed to the user on the terminal of the system. Once a new prediction model has been learnt, the rules are inserted into the image interpretation part for further automatic interpretation after approval of the user.

Besides that there is an archiving and management part that controls the whole system and stores information for long-term archiving.

Images can be processed automatically or semi-automatically. In the first case a set of images specified by the expert is automatically segmented into the background and objects of interest and for all these objects features based on the feature-extracting procedures installed in the system are automatically calculated. In this way as many features as possible are calculated, regardless of whether they make sense for a specific application or not. This requires feature-subset selection methods later on. In the second case, an image from the image archive is selected by the expert and then displayed on the monitor (Figure 3). In order to perform image processing an expert communicates with a computer. In this mode he has the option to calculate features based on the feature-extracting procedures and/or insert symbolic features based on his expert knowledge. In both modes the features are stored into the database.

CBR Image Segmentation

To ensure that the system is as flexible as possible we have installed a case-based image-segmentation procedure (Perner, 1999) into our image-analysis part.

The CBR-based image segmentation consists of a case base in which formerly processed cases are stored. A case comprises image information, non-image information (e.g., image acquisition parameters, object characteristics, and so on), and image-segmentation parameters. The task is now to find the best segmentation for the current image by looking in the case base for similar cases. Similarity determination is based on both non-image information and image information. The evaluation unit will take the case with the highest similarity score for further processing. If there are two or more cases with the same similarity score, the case that appear first will be taken. After the closest case has been chosen, the image-segmentation parameters associated with the selected case will be given to the image-segmentation unit, and the current image will be segmented. Images having similar image characteristics will show similar good segmentation results when the same segmentation parameters are applied to these images.

This procedure finds the objects-of-interest in the image and labels them. In the automatic mode the labeled objects are automatically calculated and given to the image-interpretation part, whilst in the off-line mode the found object is displayed on a monitor to the user.

Feature Extraction

The expert can calculate image features for the labeled objects. These features are composed of statistical gray level features, the object contour, square, diameter, shape (Zamperoni, 1996) and a novel texture feature that is flexible enough to describe different textures on objects (Perner, Perner, & Müller, 2002). The expert evaluates or calculates image features and stores their values in a database of image features. Each entry in the database presents features of the object of interest. These features can be numerical (calculated on the image) and symbolical (determined by the expert as a result of image reading by the expert). In the latter case the expert evaluates object features according to the attribute list, which has to be specified in advance for object description, or is based on a visual ontology available for visual content description. Then the expert feeds these values into the database.

When the expert has evaluated a sufficient number of images, the resulting database can be used for the mining process.

Decision Tree Induction Unit

The stored database can easily be loaded into the image mining tool (Figure 4).

The tool fulfills a decision-tree induction as well as case-based reasoning and clustering. Decision-tree induction allows one to learn a set of rules and basic features necessary for decision-making in a specified diagnostic task. The induction process does not only act as a knowledge-discovery process, it also works as a feature selector, discovering a subset of features that is the most relevant to the problem solution.

Decision trees partition decision space recursively into sub-regions based on the sample set. By this way the decision trees recursively break down the complexity of the decision space. The outcome has a format which naturally presents the cognitive strategy that can be used for the human decision-making process.

For any tree all paths lead to a terminal node, corresponding to a decision rule that is a conjunction (AND) of various tests. If there are multiple paths for a given class, then the paths represent disjunctions (ORs).

The developed tool allows choosing different kinds of methods for feature selection, feature discretization, pruning of the decision tree and evaluation of the error rate. It provides an entropy-based measure, a gini-index, gain-ratio and chi square method for feature selection (Perner, 2002).

The following methods for feature discretization are provided: cut-point strategy, chi-merge discretization, minimum description-length-principal based discretization method and lvq-based method (Perner, 2002). These methods allow one to make discretization of the feature values into two and more intervals during the process of decision-tree building. Depending on the chosen method for attribute discretization, the result will be a binary or n-ary tree, which will lead to more accurate and compact trees.

The *ImageMiner* allows one to chose between cost-complexity pruning, error-reduction-based methods and pruning by confidence-interval prediction. The tool also provides functions for outlier detections.

To evaluate the obtained error rate one can choose test-and-train and n-fold cross validation. Missed values can be handled by different strategies (Perner, 2002).

The user selects the preferred method for each step of the decision tree induction process. After that the induction experiment can start on the acquired database. A resulting decision tree will be displayed to the user. The user can evaluate the tree by checking the features used in each node of the tree and comparing them with his/her domain knowledge.

Once the diagnosis knowledge has been learnt, the rules are provided either in txt-format or XML format for further use in the image-interpretation system or the expert can use the diagnosis component of the *ImageMiner tool* for interactive work. It has a user-friendly interface and is set up in such a way that non-computer specialists can handle it very easily.

Case-Based Reasoning Unit

Decision trees are difficult to utilize in domains where generalized knowledge is lacking. But often there is a need for a prediction system even though there is not enough generalized knowledge. Such a system should (a) solve problems using the already stored knowledge and (b) capture new knowledge making it immediately available to solve the next problem. To accomplish these tasks case-based reasoning is useful. Case-based reasoning explicitly uses past cases from the domain expert's successful or failing experiences.

Therefore, case-based reasoning can be seen as a method for problem solving as well as a method to capture new experience in incremental fashion and make it immediately available for problem solving. It can be seen as a learning and knowledge-discovery approach, since it can capture from new experience some general knowledge such as case classes, prototypes and some higher level concept. We find these methods especially applicable for inspection and diagnosis tasks (Perner, 2006). In the case of these applications people rather store prototypical images into a digital image catalogues than a large set of different images.

We have developed a unit for our *ImageMiner* that can perform similarity determination between cases, as well as prototype selection and feature weighting [Cheng, 1974]. We call $x_n^{`} \in \{x_1, x_2, ..., x_n\}$ a nearest-neighbor to x if $\min d(x_i, x) = d(x_n', x)$, where i=1,2,...,n. The instance x is classified into category C_n, if x_n is the nearest neighbor to x and x_n belongs to class C_n.

In the case of the k-nearest neighbor we require k-samples of the same class to fulfil the decision rule. As a distance measure we use the Euclidean distance. Prototype Selection from a set of samples is done by Cheng's algorithm (Cheng, 1974). Suppose a training set T is given as $T = \{t^1, ..., t^m\}$. The idea of the algorithm is as follows: We start with every point in T as a prototype. We then successively merge any two closest prototypes p^1 andm p^2 of the same class by a new prototype p if the merging will not downgrade the classification of is patterns in T. The new prototype p may simply be the average vector of p^1 and p^2. We continue the merging process until the number of incorrect classifications of patterns in T starts to increase.

The wrapper approach (Wetterschererk & Aha, 1995) is used for selecting a feature subset from the whole set of features. This approach conducts a search for a good feature subset by using the k-NN classifier itself as an evaluation function. The 1-fold cross validation method is used for estimating the classification accuracy and the best-first search strategy is used for the search over the state space of possible feature combination. The algorithm terminates if we have not found an improved accuracy over the last k search states. The feature combination that gave the best classification accuracy is the remaining feature subset. After we have found the best feature subset for our problem, we try to further improve our classifier by applying a feature weighting technique.

The weights of each feature w_i are changed by a constant value δ:w_i:=w_i±δ. If the new weight causes an improvement of the classification accuracy, then the weight will be updated accordingly; if not, the weight will remain as it is. After the last weight has been tested the constant δ will be divided into half and the procedure repeats. The procedure terminates if the difference between the classification accuracy of two iterations is less than a predefined threshold.

Conceptual Clustering

The intention of clustering is to find groups of similar cases among the data according to the observation. This can be done based on one feature or a feature combination. The resulting groups give an idea how data fit together and how they can be classified into interesting categories.

Classical clustering methods only create clusters but do not explain why a cluster has been established. Conceptual clustering methods build clusters and explain why a set of objects forms a cluster. Thus, conceptual clustering is a type of learning by observations and it is a way of summarizing data in an understandable manner (Fisher, 1987; Gennari, Langley, & Fisher, 1989). In contrast to hierarchical clustering methods, conceptual clustering methods build the classification hierarchy not only based on merging two groups. The algorithmic properties are flexible enough to dynamically fit the hierarchy to the data. This allows incremental incorporation of new instances into the existing hierarchy and updating this hierarchy according to the new instance.

A concept hierarchy is a directed graph in which the root node represents the set of all input instances and the terminal nodes represent individual instances. Internal nodes stand for sets of instances attached to the nodes and represent a super-concept. The super-concept can be represented by a generalized representation of this set of instances such as the prototype, the mediod or a user-selected instance. Therefore a concept C, called a class, in the concept hierarchy is represented by an abstract concept description and a list of pointers to each child concept $M(C)=\{C_1, C_2, ..., C_i, ..., C_n\}$, where C_i is the child concept, called subclass of concept C.

Our conceptual clustering algorithm presented here is based on similarities (Jänichen & Perner, 2006; Perner, 2003). Due to its concept description, it explicitly supplies for each cluster a generalized shape case which represents this group and a measure for the degree of its generalization. The result will be a sequence of partitions (concept hierarchy), where the root node contains the complete set of input cases and hence it follows that this node is represented by the most generalized case. The nodes in lower hierarchy levels are comprised of fewer cases (at least one) and are more specific.

In addition to *create* and *add*, we also introduced the operators *split* and *merge* into the algorithm. We prefer to apply these local operators because they preserve the incremental fashion of the algorithm. Order dependency also becomes decreased sufficiently, even it is not guaranteed that the local changes have a sufficiently strong effect on the global data.

The algorithm implements a top-down method. Initially the concept hierarchy only consists of an empty root node. A new case is placed into the actual concept hierarchy level by level, beginning with the root node until a terminal node is reached. In each hierarchy level one of these four different kinds of operations is performed:

- The new case is incorporated into an existing child node.
- A new empty child node is created where the new case is incorporated.
- Two existing nodes are merged to form a single node where the new case is incorporated.
- An existing node is split into its child nodes.

The new case is tentatively placed into the next hierarchy level by applying all of these operations. Finally that operation is performed which gives the best score to the partition according to the evaluation criteria. A proper utility function prefers compact and well- separated clusters. These are clusters with small inner-cluster variances and high inter-class variances. Thus we calculate the score of a partition comprised of the clusters $\{X_1, X_2, \ldots X_m\}$ by: $SCORE = \frac{1}{m}\sum_{i=1}^{m} p_i\ (SB_i - SW_i)$, where m is the number of clusters in this partition, p_i is the relative frequency of the i-th cluster, SB_i is the inter-cluster variance and SW_i is the inner-cluster variance of the i-th cluster. The normalization according to m is necessary to compare partitions of different size. The relative frequency p_i of the i-th cluster is: $p_i = \frac{n_i}{n}$, where n_i is the number of cases in the i-th cluster and n is the number of cases in the parent node. The inter-cluster variance of a cluster X is calculated by: $SB_X = \frac{1}{n_x}\sum_{i=1}^{n_x} d(x_i, \bar{\mu}_P)^2$, where $\bar{\mu}_P$ is the cluster centre of the parent node and x_i are the instances in all child nodes.

The output of our algorithm for applying eight exemplary shape cases of fungal strain *Ulocladium Botrytis* is shown in Figure 5. On the top level the root node is shown which comprises of all input cases. Successively the tree is partitioned into nodes until each input case forms its own cluster.

The main advantage of our conceptual clustering algorithm is that it brings along a concept description. Thus, in comparison to agglomerative clustering methods, it is easy to understand why a set of cases forms a cluster. During the clustering process the representative case, and also the variances and maximum distances in relation to this representative case are calculated, since they are part of the concept description. The algorithm is of incremental fashion, because it is possible to incorporate new cases into the existing learnt hierarchy.

THE OVERALL IMAGE MINING PROCEDURE

The whole procedure for image mining is summarized in Figure 6. It is partially based on our developed methodology for image-knowledge engineering (Perner, 1994). The process can be divided into five major steps: (1) brain storming, (2) interviewing process, (3) collection of image descriptions into the data base, (4) mining experiment and (5) review.

Brain storming is the process of understanding the problem domain and identifying the important knowledge pieces on which the knowledge-engineering process will focus.

For the interviewing process we used our developed methodology for image-knowledge engineering described in Perner (1994) in order to elicit the basic attributes as well as their attribute values. Then the proper image processing and feature-extraction algorithms are identified for the automatic extraction of the features and their values.

Based on these results we then collected into the data base image readings done by the expert and done by the automatic image-analysis and feature-extraction tool. The resulting data base is the basis for our mining experiment. The error rate of the mining result was then determined based on sound statistical methods such as cross validation. The error rate as well as the rules were then reviewed together with the expert and depending on the quality of the results the mining process stops or goes into a second trail, starting either at the top with eliciting new attributes or at a deeper level, for example with reading new images or

Figure 5. Concept hierachy for 2-D forms of biological objects

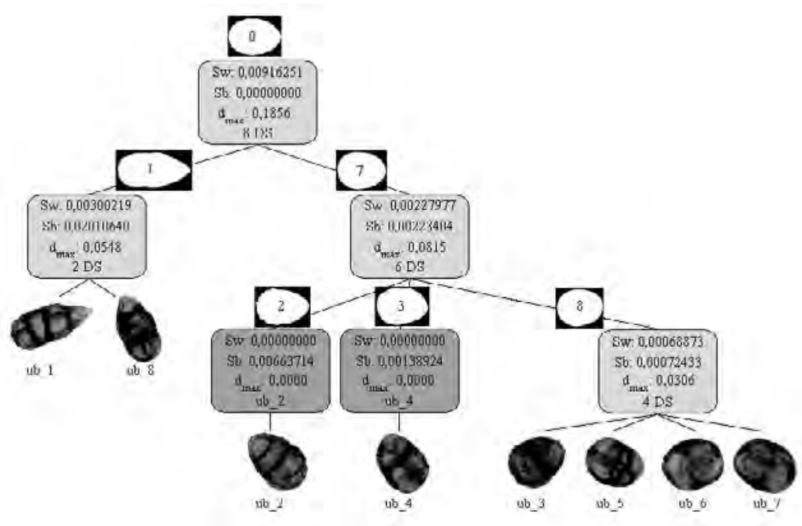
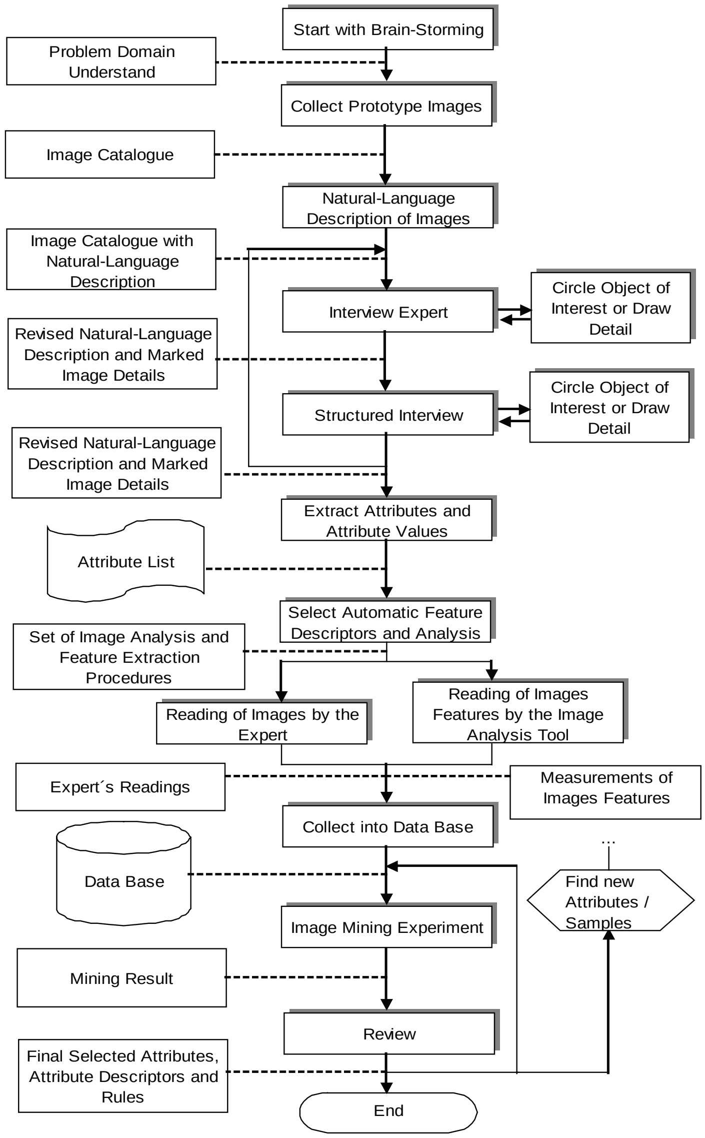


Figure 6. Procedure of the image mining process



incorporating new image-analysis and feature-extraction procedures. The incorporation of new image-analysis and feature-extraction procedures seems to be an interactive and iterative process at the moment, since it is not possible to provide ad-hoc sufficient image-analysis procedures for all image features and details appearing in the real world. The mining procedure stops as soon as the expert is satisfied with the results.

A CASE STUDY

The Application

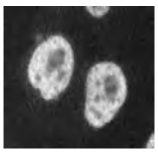
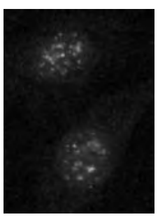
We will describe the usage of the image-mining tool based on the task of HEp-2 cell classification. HEp-2 cells are used for the identification of antinuclear autoantibodies (ANA). They allow the recognition of over 30 different nuclear and cytoplasmic patterns which are given by upwards of 100 different autoantibodies. The identification of these patterns has up to now been done manually by a human inspecting the slides with the help of a microscope. The lacking automation of this technique has resulted in the development of alternative techniques based on chemical reactions, which do not have the discrimination power of the ANA testing. An automatic system would pave the way for a wider use of ANA testing.

Recently, the various HEp-2 cell images occurring in medical practice are being collected into a data base at the university hospital of Leipzig. The images were taken by a digital image-acquisition unit consisting of a microscope AXIOSKOP 2 from Carl Zeiss Jena, coupled with a color CCD camera Polariod DPC. The digitized images were of 8-bit photometric resolution for each color channel with a per pixel spatial resolution of 0.25 µm. Each image was stored as a color image on the hard disk of the PC but is transformed into a gray-level image before being used for automatic image analysis.

The scope of our work was to mine these images for proper classification knowledge, so that it can be used in medical practice for diagnosis or for teaching novices. Besides that it should give us the basis for the development of an automatic image-diagnosis system.

Our experiment was supported by an immunologist who is an expert in the field and acts as a specialist to other laboratories in case of diagnostically complex cases.

Figure 7. Image catalogue and expert's description

Class	Image	Description
Fine Speckled 200 000		Smooth and uniform fluorescence of the nuclei Nuclei sometimes dark Chromosomes fluoresced weak up to extreme intensive
Fine dotted (speckled) nuclei fluorescence 320 200		Dense fine speckled fluorescence Background diffuse fluorescent
Homogeneous Nuclear 100 000		A uniform diffuse fluorescence of the entire nucleus of interphase cells. The surrounding cytoplasm is negative.
...	...	...
Centromere 500 000		Nuclei weak uniform or fine granular, poor distinction from background

Brainstorming and Image Catalogue

First, we started with a brain storming process that helped us to understand the expert's domain and to identify the basic pieces of knowledge. We could identify mainly four pieces of knowledge: Hep-2 cell atlas, the expert, slide preparation and a book describing the basic parts of a cell and their appearance.

Table 1. Attribute list

Attribute	Code	Attribute Values
Interphase Cells	0	Undefined
	1	Fine speckled
	2	homogeneous
	3	Coarse Speckled
	4	Dense fine speckled Fluorescence
Nucleoli	0	Undefined
	1	Dark area
	2	fluoresce
Background	0	Undefined
	1	Dark
	2	Fluorescence
Chromosomes	0	Undefined
	1	Fluorescence
	2	Dark
Cytoplasm	0	Undefined
	1	Speckled Fluorescence
Classes	100 000	Homogeneous
	100 320	Homogeneous fine speckled
	200 000	Nuclear
	320 000 320 200	Fine speckled Fine speckled nuclear

Then the expert collected prototype images for each of the six classes appearing most frequently in his daily practice. The expert wrote down a natural-language description for each of these images. As a result we obtained an image catalogue having a prototype image for each class and associated to each image is a natural-language description of the expert (see Figure 7).

Interviewing Process

Based on these image descriptions we started our interviewing process. First, we only tried to understand the meaning of the expert description in terms of image features. We let the individual circle the interesting object in the image to understand the meaning of the description. After having done this, we went into a structured interviewing process asking for specific details such as: "Why do you think this object is *fine-speckled* and the other one is not. Please describe the difference between these two." It helped us to verify the expert description and to make the object features more distinct.

Finally, we could extract from the natural-language description the basic vocabulary (attributes and attribute values, see Table 1) and associate the meaning to each attribute.

In a last step we reviewed the chosen attributes and the attribute values with the expert and found a common agreement on the chosen terms. The result was an attribute list which is the basis for the description of object details in the images. Furthermore, we identified from the whole set of feature descriptors the set of a feature descriptors which might be useful for the objective measurement of image features. In our case we found that describing the cells by their boundary and calculating the size and the contour of the cell might be appropriate. The different descriptors of the nuclei of the cells might be sufficiently described by the texture descriptor of our image-analysis tool.

Collection of Image Descriptions into the Data Base

Now we could start to collect a data base of image descriptions based on these attributes and attribute values as well as on feature measurements calculated with the help of the image-analysis tool. For our experiment we used a dataset of 110 images. The dataset contained 6 classes, each equally distributed. For each class we had 20 images. The expert used the image-analysis tool shown in Figure 3 and displayed one after another each image from our data base. He watched the images on display and described the image content on the basis of our attribute list and fed the attribute values into the data base. Besides that he marked the objects of interest in the image on display and used the necessary feature descriptors selected during the interviewing process and provided by the image-analysis unit to measure the image features such as size, contour, and texture. The resulting values for these features are automatically fed into the data base and stored together with the expert's image description into the data base (see Figure 8).

The Image Mining Experiment

The collected dataset was then given to the tool *ImageMiner*. The decision-tree induction algorithm that showed the best results on this dataset is based on the entropy-criterion for the attribute selection, cut-point strategy for the attribute discretization and minimal error-reduction pruning. We carried out three experiments. First, we learnt a decision tree only based on the image reading by the expert, then learnt a decision tree only based on the automatic calculated images features, and finally, we learnt a decision tree based on a data base containing both feature descriptions. The resulting decision tree for the expert's reading is shown in Figure 9 and the resulting decision tree for the expert's reading together with the measured image features is shown in Figure 10. We do not show the tree for the measured image features, since the tree is too complex. The error rate was evaluated by leave-one-out cross-validation.

The error rate of the decision trees from the first two experiments is higher than the error rate made by the expert (see Table 2). None of the trees, whether based on the expert's reading or based on the measured image features, give a sufficiently low error rate. Only the combined data base from the expert's reading and measured image features gives us an error rate that comes close to an expert's error rate.

Review

The tree created based on the image readings from the expert has an error rate of 27.9% (see Table 2). Under the assumption that the class labels represent the true class (gold standard), we can only conclude that there is a knowledge gap.

Figure 8. Excerpt from database

Class	Contour (Kontur)	Area	Shape Factor (Form)	MEAN	VAR	SKEW	CURT	VC	ENERGY	...	NUCLEOLI	CHROMO	CYTOPLA (Zytopla)	Background (Hintergrund)	
100000	14,3734	14,3189	144,2812	87,1507	244,3043	1,1233	7,5139	0,1793	0,0209	...	1	1	0	1	
100320	10,3675	7,2986	147,2687	144,6974	282,0444	-0,6999	2,4243	0,1161	0,0238	...	1	0	0	0	
320200	11,9142	9,4348	150,4512	132,5286	675,6562	0,1685	-0,5039	0,1961	0,0119	...	2	0	0	1	
200000	9,0332	5,2114	156,5795	94,5199	1400,9983	0,6564	-0,3728	0,3960	0,0100	...	2	0	0	1	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	

Figure 9. Decision tree obtained from expert's readings

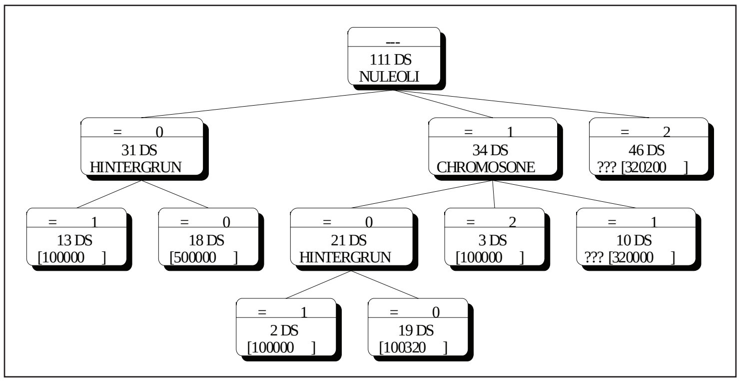


Figure 10. Decision tree obtained from expert's readings and image readings

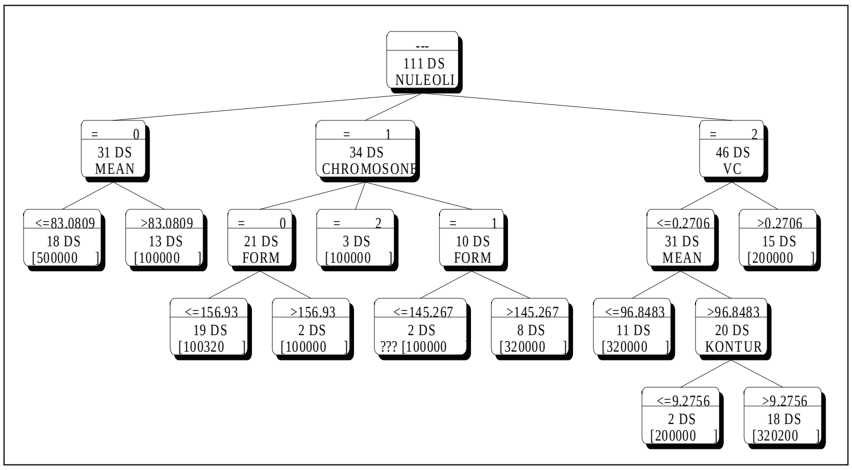


Table 2. Error rate for decision trees obtained from the different databases

	Error Rate		
Data Set	**Expert**	**Unpruned Tree**	**Pruned Tree**
Original Data Set	25 %		
Expert Readings		27 %	27 %
Automatic Feature Analysis		27 %	27 %
Combined Data Set		6,9 %	6,9 %

There must be some hidden knowledge which the expert is using during decision making, but the expert could not make this knowledge explicit during the interviewing process. Here we have an example for the problem "difference between showing and naming" mentioned in the second section. However, the expert's error rate is also high.

Our first objection was: Is the assumption that the class label is the true class label true or not? As far as we know the chemical investigation of the serum which was used to determine the gold standard does not so accurately discriminate between the different classes.

The experiment based on the features automatically measured in the images gives us no better results. The resulting tree is very bushy and deep and uses almost all attributes.

Only the combination between the expert's readings and the readings by the image-analysis unit shows us reasonable results. The feature nucleoli is the most important feature and the correct description of the nucleoli will improve the results dramatically. During the image-analysis phase we did not describe this object separately. The hope was that the texture descriptor for the whole cell is sensitive enough to model the different visual appearances of the different cells. The experiment shows that only the combination of basic image descriptors from the image analysis with expert reading gave sufficient results. Therefore, we believe that our first objection concerning the true class label does not hold any more. We rather think that in order to improve the accuracy of the classifier we must find a good feature descriptor for the different appearances of the object nucleoli.

LESSONS LEARNED

We have found out that our methodology of data mining allows a user to learn the decision model and the relevant diagnostic features. A user can independently use such a methodology of data mining in practice. The user can easily perform different experiments until the user is satisfied with the result. By practicing this, the user can explore the application and find out the connection between different knowledge pieces.

However, some problems should be taken into account for the future system design.

As we have already pointed out in a previous experiment (Perner, 2000), an expert tends to specify symbolic attributes by means of a large number of attribute values. For example in this experiment the expert specified for the attribute "margin" 15 attribute values such as "non-sharp," "sharp," "non-smooth," smooth," and so on. A large number of attribute values will result in small sub-sample sets soon after the tree-building process started. It will result in a fast termination of the tree-building process. This is also true for small sample sets that are usual in medicine. Therefore, a careful analysis of the attribute list should be done after the physician has specified it.

During the process of building the tree, the algorithm picks the attribute with the best attribute-selection criteria. If two attributes have both the same value, the one that appears first in the attribute list will be chosen. That might not always be the attribute the expert himself would choose. In order to avoid this problem, we think that in this case we should allow the expert to choose manually the attribute that the expert prefers. We expect that this procedure will bring the resulting decision model closer to the expert's ones.

The developed image-analysis tool allows extracting image features (see the fourth section). It supported the analysis and exploration of other image-diagnosis tasks, such as the analysis of sheep follicle and lymph nodule analysis. It proved to be very useful for the analysis of microscopic images of different cell-based assays. New applications might require further feature descriptors. Therefore the image-analysis tool must have an open architecture that allows incorporating new feature descriptors into the tool.

The described method of image mining had been applied to a wide range of applications. This includes the analysis of sheep follicle based on a texture descriptor, evaluation of imaging effects of radiopaque material for lymph-nodule analysis, mining knowledge for IVF therapy, transplantation medicine, and inspection of microscopic images of cell-based assays and for the diagnosis of breast carcinoma in MR images. In all these tasks we did not have a well-trained expert. These were new tasks and reliable decision knowledge had not been built up in practice yet. The physicians did the experiments by themselves. They were very happy with the obtained results, since the learnt rules gave them deeper understanding of their problems and helped to predict new cases. It helped the physicians to explore their data and inspired them to think about new improved ways of diagnosis. In some of the tasks we were able to build an automatic image-interpretation system based on the discovered visual features and knowledge.

Compared to our first version of the image-mining tool the new features of case-based image segmentation, case-based reasoning and clustering gave a new flexibility to the image mining process depending on the characteristics of the data of the particular application. The clustering method allowed discovering new groups according to the considered observation(s) that could be used further for the construction of the prediction model. Case-based image segmentation gave the flexibility needed to discover objects of interest in different image modalities and qualities. In the case of rare data or image catalogues case-based reasoning was the right method to construct a decision model and acquire new images in incremental fashion.

CONCLUSION AND FURTHER WORK

In this chapter we have presented our methods and the methodology for image mining. Based on that, we built our system *ImageMiner Version 2.1*. This tool has shown excellent performance for a wide range of image-mining tasks. The basis for the image mining experiment is a sufficiently large database with images, feature description and/or expert descriptions. Such databases result for example from the broad use of image databases in many fields.

We were able to learn the important attributes needed for image interpretation and to understand the way in which these attributes were used for decision-making by applying data- mining methods to the database of image descriptions. We showed how the domain vocabulary should be set up to get good results and which techniques should be used in order to check the reliability of the chosen features.

The explanation capability of the induced tree was reasonable. The attributes included into the tree represented the expert knowledge. Finally, we can say that image-archiving systems in a combination with image-mining methods open a possibility for advanced computer-assisted diagnosis-system development.

We have recently been going on to apply the system to video microscopy and develop more feature extractions and image-mining procedures that can further support the image-mining process.

REFERENCES

Bhanu, B., & Dong, A. (2001). Concepts learning with fuzzy clustering and relevance feedback. In P. Perner (Ed.), *Machine learning and data mining in pattern recognition* (pp. 102-116). Berlin; Heidelberg: Springer-Verlag.

Bouguila, N., & Ziou, D. (2005). MML-based approach for finite dirichlet mixture estimation and selection. In P. Perner (Ed.), *Machine learning and data mining in pattern recognition* (pp. 42-52). Berlin; Heidelberg: Springer-Verlag.

Burl, M. C., & Lucchetti, D. (2000) Autonomous visual discovery. In B. V. Dasarathy (Ed.), *Data mining and knowledge discovery: Theory, tools, and technology, Proceedings of SPIE* (Vol. 4057, pp. 240-250).

Chang, C. L. (1974). Finding prototypes for nearest neighbor classifiers. *IEEE Transactions on Computers, C-23*(11), 1179-1184.

Eklund, P. W., You, J., & Deer, P. (2000). *Mining remote sensing image data: An integration of fuzzy set theory and image understanding techniques for environmental change detection*. In B. V. Dasarathy (Ed.), Data mining and knowledge discovery: Theory, tools, and technology, Proceedings of SPIE (Vol. 4057, pp. 265-273).

Fisher, D. H. (1987). Knowledge acquisition via incremental conceptual clustering. *Machine Learning, 2*(2), 139-172.

Fischer, S., & Bunke, H. (2001). Automatic identification of diatoms using decision forests. In P. Perner (Ed.), *Machine learning and data mining in pattern recognition* (pp. 173-183). Berlin; Heidelberg: Springer-Verlag.

Gennari, J. H., Langley, P., & Fisher, D. H. (1989). Models of incremental concept formation. *Artificial Intelligence, 40*(1-3), 11-61.

Giacinto, G., & Roli, F. (1999). Automatic design of multiple classifier systems by unsupervised learning. In P. Perner & M. Petrou (Eds.), *Machine learning and data mining in pattern recognition* (pp. 131-143). Berlin; Heidelberg: Springer-Verlag.

Imiya, A., & Ootani, H. (2001). PCA-based model selection and fitting for linear manifolds. In P. Perner (Ed.), *Machine learning and data mining in pattern recognition* (pp. 278-292). Berlin; Heidelberg: Springer-Verlag.

Indurkhya, N., & Weiss, S. M. (2001). Rule-based ensemble solutions for regression. In P. Perner (Ed.), *Machine learning and data mining in pattern recognition* (pp. 62-72). Berlin; Heidelberg: Springer-Verlag.

Jahn, H. (1999). Unsupervised learning of local mean gray values for image processing. In P. Perner & M. Petrou (Eds.), *Machine learning and data mining in pattern recognition* (pp. 64-74). Berlin; Heidelberg: Springer-Verlag.

Jänichen, S., & Perner, P. (2006). Conceptual clustering and case generalization of 2-dimensional forms. *International Journal of Computational Intelligence, 22*(3-4), 177-193.

Kehoe, A., & Parker, G. A. (1991). An IKB defect classification system for automated industrial radiographic inspection. *IEEE Expert Systems, 8*, 149-157.

Malerba, D., Esposito, F., Lanza, A., & Lisi, F. A. (2001). First-order rule induction for the recognition of morphological patterns in topographic maps. In P. Perner & M. Petrou, (Eds.), *Machine learning and data mining in pattern recognition* (pp. 88-101). Berlin; Heidelberg: Springer Verlag.

Matsuo, Y., Amano, M., & Uehara, K. (2002). Mining video editing rules in video streams. *ACM Multimedia*, 255-258.

Martinez, J. (2001). *Overview of the MPEG-7 Standard* (version 5.0), ISO/IEC JTC1/SC29/WG11 N4031, Singapore. Retrieved October 19, 2006, from http://www.cselt.it/mpeg/standards/mpeg-7/mpeg-7.htm

Megalooikonomou, K., Davatzikos, C., & Herskovits, E. (1999). Mining lesion-defect associations in a brain image database. In *Proceedings from the International Conference Knowledge Discovery and Data Mining (KDD99)*, San Diego, CA (pp. 347-351).

Palenichka, R. M., & Volgin, M. A. (1999). Extraction of local structural features in images by using a multi-scale relevance function. In P. Perner & M. Petrou (Eds.), *Machine learning and data mining in pattern recognition* (pp. 87-102). Berlin; Heidelberg: Springer-Verlag.

Perner, P. (1994). A knowledge-based image inspection system for automatic defect recognition, classification, and process diagnosis. *International Journal on Machine Vision and Applications, 7*, 135-147.

Perner, P. (1998). Case-based reasoning for the low-level and high-level unit of an image interpretation system. In S. Singh (Ed.), *Advances in pattern recognition* (pp. 45-54). Berlin; Heidelberg: Springer-Verlag.

Perner, P. (1999). An architeture for a CBR image segmentation system. *Engineering Applications of Artificial Intelligence, 12*(6), 749-759.

Perner, P. (2000). Mining knowledge in medical image databases. In B. V. Dasarathy (Ed.), *Data mining and knowledge discovery: Theory, tools, and technology, Proceedings of SPIE* (Vol. 4057, pp. 359-369).

Perner, P., Zscherpel, U., & Jacobsen, C. (2001). A comparison between neural networks and decision trees based on data from industrial radiographic testing. *Pattern Recognition Letters, 2*, 47-54.

Perner, P., Perner, H., & Müller, B. (2002). Mining knowledge for Hep-2 cell image classification. *Journal Artificial Intelligence in Medicine,* 26, 161-173.

Perner, P. (2002). *Data mininig on multimedia data*. Berlin; Heidelberg: Springer-Verlag.

Perner, P. (2003). Incremental learning of retrieval knowledge in a case-based reasoning system. In K. D. Ashley & D. G. Bridge (Eds.), *Case-based reasoning—Research and development* (pp. 422-436). London: Springer-Verlag.

Perner, P. (2005). Case-based reasoning for image analysis and interpretation. In C. H. Chen & P. S. P. Wang (Eds.), *Handbook of pattern recognition and computer vision* (3rd ed., pp. 95-114). World Scientific Publishing.

Perner, P. (in press). A comparative study of catalogue-based classification. In *Proceedings ECCR 2006.* Berlin; Heidelberg.

Petrushin, V. A. (2005). Mining rare and frequent events in multi-camera surveillance video using self-organizing maps. In *Proceedings of KDD2005* (pp. 794-800).

Satyanarayana, A., & Davidson, I. (2005). A dynamic adaptive sampling algorithm (DASA) for real world applications: Finger print recognition and face recognition. In M.-S. Hacid, N. V. Murray, Z. W. Ras, & S. Tsumoto (Eds.), *Foundations of intelligent systems* (pp. 631-640). New York: Springer-Verlag.

Schmidt, R., & Gierl, L. (2001). Temporal abstraction and case-based reasoning for medical course data: Two prognostic applications. In P. Perner (Ed.), *Machine learning and data mining in pattern recognition* (pp. 23-33). Berlin; Heidelberg: Springer Verlag.

Schröder, S., Niemann, H., & Sagerer, G. (1988). Knowledge acquisition for a knowledge-based image analysis system. In J. Bosse & B. Gaines (Eds.), *Proceedings of the European Knowledge-Acquisition Workshop EKAW 88*, Sankt Augustin.

Sivic, J., & Zisserman, A. (2004). Video data mining using configurations of viewpoint invariant regions. *Computer Vision and Pattern Recognition, 1*, 488-495.

Wettschereck, D., & Aha, D. W. (1995). Weighting features. In M. M. Veloso & A. Aamodt (Eds.), Case-based reasoning research and development (pp. 347-358). Berlin; Heidelberg: Springer-Verlag.

Zhu, X., Wu, X., Elmagarmid, A. K., Feng, Z., & Wu, L. (2005). Video data mining: Semantic indexing and event detection from the association perspective. *IEEE Transaction on Knowledge and Data Engineering, 17*(5), 665-677.

Zaiane, O. R., & Han, J. (2000) Discovery spatial associations in image. In B. V. Dasarathy (Ed.), *Data mining and knowledge discovery: Theory, tools, and technology* (pp. 138-148).

Zamperoni, P. (1996). Feature extraction. In H. Maitre & J. Zinn-Justin (Eds.), *Progress in picture processing* (pp. 121-184). Elsevier Science.

This work was previously published in Knowledge Discovery and Data Mining: Challenges and Realities, edited by X. Zhu and I. Davidson, pp. 75-97, copyright 2007 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 2.24
Efficient Method for Image Indexing in Medical Application

Richard Chbeir
University of Bourgogne, France

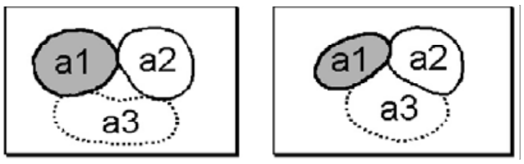
INTRODUCTION

In last two decades, image retrieval has seen a growth of interests in several domains. As a result, a lot of work has been done in order to integrate it in the standard data processing environments (Rui, Huang, & Chang, 1999; Smeulders, Gevers, & Kersten, 1998; Yoshitaka & Ichikawa, 1999). To retrieve images, different methods have been proposed in the literature (Chang & Jungert, 1997; Guttman, 1984; Lin, Jagadish, & Faloutsos, 1994). These methods can be grouped into two major approaches: metadata-based and content-based approaches. The metadata-based approach uses alphanumeric attributes and traditional techniques to describe the context and/or the content of the image such as title, author name, date, and so on. The content-based approach uses image processing algorithms to extract low-level features of images such as colors, textures, and shapes. Image retrieval using these features is done by methods of similarity and hence is a non-exact matching.

The requirement of each method depends on the application domain. In this paper, we address the domain of medicine where image retrieval in particular is very complex and should consider:

- Both content-based and metadata representations of images and salient objects. This guarantees a pertinent integration of all the aspects of image in order to capture pertinent information and to assure the relevance of all query types (Chbeir, Atnafu, & Brunie, 2002).
- High-precision description of images. For example, the spatial data in surgical or radiation therapy of brain tumors is decisive because the location of a tumor has profound implications on a therapeutic decision (Chbeir, Amghar, & Flory, 2001; Chbeir et al., 2002). Furthermore, it is crucial to distinguish between similar situations. Figure 1 shows two different images of three salient objects that are traditionally described by the same spatial relations in both cases: topological relations: a1 Touch a2, a1 Touch

Figure 1. Two different spatial situations

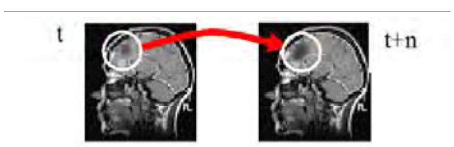


a3, a2 Touch a3; and directional relations: a1 Above a3, a2 Above a3, a1 Left a2.

- The evolutionary aspect of image content (Chbeir, Amghar, Flory, & Brunie, 2001) such as tumor development in brain (Figure 2), virus changes, and so on. The detection of the evolutionary aspects of objects (displacement, deformation, contraction, rotation, etc.) can significantly help physicians to establish an appropriate diagnosis or to make a therapeutic or surgical decision. An example for such a query is: "Find treatments of lesion detected inside brain images where a size increasing has been observed at every examination between time t and t+n".

In this article, we address the spatial and evolutionary issues of images. We propose a novel method that considers different types of relations. This method allows providing a highly expressive and powerful mechanism for indexing images.

Figure 2. Tumor growth inside the brain



The rest of this article is organized as follows: the next section is devoted to detail the related work. In the following section, we define our method of computing the different relations and we show how image indexing can be done. The subsequent section demonstrates how our method can adequately index medical images. Finally, we conclude and give future work orientations.

RELATED WORK

The problem of image retrieval is strongly related to image representation. Computing relations between either salient objects, shapes, points of interests, etc. have been widely used in image representation such as R-tree and its variants (Beckmann, 1990; Guttman, 1984), hB-tree (Lomet & Salzberg, 1990), ss-tree (White & Jain, 1996), TV-tree (Lin et al., 1994), 2D-String and its variants (Chang & Jungert, 1997; Chang & Jungert, 1991; Chang, Shi, & Yan, 1987), and so on. Spatial relations are mostly used for indexing and retrieval purposes for its automatic detection capability.

Three major types of spatial relations are generally proposed in image representation (Egenhofer, Frank, & Jackson, 1989):

- Metric relations measure the distance between salient objects (Peuquet, 1986). For instance, the metric relation "far" between two objects A and B indicates that each pair of points A_i and B_j has a distance grater than a certain value d.
- Directional relations describe the order between two salient objects according to a direction, or the localisation of salient object inside images (El-kwae & Kabuka, 1999). In the literature, fourteen directional relations are considered:
  - **Strict:** north, south, east, and west.
  - **Mixture:** north-east, north-west, south-east, and south-west.

- **Positional:** left, right, up, down, front and behind.

Directional relations are rotation variant and there is a need to have referential base. Furthermore, directional relations do not exist in certain configurations.

- Topological relations describe the intersection and the incidence between objects. Egenhofer and Herring (1991) have identified six basic relations: disjoint, meet, overlap, cover, contain, and equal. Topological relations present several characteristics that are exclusive to two objects (i.e., there is one and only one topological relation between two objects). Furthermore, topological relations have absolute value because of their constant existence between objects. Another interesting characteristic of topological relations is that they are transformation, translation, and scaling invariant.

In spite of all proposed work to represent complex visual situations, several shortcomings exist in the methods of spatial relation computations. Particularly, traditional methods do not have the required expressive power to distinguish between similar situations in critical domains such as in medicine.

On the other hand, the evolution of image content needs to be taken into consideration in several domains. Any evolution needs to consider time and thus temporal relations. For that reason, two paradigms are proposed in the literature in order to compute temporal relations:

- The first paradigm consists of representing the time as a set of instants: $t_1,..t_i,..t_n$. Traditionally, only three temporal relations are possible between two objects: before, its symmetric relation after, and equal.
- The second paradigm considers the time as a set of intervals $[t_i, t_j]$. Allen relations (Allen, 1983) are often used to represent temporal relations between intervals. Allen proposes 13 temporal relations (Figure 3), in which six are symmetrical.

For instance, in geographic applications, spatio-temporal queries (Bonhomme, Trepied, Aufaure, & Laurini, 1999) are used more and more to study the translation of a mobile object, the evolution of spatial objects, and so on. In the medical domain, images are evolutionary in nature. Their content evolution description can provide an important support for the treatment of diseases. To the best of our knowledge, the evolutionary content of medical images was only studied by Cárdenas, Ieong, Taira, Barker, and Breant, (1993); Chu, Hsu, Cárdenas et al. (1998). In Cardenas et al. (1993), the authors study the development of bones structures. They define a temporal evolutionary data model (TEDM). It extends traditional constructs. However, several evolutionary situations (such as deformation, expansion, contraction, etc.) are not considered.

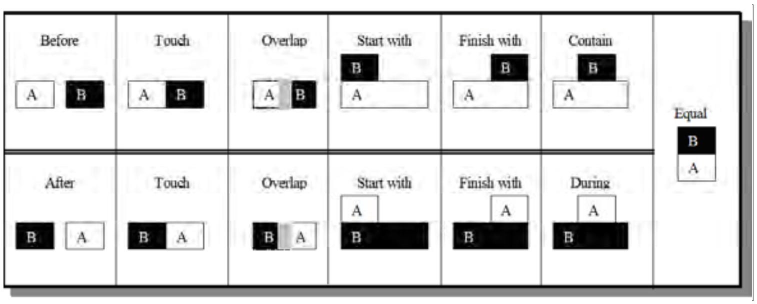
In this paper, we present our indexing method that is capable to represent spatial relations in highly expressive manner and to provide efficient technique to detect evolutionary content of images.

META-MODEL OF RELATIONS

Our proposal represents a generalized extension of the 9-Intersection model and its variants (Egenhofer & Franzosa, 1991). It provides a method for computing not only topological relations but also other types of relations with higher precision.

The idea is to identify relations of two values (or instances) on the basis of the same feature (such as shape, position, time, etc.). For example, the shape feature expresses spatial relations, the time feature provides temporal relations, and so on. To identify a relation between two feature values, we use an intersection matrix between several sets defined below.

Figure 3. Allen relations



Definition of Intersection Sets

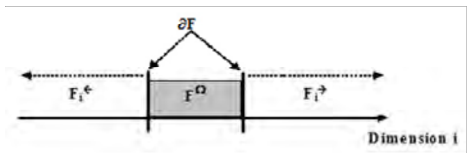
Let us first consider a feature F. We define its intersection sets as follows:

- The interior F^{Ω} contains all elements that cover the interior or the core of F. In particular, it contains the barycentre of F. The definition of this set has a great impact on the other sets. F^{Ω} may be empty ($\emptyset$).
- The boundary ∂F contains all elements that allow determining the frontier of F. ∂F is never empty ($\neg\emptyset$).
- The exterior F^{-} is the complement of $F^{\Omega}\cup\partial F$. It contains at least two elements: the minimum value and the maximum value. F^{-} can be divided into several disjoint exterior subsets. This decomposition depends on the number of the feature dimensions.

If the feature has only one dimension i (such as the acquisition time of a frame in a video), two exterior intersection subsets are defined (Figure 4):

- $F_i^{\leftarrow}$ (or inferior) contains elements of F^{-} that do not belong to any other intersection set and inferior to ∂F elements on the basis of i dimension.
- $F_i^{\rightarrow}$ (or superior) contains elements of F^{-} that do not belong to any other intersection set and superior to ∂F elements on the basis of i dimension.

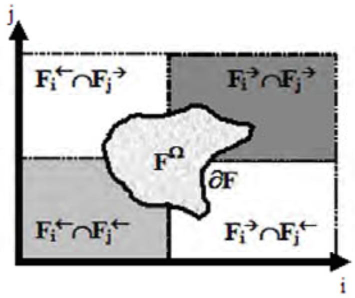
Figure 4. Intersection sets of one-dimensional feature



If we consider a feature of two dimensions i and j (as the 2D shape of a salient object in Figure 5), we then define four exterior intersection subsets:

- $F_i^{\leftarrow}\cap F_j^{\leftarrow}$ (or inferior) contains elements of F^{-} that do not belong to any other intersection set and inferior to F^{Ω} and ∂F elements on the basis of i and j dimensions.
- $F_i^{\rightarrow}\cap F_j^{\rightarrow}$ (or superior) contains elements of F^{-} that do not belong to any other intersection set and superior to F^{Ω} and ∂F elements on the basis of i and j dimensions.
- $F_i^{\leftarrow}\cap F_j^{\rightarrow}$ contains elements of F^{-} that do not belong to any other intersection set and inferior to F^{Ω} and ∂F elements on the basis of i dimension, and superior to F^{Ω} and ∂F elements on the basis of j dimension.
- $F_i^{\rightarrow}\cap F_j^{\leftarrow}$ contains elements of F^{-} that do not belong to any other intersection set and superior to F^{Ω} and ∂F elements on the basis of i dimension, and inferior to F^{Ω} and ∂F elements on the basis of j dimension.

Figure 5. Intersection sets of polygonal shape



More generally and using the same reasoning, we are able to determine intersection sets (2n) of n dimensional feature.

In addition, we use a tolerance degree in the feature intersection sets definition in order to represent separations between sets and to provide a simple flexibility parameter for the computing process. For this purpose, we use two tolerance thresholds:

- Internal threshold ε^i: defines the distance between F^{Ω} and ∂F,
- External threshold ε^e: defines the distance between subsets of F^-.

Relations Computing Via Intersection Matrix

To calculate relation between two feature values, we establish an intersection matrix of their corresponding feature intersection sets. Matrix cells have binary values:

- 0 whenever intersection between sets is empty
- 1 otherwise

For two-dimensional feature (such as the shape) of two values A and B, we obtain the following intersection matrix.

The indexing relation between two values is expressed then by a binary value which is the juxtaposition of each raw of the corresponding intersection matrix. This is very important for indexing purposes.

Indexing Images

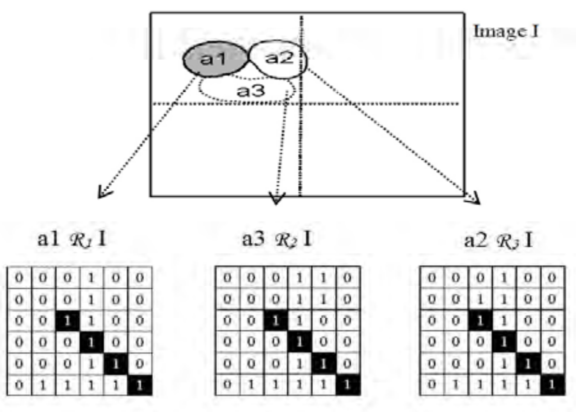
To index an image, we proceed as follows. First, we identify the spatial relations between the whole image and its salient objects. This allows determining the relative position of each object inside the image. In this case, the shape is used as a feature. The image is considered as a special object[1] that encloses all salient objects. We compute the intersection matrix of each pair of Image/Salient object. Figure 7 shows the result of this step on a 2D image I: three different spatial relations between the image I and its salient objects (a1, a2, and a3) are identified. It is important to mention that these relations are traditionally described by only one expression (i.e. northwest). We can see that, even at this level, our method provides a higher expression capability.

The second step consists of computing spatial relations between each pair of salient objects. Our

Figure 6. Intersection matrix of two values A and B on the basis of two-dimensional feature

$A^{\Omega}\cap B^{\Omega}$	$A^{\Omega}\cap\partial B$	$A^{\Omega}\cap B_1^{\leftarrow}\cap B_2^{\leftarrow}$	$A^{\Omega}\cap B_1^{\leftarrow}\cap B_2^{\rightarrow}$	$A^{\Omega}\cap B_1^{\rightarrow}\cap B_2^{\rightarrow}$	$A^{\Omega}\cap B_1^{\rightarrow}\cap B_2^{\leftarrow}$
$\partial A\cap B^{\Omega}$	$\partial A\cap\partial B$	$\partial A\cap B_1^{\leftarrow}\cap B_2^{\leftarrow}$	$\partial A\cap B_1^{\leftarrow}\cap B_2^{\rightarrow}$	$\partial A\cap B_1^{\rightarrow}\cap B_2^{\rightarrow}$	$\partial A\cap B_1^{\rightarrow}\cap B_2^{\leftarrow}$
$A_1^{\leftarrow}\cap A_2^{\leftarrow}\cap B^{\Omega}$	$A_1^{\leftarrow}\cap A_2^{\leftarrow}\cap\partial B$	$A_1^{\leftarrow}\cap A_2^{\leftarrow}\cap B_1^{\leftarrow}\cap B_2^{\leftarrow}$	$A_1^{\leftarrow}\cap A_2^{\leftarrow}\cap B_1^{\leftarrow}\cap B_2^{\rightarrow}$	$A_1^{\leftarrow}\cap A_2^{\leftarrow}\cap B_1^{\rightarrow}\cap B_2^{\rightarrow}$	$A_1^{\leftarrow}\cap A_2^{\leftarrow}\cap B_1^{\rightarrow}\cap B_2^{\leftarrow}$
$A_1^{\leftarrow}\cap A_2^{\rightarrow}\cap B^{\Omega}$	$A_1^{\leftarrow}\cap A_2^{\rightarrow}\cap\partial B$	$A_1^{\leftarrow}\cap A_2^{\rightarrow}\cap B_1^{\leftarrow}\cap B_2^{\leftarrow}$	$A_1^{\leftarrow}\cap A_2^{\rightarrow}\cap B_1^{\leftarrow}\cap B_2^{\rightarrow}$	$A_1^{\leftarrow}\cap A_2^{\rightarrow}\cap B_1^{\rightarrow}\cap B_2^{\rightarrow}$	$A_1^{\leftarrow}\cap A_2^{\rightarrow}\cap B_1^{\rightarrow}\cap B_2^{\leftarrow}$
$A_1^{\rightarrow}\cap A_2^{\rightarrow}\cap B^{\Omega}$	$A_1^{\rightarrow}\cap A_2^{\rightarrow}\cap\partial B$	$A_1^{\rightarrow}\cap A_2^{\rightarrow}\cap B_1^{\leftarrow}\cap B_2^{\leftarrow}$	$A_1^{\rightarrow}\cap A_2^{\rightarrow}\cap B_1^{\leftarrow}\cap B_2^{\rightarrow}$	$A_1^{\rightarrow}\cap A_2^{\rightarrow}\cap B_1^{\rightarrow}\cap B_2^{\rightarrow}$	$A_1^{\rightarrow}\cap A_2^{\rightarrow}\cap B_1^{\rightarrow}\cap B_2^{\leftarrow}$
$A_1^{\rightarrow}\cap A_2^{\leftarrow}\cap B^{\Omega}$	$A_1^{\rightarrow}\cap A_2^{\leftarrow}\cap\partial B$	$A_1^{\rightarrow}\cap A_2^{\leftarrow}\cap B_1^{\leftarrow}\cap B_2^{\leftarrow}$	$A_1^{\rightarrow}\cap A_2^{\leftarrow}\cap B_1^{\leftarrow}\cap B_2^{\rightarrow}$	$A_1^{\rightarrow}\cap A_2^{\leftarrow}\cap B_1^{\rightarrow}\cap B_2^{\rightarrow}$	$A_1^{\rightarrow}\cap A_2^{\leftarrow}\cap B_1^{\rightarrow}\cap B_2^{\leftarrow}$

Figure 7. The spatial relations between the image and its 3 salient objects



method allows combining both directional and topological relation into one binary relation. The directional and topological relations are replaced by an expressive binary spatial relation between two salient objects. Figure 8 shows the result of this step where a distinguished[2] spatial relation is identified for each pair of salient objects of the image I.

As a result, the image I is indexed as:

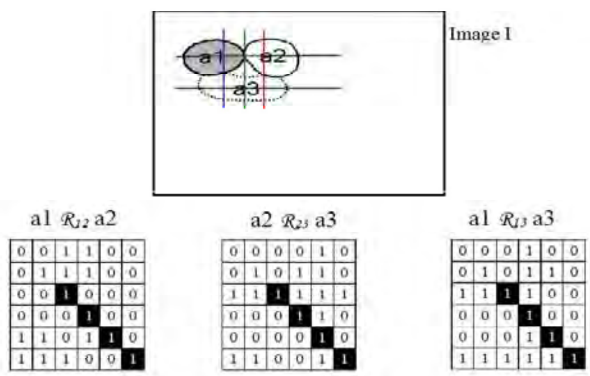
$$(a1\ R_1\ I + a2\ R_2\ I + a3\ R_3\ I) +$$
$$(a1\ R_{12}\ a2 + a2\ R_{23}\ a3 + a1\ R_{13}\ a3)$$

Indexing Evolutionary Content

To index evolutionary content of images, we compare some of their features values. We consider that calibrating techniques are already applied before comparing. In Chbeir et al. (2001b), we identified several types of evolutionary possibilities in the medical domain. In this paper, we address only two major possibilities: the displacement and the transformation of salient object.

The displacement represents the evolution of salient object position between two images. For example, the displacement of a fibrinocruorique clot through a deep vein of the lower limb is very common in medicine. The displacement of a salient object can be captured by comparing its relative spatial relations with images, and then its spatial relations with other salient objects. Using our method, even similar and complex spatial situations are detected (see following below).

Figure 8. The spatial relations between salient objects



The transformation represents the evolution of an object shape between two images. For example, the expansion of tumor size inside the left lobe of a brain is an example of this type of evolution. The transformation can be expressed as deformation, contraction, stability, and so on. To detect salient object transformation, we consider the surface as a feature. In this case, the exterior set F^- is considered as indivisible. We determine then the relation between the surface changes of the same salient object detected in two images. In the medical domain, it is obvious that the two images must belong to the same patient. Figure 9 snapshots some of the possible evolutionary relations that can be detected.

APPLICATION EXAMPLE

Let us consider a medical image with three salient objects a1, a2, and a3 (Figure 10) where shapes have changed during a period of time. These two situations resemble similar but actually different. The distinction between them must be well expressed especially in the application domains that require precision (such as in medicine). As shown in Figure 10, our method is capable to distinguish the different situations that are traditionally impossible to detect. The relations R_{12} and R'_{12} are equal but both relations R_{23} / R'_{23}, and R_{13} / R'_{13} are clearly distinguished.

Furthermore, our indexing method allows detecting the salient objects transformations: the deformation of a3, and the contraction of both a1 and a2.

CONCLUSION

We presented the domain of spatio-temporal image indexing and an original method capable of considering different types of relations in image representation. With our method, it is possible to homogenize, reduce and optimize the relations in

Figure 9. Several evolutionary relations detected

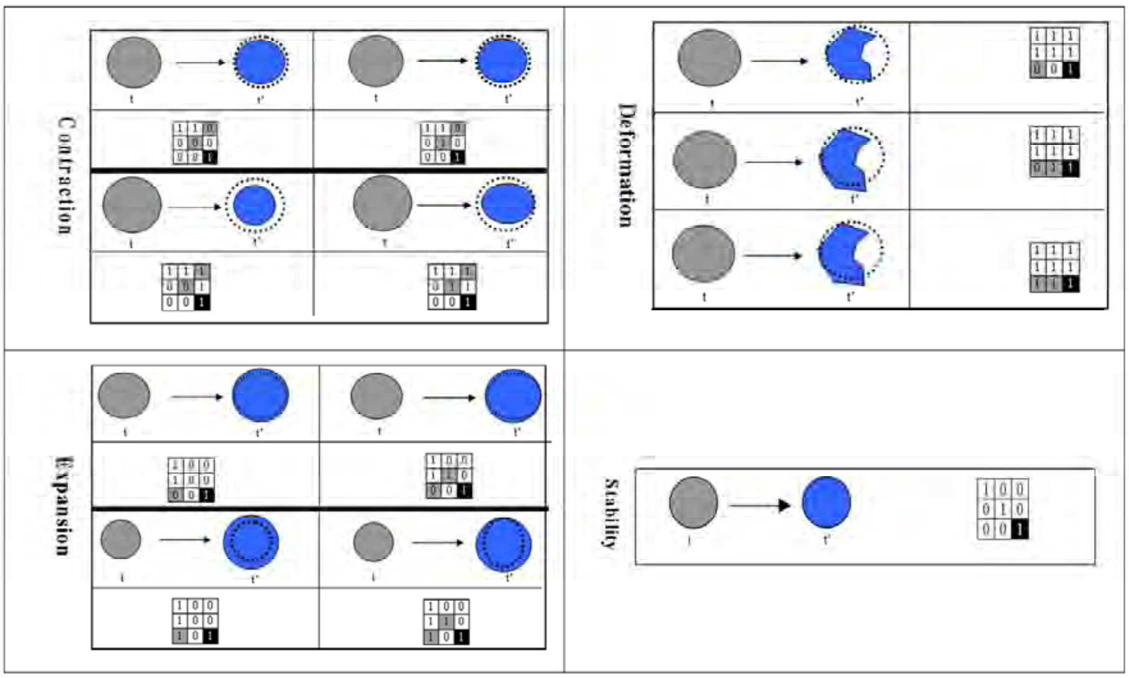


Figure 10. Using our method, the identification of similar spatial situations is possible

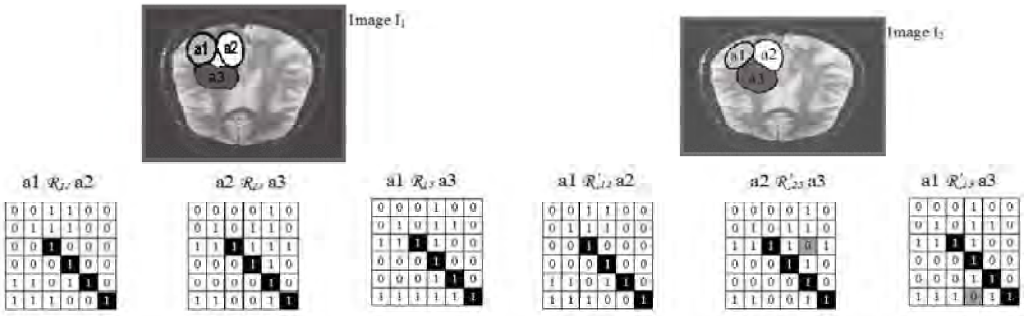


image description models (Chbeir et al., 2002). In this article, we showed how to provide a highly expressive power to spatial relations that can be applied to describe images and then to formulate complex visual queries. We also showed how to detect evolutionary content of images. The example in medical domain demonstrates how image indexing can be improved.

Currently, we are working on integrating such indexing method in our prototype EMIMS (Chbeir et al., 2001a; Chbeir et al., 2002). Future work includes considering indexing of low-level features (color, texture, etc.) and more intense experiments in complex environment where large number of feature dimensions (2,5D[3] and 3D images) and salient objects exist. Our method will also be studied to see if it can used in datagrid computing.

REFERENCES

Allen, J.F. (1983). Maintaining knowledge about temporal Intervals. *Communications of the ACM, 26*(11), 832-843.

Beckmann, N. (1990). The R*-tree: An efficient and robust access method for points and rectangles. *SIGMOD Record, 19*(2), 322-331.

Bonhomme, C., Trepied, C., Aufaure, M.A., & Laurini, R. (1999). A visual language for querying spatio-temporal database. *Proceedings of the 7th International Symposium on GIS, ACMGIS' 99*, Kansas City, (pp. 34-39).

Cardenas, A.F., Ieong, I.T., Taira, R.K., Barker, R., & Breant, C. (1993). The knowledge-based object-oriented PICQUERY+ language. *IEEE-Transactions-on-Knowledge-and-Data-Engineering, 5*(4), 644-657.

Chang, S.K. & Jungert, E. (1991). Pictorial data management based upon the theory of symbolic projections. *Journal of Visual Languages and Computing, 2*(3), 195-215.

Chang, S.K. & Jungert, E. (1997). Human- and system-directed fusion of multimedia and multimodal information using the sigma-tree data model. *Proceedings of the 2nd International Conference on Visual Information Systems*, San Diego, (pp. 21-28).

Chang, S.K., Shi, Q.Y., & Yan, C.W. (1987). Iconic indexing by 2-D strings. *IEEE-Transactions-on-Pattern-Analysis-and-Machine-Intelligence, PAMI-9*(3), 413-428.

Chbeir, R., Amghar, Y., & Flory, A. (2001a). A prototype for medical image retrieval. *International Journal of Methods of Information in Medicine, Schattauer,* 3, 178-184.

Chbeir, R., Amghar, Y., Flory, A., & Brunie L. (2001b). A hyper-spaced data model for content and semantic-based. *Proceedings of ACS/IEEE International Conference on Computer Systems and Applications*, Beirut, Lebanon (pp. 161-167).

Chbeir, R., Atnafu, S., & Brunie, L. (2002). Image data model for an efficient multimcriteria query: A case in medical databases. *The 14th International Conference on Scientific and Statistical Database Management, IEEE Computer Society Press*, Edinburgh, Scotland, July 24-26 (pp. 165-174).

Chu, W.W., Hsu, C.C., Cárdenas, A.F., et al. (1998). Knowledge-based image retrieval with spatial and temporal constructs. *IEEE Transactions on Knowledge and Data Engineering, 10*(6), 872-888.

Egenhofer, M. & Franzosa, R. (1991). Point-set topological spatial relations. *International Journal of Geographical Information Systems*, *5*(2), 161-174.

Egenhofer, M. & Herring, J. (1991). *Categorising binary topological relationships between regions, lines, and points in geographic databases, a framework for the definition of topological relationships and an algebraic approach to spatial reasoning within this framework.* Technical Report 91-7, National center for Geographic Information and Analysis, University of Maine, Orono.

Egenhofer, M., Frank, A. & Jackson, J. (1989). A topological data model for spatial databases. Symposium on the Design and Implementation of Large Spatial Databases, Santa Barbara, CA. *Lecture Notes in Computer Science*, *409*, 271-286.

El-kwae, M.A. & Kabuka, M.R. (1999). A robust framework for content-based retrieval by spatial similarity in image databases. *ACM Transactions on Information Systems, 17*(2), 174-198.

Guttman, A. (1984). R-trees: A dynamic index structure for spatial searching. *SIGMOD Record, 14*(2), 47-57.

Lin, K.I., Jagadish, H.V., & Faloutsos, C. (1994). The TV-tree: An index structure for high dimensional data. *Journal of Very Large DataBase, 3*(4), 517-549.

Lomet, D.B. & Salzberg, B. (1990). The hb-tree: A multiattribute indexing method with good guaranteed performance. *ACM Transactions on Database Systems, 15*(4), 625-658.

Peuquet, D. J. (1986). The use of spatail relationsships to aid spatial database retrieval. *Proceedings of the 2nd International Symposium on Spatial Data Handling*, Seattle (pp. 459-471).

Rui, Y., Huang, T.S., & Chang, S.F. (1999). Image retrieval: Past, present, and future. *Journal of Visual Communication and Image Representation*, *10*, 1-23.

Smeulders, A.W.M., Gevers, T., & Kersten, M.L. (1998). Crossing the divide between computer vision and databases in search of image databases. *Proceedings of the Visual Database Systems Conference*, Italy (pp. 223-239).

White, D.A. & Jain, R. (1996). Similarity indexing with the SS-tree. *Proceedings of the 12th International Conference on Data Engineering*, New Orleans, Louisiana (pp. 516-523).

Yoshitaka, A. & Ichikawa, T. (1999). A survey on content-based retrieval for multimedia databases. *IEEE Transactions on Knowledge and Data Engineering*, *11*(1), 81-93.

KEY TERMS

Directional Relation: Describes the order between two salient objects according to a direction, or the localization of salient object inside images. In the literature, 14 directional relations are considered:

- **Strict:** north, south, east, and west.
- **Mixture:** north-east, north-west, south-east, and south-west.
- **Positional:** left, right, up, down, front and behind.

Image Indexing: Consists of assigning concise and significant descriptors to an image. Objects shapes and positions are used in several indexing approaches where image is represented as a graph or tree (R-tree, B-tree, etc.).

Metric Relation: Measures the distance between salient objects. For instance, the metric relation "far" between two objects A and B indicates that each pair of points A_i and B_j has a distance grater than a certain value d.

Multimedia Document: Represents a document containing not only textual data but also multimedia ones such as images, videos, songs, and so on.

Salient Object: Interesting or significant object in an image (sun, mountain, boat, etc.). Its computing changes in function of the application domain.

Spatio-Temporal Relation: A combination of two spatial and temporal relations into one relation used to index video-documents and objects evolution between two images (e.g., tumor evolution).

Topological Relation: Describes the intersection and the incidence between objects. Six basic relations have been identified in the literature: *disjoint, meet, overlap, cover, contain, and equal*.

ENDNOTES

1. Where the interior set is empty, the boundary set is the barycentre, and the exterior is divided into four sets.
2. Traditionally, the three salient objects are described by non expressive spatial relations (same topological and directional relations).
3. Sequence of 2D images in medical applications (i.e., scanner).

This work was previously published in Encyclopedia of Multimedia Technology and Networking, edited by M. Pagani, pp. 257-264, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global)

Section III
Tools and Technologies

This section presents extensive coverage of the interaction between healthcare and the various technologies that both derive from and support the research of practitioners, students, and patients alike. These chapters provide an in-depth analysis of mobile technologies and their use in healthcare, while also providing insight into new and upcoming technologies, theories, and instruments that will soon be commonplace. Within these rigorously researched chapters, readers are presented with countless examples of the tools that facilitate and support medical innovations. In addition, the successful implementation and resulting impact of these various tools and technologies are discussed within this collection of chapters.

Chapter 3.1
Application of Mobile Technologies in Healthcare Diagnostics and Administration

Ketan Vanjara
Microsoft Corporation, India

ABSTRACT

This chapter explores various advancements in mobile devices and corresponding software applications that enhance diagnostics and administration in the healthcare domain. This chapter further proposes networking and integration of these devices with the existing networks and devices as further development in healthcare.

A ROAD ACCIDENT SCENARIO

In 1992, a colleague of mine was traveling from our offices in western India to our manufacturing facilities about 50 kilometers away. He had worked late the previous night and again come to the office early in the morning. We had our lunch together and he left for the manufacturing facilities (factory) around noon, driving his own car. Even after two hours, he had not reached our factory and, increasingly, the office and the factory administrators were panicking, trying to locate this employee of ours. Around 6:30 p.m. we received a call from the general hospital of the city that a semi-conscious accident patient had been admitted to the hospital. In short, my colleague had met with an accident around 1 p.m., and he was admitted to the hospital only around 6 p.m. The treatment could not begin before 8 p.m. The end result was that he became paraplegic. Based on the information provided by the doctor on duty, had he been treated within three hours of the accident, paraplegia could have been avoided. Did mobile devices—or their lack of—have a role to play in this incident? The answer has to be a 'yes'.

Though just a little over a decade ago, this era was devoid of any mobile devices in India. My colleague did not have a cell phone, nor did his family (it took a while to contact them). His car did not have a Global Positioning System (GPS). The hospital and doctors too did not have any mobile devices. Finally, even after arriving at the hospital, the doctors had to wait for the family before starting treatment, as they needed medical information of the patient such as existing diseases, allergies, sensitivities, and so on. Overall, the above scenario shows the lack of our ability to manage emergency medical situations without immediate communication. Today, though, with the advancement of mobile technologies, such events are better managed. However, there is still a need to provide integration of various mobile tools with the existing devices, technologies, and networks in order to reap full benefits of mobility in the healthcare sector in general and emergency medical situations in particular.

This chapter explains how mobile technologies are shaping the face of healthcare today. This chapter also discusses the potential for improved healthcare through mobility in the future. The primary focus of this chapter is on the integration of mobile devices/gadgets in diagnostics, healthcare administration, and healthcare information systems. Finally, this chapter also proceeds to identify issues and challenges involved in such integrations and how to handle them.

THE HEALTHCARE LANDSCAPE

Healthcare is one of the major domains in life that is full of amazing advancements. However, it also comes with its ever evolving challenges. For example, of late this domain has been mired in controversies on ethical issues like cloning and stem cell research. While medical science still continues to struggle with challenges like eradication of malaria and treatment of cancer and AIDS, its achievements have not been mean by any standards.

The progress made in the last 20 years (is) amazing if one looks at it now. In short, the advances made in the last 10-15 years are equal or even surpass the advances made in the last 100 years. (Majeed, 2005)

Medical science has discovered new diseases and invented new treatments and therapies. In fact, a number of medical scientific disciplines like Pharmacogenomics that allows creation of personalized medicine, toxicokinetics, proteomics, therapeutic vaccines, stem cell research, bioinformatics, and telemedicine, to name a few, evolved in the last 20 years.

Complex diagnostics, short-duration intensive care, acute medicines, and micro-surgeries are all becoming common in today's healthcare.

Another major contributor to advancements in medical sciences has been the developments in digital technologies. Be it bioinformatics or telemedicine, the need for modern healthcare to offer the best treatment for complex diseases in the fastest possible manner with resource constraints makes it highly dependent on digital technologies. Data-intensive areas like bioinformatics also need to depend heavily on high-end computing devices.

Thus, today's healthcare is characterized by the following:

- **Multiplicity and Complexity of Diseases and Treatments:** The breadth and depth of knowledge on various medical disciplines, diseases, their symptoms, causes, and various alternatives of treatment have been growing exponentially in the last couple of decades. This has led to increased complexity in the fields of diagnostics and therapies. So much has been the growth in knowledge and complexity that it is almost impossible for a medical practitioner to function effectively without supporting devices like computers and PDAs.
  An indicator of this fact is the vast amount of literature available on all the medical disciplines—a simple search on Google (2005) returns thousands of medical journals, 1,440 of which are free. Hundreds of viruses and their mutations have been identified for something as simple as the common cold (Health-Cares, 2005).
- **Quick Response:** Given that the number of accidents is increasing every day, the demand for quick response from medical services is growing too. Continuous efforts

are on to reduce the interval between the time a person needs medical attention to the time when treatment is provided to her. This has led to provisioning of preliminary treatment as soon as the patient is contacted—at the accident site, at the patient's home, or even in the ambulance taking the patient to a clinic or hospital. Mobile clinics and surgery ambulances are quite common nowadays to meet this need of quick response.

- **Quick Cure:** Quick response alone is not enough. A typical patient today expects to get cured much faster so that he can get back to work and his normal life. This need drives the entire industry, especially pharmaceutical companies, to constantly research and produce better drugs and come up with drug delivery systems that offer faster cure (Howard, 2004).
- **Mobility and Remote Assistance:** Mankind is more mobile today than it has ever been before. This increase in mobility has given rise to the need for remote medical assistance. Despite their mobility, a patient and a doctor need to stay in touch with each other to get/provide remote assistance. Such remote assistance can take various forms, ranging from simple things like monitoring a patient's progress and response to treatment, to more complex and technology-intensive issues like telemedicine or tele-surgery.
- **Cost-Price Paradox:** While it is quite obvious that the emerging nature of medical services as well as demands on them result in increased costs, there is constant pressure to drive the price of healthcare southwards. Besides providing better value for money to patients, reduced price also makes healthcare affordable to many more people. Thus, it is imperative upon various players in the healthcare sector to provide superior and continuously improving products and services at continuously reducing costs. This requires a lot of automation and innovation.
- **Resource Constraint:** Increase in multiplicity and complexity of medical disciplines, diseases, and therapies requires specialized skills. Thus, we have specialists in every branch and sub-branch of medicine. By their very nature, specialist skills are scarce. However, the demand on such specialist skills keeps on rising. This creates a resource constraint and a need to be able to utilize the same resources effectively across different locations or geographies.
- **Data Driven:** Given the plethora of variables that need to be considered and evaluated to provide effective treatment, the medical profession has become highly data driven. Besides supporting research, such data actually influences, if not drives, treatment. Such dependence on volumes of data necessitates use of digital devices.
- **Return of Alternative Therapies:** These include medical disciplines like the age old ayurveda, naturopathy, homeopathy, and so on. While their acceptance is growing among many people all over the globe, technology has not made as many inroads into these branches of medicine as others. This presents a huge opportunity. We shall discuss one such potential in this chapter.

THE MOBILE TECHNOLOGY LANDSCAPE

After surveying the healthcare landscape, let us take a look at the mobile technology landscape as well. The handheld devices of today, like cell phones and PDAs, are far more powerful than the large supercomputers of yesteryears that occupied large space. In addition to this shrinking of sizes, there are quite a few other significant developments taking place in this arena. Some of the recent developments that help us understand the mobile technology landscape are as follows:

- **Nanotechnology—Miniaturization Shift:** Handheld devices and various other gadgets related to mobile technologies are getting smaller and smaller. The scale is moving from micro to nano. Nanotechnology, considered to be one of the foundations of future technology revolutions, enables many unthinkable things like storing several gigabytes of data on a millipede nanodrive of the size of a postage stamp (Global Change, 2005). We already get 60 GB of storage in an iPod (http://www.apple.com/ipod/). PDA PC watches and pocket-sized holographic projectors are on their way to the market. One of the constraints of adopting mobile technologies has been smaller output interfaces in the form of display screens. Hologram projectors resolve this problem effectively (http://www.lightblueoptics. com/index.html; Howard, 2004).
- **Software Power:** While hardware devices are offering more (functionality) in less (smaller devices), software is not to be left behind. There are thousands of applications created for these small devices that have larger functionality and very small footprint. Be it Win CE operating system, various MS Office applications, or entertainment software like iTunes_all of them provide extensive functionality while occupying much less space on the devices (Microsoft, 2005).
- **Convergence:** The world talked a lot about convergence of information and communication technologies in late nineties. However, such convergence is an ongoing process that is still continuing. What we are seeing today is convergence of various advanced technologies like computing, mobile telephony (cell phones), and connectivity (Internet) on a single handheld device like the HP iPAQ 6315 (2005) Pocket PC–Phone Edition. Pretty soon, positioning (GPS) too will be integrated. Handheld GPS receivers are already available in the market. Microsoft's product *Streets & Trips 2005* with handheld GPS locator combines Streets & Trips software with a GPS plug-n-play receiver hardware component, to create a planning and driving tool that turns a laptop into a real-time GPS tracking device (Microsoft Streets & Trips 2005 GPS, 2005). Very soon we may see all these features available on nano devices, resulting in the handheld devices having more processing power, superior features, larger storage, longer battery life, and becoming more universally accessible than the tablet PC of today. Such convergence of technologies will empower and enable users to work far more efficiently and effectively than ever_and from wherever.
- **Natural Interfaces.** Another major area of advancement in technologies is natural interfaces. These include voice recognition, handwriting recognition, and visual recognition. One of the greatest hindrances to adoption of mobile devices has been the input interfaces in the form of the keyboard or mouse. The smaller keypads of handheld devices make it difficult to input large content. Voice and handwriting recognition address this issue. Visual recognition addresses issues related to security. Very soon we will also see interactive surfaces like floors, tables, walls, and so on (http://www.naturalinteraction.org, http://www.hcirn.com/, http://nooface.net).
- **Function-Specific Handheld or Nano Devices.** While information and communication technology industries are creating newer, smaller, and smarter versions of technology-driven handheld devices like cell phones, pocket PCs, and PDAs, industries like healthcare are creating and continuously improving functionality-driven handheld devices to perform specific functions related to healthcare. These include diagnostic devices like Smart Pill from Smart Pill

Diagnostics and Amplichip from Roche (Medgadget, 2005a).

- **Mobility and Wireless:** With mobility comes the need of wireless. One cannot be truly mobile with wired devices. In addition to wireless devices, we are also seeing the creation of connectivity infrastructure in the form of cellular and Wi-Fi networks.

  A cellular radio network is a radio network made up of a number of radio cells (or just cells) each served by a fixed transmitter, normally known as a base station. These cells are used to cover different areas in order to provide radio coverage over a wider area than the area of one cell. (Cellular Network, 2005)

  Wi-Fi networks (also known as wireless LAN) use radio technologies called IEEE 802.11b or 802.11a to provide a secure, fast, and reliable wireless connection. Wi-Fi functions through a transmitting antenna which is usually linked to a DSL or high-speed land-based Internet connection and uses radio waves to beam signals. Another antenna, which is in the laptop or PC, catches the signal. The signal, usually, has a range of about 300 feet for most home connections. The farther the user is from the signal, the slower the connection speed. (Shaheen, 2005)

  Our cell phones work on a cellular network, while mobile computers use various other wireless technologies to network, like infrared, Wi-Fi, radio-frequency, Bluetooth, and so on. Recently, cell phones that can work on both cellular and Wi-Fi networks, and also seamlessly switch between the two, have also arrived (Computerworld, 2004).
- **Cost Paradox:** The cost paradox in the technology landscape is quite opposite of that in the healthcare space. In technology areas, the normal trend is to get more for less. From memory to storage to functionality—hardware or software—we either get the same product cheaper or a better product at the same price with passing time. This puts technology in an enviable position where it can be and it is complementary to all other industries or needs.
- **Business Intelligence:** Due to increasing specialization and complexity in various domains today, decision making in today's world is a function of a multitude of variables. This has necessitated creation of huge databases. These databases need to be analyzed almost real time to support decision making. Data mining and business intelligence come in handy here.

APPLYING MOBILE TECHNOLOGIES IN HEALTHCARE

Having surveyed the healthcare and mobile technologies landscape, we now move on to explore the possibilities of convergence of both. Though the possibilities are endless, we will keep our exploration limited to:

- **Diagnostics:** With the exception of certain alternate medical sciences, treatment of therapy nowadays is seldom provided without diagnosing the symptoms through a plethora of tests in the areas of pathology and radiology, as well as physical examination. The field of diagnostics, therefore, offers many challenges—in other words opportunities—for mobile technologies.
- **Healthcare Administration:** Healthcare administration includes the complete cycle of healthcare provisioning from the moment a patient seeks the service until the patient is cured or no more needs the service. This includes activities like consulting, symptom recording, diagnostics, analysis and interpretation of test reports, prescription of treatment, therapy or surgery, drug delivery, post-operative support, and so on. Due to

various characteristics of the healthcare sector mentioned earlier, mobile devices can make a lot of contribution in this area as well.

- **Healthcare Information Management Systems:** This includes creation, maintenance, and global provisioning of all healthcare-related information like medical records, test reports, receipts, payments, service levels, and so on. Not all information is critical such that it needs to be provided anytime, anywhere. However, information like medical records, test reports, service levels, and so forth do need to be universally available.

MOBILE TECHNOLOGIES IN DIAGNOSTICS

Among all the applications of mobile technologies in the healthcare space, diagnostics is the most challenging as it involves integration of multiple technologies. A patient has to go through a series of diagnostic tests to enable a doctor to provide appropriate treatment. There are two major bottlenecks around this: (1) in urban areas, where the diagnostic labs are normally located, the patient population is very mobile and also adversely affected by traffic conditions of modern urban areas; and (2) in rural areas, where the patients are not that mobile and rarely have a diagnostic lab. Mobile technologies can resolve both these problems. In addition, mobile technologies can also assist in cases of emergencies, like accidents or elderly patients that cannot travel much.

Conventional diagnostics can be broadly classified into four areas: pathology, radiology,

scans (MRI, CT, etc.), and pulse (ECG, physical pulse examination, etc.).

The good news is that today we have miniature devices that are useful in all types of diagnostics. In the beginning, personal wellness monitoring meant the bathroom scale and the mercury thermometer. Then they went electronic, and home blood-pressure devices became common. Today, it would not be shocking to see a home cardiac defibrillator sitting in the corner (HealthDay, 2005). However, only some of the diagnostics devices have become popular so far. Others are at various stages of development, clinical trial, or market deployment.

Let us briefly look at some of the handheld or miniature devices for diagnostics that are already being used or about to be deployed in the healthcare market. A brief description is given on the functionality of each device (http://www.medgadget.com/archives/diagnostics/index.html).

- **Blood Glucose Monitor Systems/Diabetes Systems:** These devices help in monitoring/determining the glucose levels in blood. They have become quite sophisticated over a period of time, and the latest models have many features like no strip handling, 250-test memory, and so on.
- **Wrist BP Monitors:** They look very similar to electronic wrist watches and are used in determining the systolic and diastolic blood pressure of a patient. The advanced versions can store up to 90 readings and are very lightweight, operating on AA or AAA batteries.
- **Oximeter:** These devices determine the degree to which hemoglobin on the red blood cell is loaded with oxygen ("oxygen saturation"). Sophisticated models have additional features like measurement of pulse rate and SpO2, verification of endotracheal tube placement, and changes in carbon dioxide level with the help of a CO_2 sensor.
- **Automatic External Defibrillators/Pacemakers/CHF Monitor/Fluid Status Monitoring:** These are electrical devices used to counteract fibrillation of the heart muscle and restore normal heartbeat by applying a brief electric shock. In addition, they have built-in monitors that track the problem to

prevent crisis. The fluid status monitoring device sits under the patient's left collarbone and sends out electrical impulses to monitor liquid levels; whenever the level is near critical, the device alerts the patient to get help before the problem turns critical. Doctors also can check the monitors by phone, using computerized hookups in their patients' homes.

- **iPod:** The iPod is not just for music any more. Radiologists in the United States, Europe, and Australia are now using iPod devices to store medical images. "This is what we call using off the shelf, consumer market technology," says Osman Ratib, MD, PhD, professor and vice-chairman of radiologic services at UCLA. "Technology coming from the consumer market is changing the way we do things in the radiology department" (Ratib, 2004). A software (OsiriX) has also been developed to automatically recognize and search for medical images on the iPod. When it detects the images, they automatically appear on the list of image data available—similar to the way music files are accessible by the iTunes music application.
- **Portable Patient Monitors:** These devices monitor and provide digital display of mean and diastolic blood pressures, heart rate and respiratory rate, temperature (a, b, or differential A-B) based on two temperature inputs, blood pressure, and ECG trace. Several alarms can also be set in these monitors based on over/under pulse rate, pressure, or temperature (http://www.pemed.com/pmonitor/pmon.htm).
- **GI Monitoring System:** A patient swallows this slightly larger-than-multi-vitamin-sized capsule containing two subminiature radio transmitters. The capsule's transmissions are received by a small, patient-worn mobile receiver/controller. Site-specific data is captured in real time as the capsule passes through the GI tract. It also measures pH, temperature, and pressure. The capsule is not absorbed, nor does it interact with the GI tract in any way other than its propulsion via peristalsis, eventually exiting through the colon. The future-generation systems are expected to enable real-time, site-specific tracking of the capsule's position within the GI tract, transmit pictures, and also facilitate drug delivery.
- **Handheld MRI:** The primary reason for the large size of conventional MRI scanners is the use of large magnets that have to be chilled by complicated cooling systems. Igor Savukov and Michael Romalis of Princeton University have shown that a device called an atomic magnetometer can detect magnetic signals from water without giant magnets or complicated cooling systems. This has given rise to a distinct possibility of handheld MRI scanners in the future that would image tissues inside the body as easily as a digital camera takes a photo.
- **Wireless/Handheld X-Ray Sensor:** A company has developed a wireless direct digital radiography system for dental x-ray examinations that communicates images it captures with a dentist's computer. There are quite a few models of handheld or portable x-ray instruments available now. They provide greater flexibility, are optimal for remote uses, capture high-quality images, transfer them digitally to computers, and reduce radiation exposure.
- **Oral Swab Kit:** Various organizations are prototyping oral swab kits or similar devices for saliva-based diagnostics that can detect exposure to a variety of substances, from narcotics to anthrax to common bacteria and viruses including those of HIV. These will be quantum enhancements to rapid saliva tests that are already in use. These devices would increase ease of detection and accelerate response time whether they are used

in the middle of a public health incident or in a busy doctor's office. The devices are expected to communicate captured data in digital form as

- **Wireless Miniature Devices for CHF Patients:** These are technologies integrated into minute implants, requiring no antenna, wires, or connecting leads, that allow a tiny device implanted deep inside the body to communicate wirelessly with other implanted devices and external systems. Such devices include a device that offers on-demand, non-invasive means to monitor intra-aneurysm pressures following endovascular graft procedures and a device for measurement of pulmonary artery pressure, which is the most important hemodynamic indicator in heart failure. A one-time, minimally invasive catheter-based procedure allows unlimited, non-invasive home-based monitoring of a CHF patient's hemodynamic status.
- **Other Handheld Wireless Diagnostic Devices:** These include holographic sensors, PathoTester (a wireless handheld system that can detect growth of microbial pathogens using nanocartridge-based technology that contains microbe-specific sensor holograms), non-invasive glucometer (that measures blood sugar in the body based on the amount of heat released from the tip of the finger—metabolic heat conformation), Glucose Monitoring Watch (which continuously displays blood glucose through use of photo acoustic waves originating in the blood vessels), and Wireless Objective Hearing Assessment Systems.
- **Pulse Meter for Ayurveda:** Ayurveda, the science of life and longevity, is an ancient Indian system of holistic medicine, which has re-emerged as an important form of holistic therapy throughout the world. Its essence is the conjunction of the body and the mind, and their coordination that achieves optimal health and happiness. This science is heavily dependent on pulse examination for diagnostics *Nadi Pariksha*. This involves determining *Prakriti* and *Vikriti* in pulse, understanding the seven levels of pulse, and integrating qualities of the pulse with signs and symptoms in clinical assessment. Future enhancements of pulse meters are expected to fulfill the needs of Ayurveda as well.

There are tele-healthcare gadgets galore, and the list can go on and on. Take for instance an all-in-one device that helps patients do the following: measure blood pressure, blood sugar, blood oxygen (SpO_2), measure temperature and weight, record peak flow, record stethoscope sounds, and take ECGs.

As of now many of these devices are not network ready for communication with other wireless devices like cell phones or PDAs, nor are all of them following any common communication standard. However, very soon they will be getting linked wirelessly to home computers and cell phones, as a health revolution is creeping quietly into our lives.

Where it is going to take us is anyone's guess, just as whether new age health monitoring will make us live longer. But the IBM Corp., for example, has developed a little electronic pillbox that may come in handy as the Baby Boomers edge into the years of forgetfulness. The box sends a signal to a mobile phone every time a pill is removed. If patients forget a pill, or take too many pills from the box, they get a friendly reminder phone call. For patients who need frequent monitoring of vital signs, an IBM wristband device measures blood pressure and heart rate at the push of a button, and these are transmitted automatically to medical personnel.

Sun Microsystems and MedicTouch have introduced the Pulse Meter, which they call the 'ideal health monitoring solution for sport enthusiasts, the elderly, rehabilitation outpatients and health-

care providers; providing health-monitoring anytime, anywhere.' A Pulse Meter user connects the sensor to a hand, starts the program on a mobile phone and the pulse is displayed within seconds on the phone screen, archived, customized for user, and transmitted. (HealthDay, 2005)

Medical monitoring and treatment are going wireless and Internet based, both at home and at the hospital. HealthDay (2005) reports about tele-healthcare gadgets designed for initial, basic in-home diagnostics.

NEXT STEPS

We are at the crossroads of mobile technologies and healthcare diagnostics today. While both of them converge here coming from two different directions, the convergence itself is expected to advance in two different directions simultaneously. These two directions are as follows:

1. **Integration of all the handheld diagnostic devices or their functionality** into a single device that is connected to the wired and wireless networks. Eventually, they would be interconnected with healthcare networks and databases (generic as well as medical records). This model would look somewhat similar to that in Figure 1.
   In fact, European researchers have just completed something very similar—the Wearable Health Care System project, otherwise known as WEALTHY (2005). In addition to being fashionable and slimming, the body suit will measure all manner

Figure 1. Network of all handheld devices (diagnostic and others) providing healthcare administration and healthcare information management systems-related services to doctors and patients. A single handheld diagnostic device with all the diagnostic functionality enables the patient to get desired healthcare services.

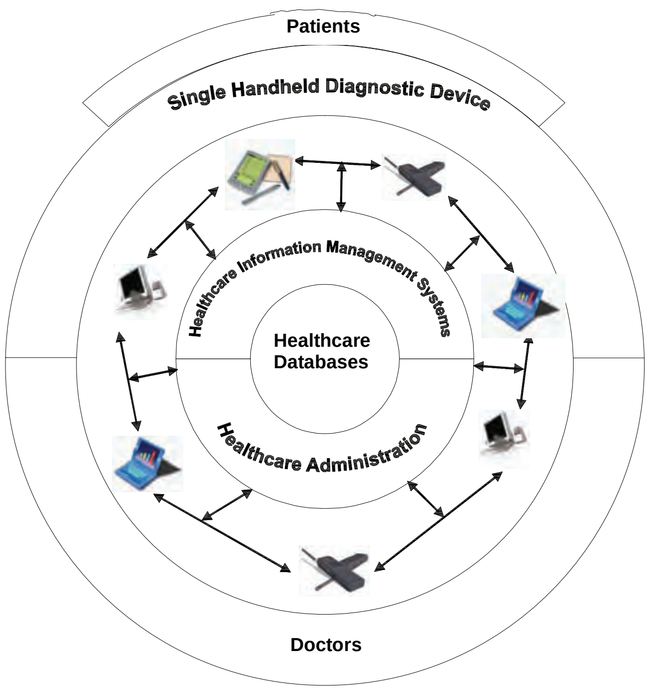


Figure 2. A representative diagram of WEALTHY

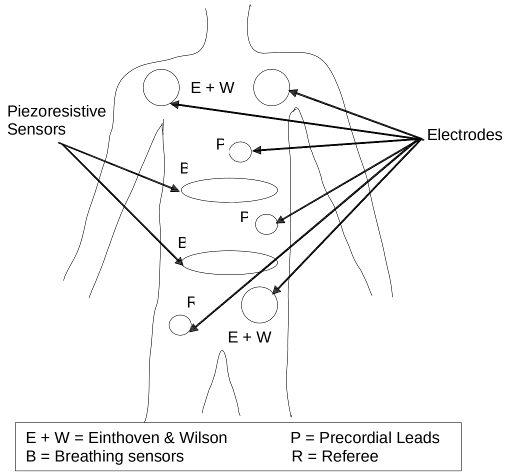
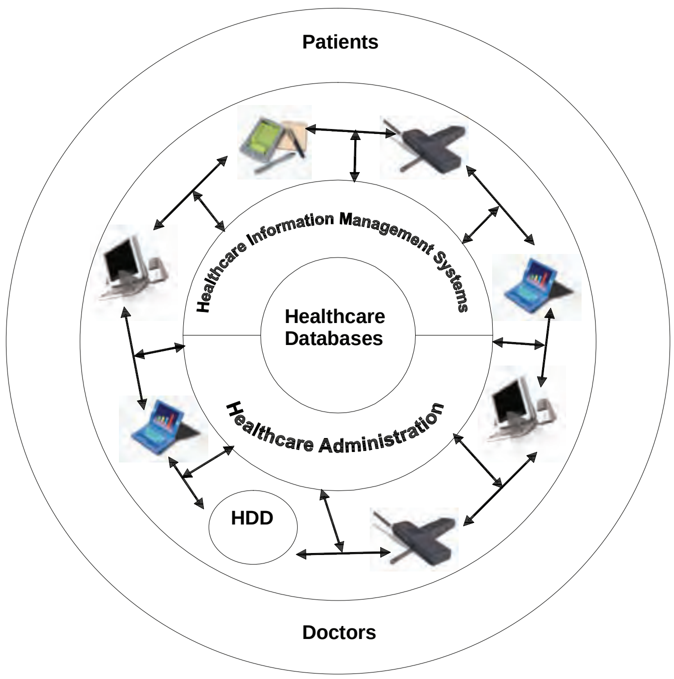


Figure 3. Network of all handheld devices (diagnostic and others) providing healthcare administration and healthcare information management systems-related services to doctors and patients. HDD = a network of handheld diagnostic devices that is integrated with wired and wireless networks. This entire network is accessible to doctors and patients.



of bodily processes, including respiration, core and surface skin temperature, position, and movement. It will also transmit all this data over the mobile phone network thanks to a miniaturized GPRS transmitter hidden in the suit. The transmitter can also alert emergency services if the wearer stops responding or shows alarming vital signs. WEALTHY is depicted in Figure 2.

2. **Networking of individual handheld diagnostic devices** amongst themselves and to wired and wireless networks. Eventually, they would also be interconnected with healthcare networks and databases (generic as well as medical records). This model would look somewhat similar to that shown in Figure 3.

MOBILE TECHNOLOGIES IN HEALTHCARE ADMINISTRATION

Healthcare administration consists of all the activities undertaken to provide healthcare services to patients. Broadly, these activities are:

- **Consultation**
  - Patient visits the doctor
  - Patient provides details of their ailment to the doctor
  - Doctor carries out physical examination for symptoms
  - Doctor may suggest further diagnostic tests—the probability of this is much higher today with the advancement of medical science
- **Diagnosis**
  - Patient visits various laboratories for diagnostic tests
  - Patient visits the laboratories again to collect the diagnostic test reports
  - Patient visits the doctor again with the diagnostic test reports
- **Treatment**
  - Doctor prescribes medication to the patient or recommends a surgery
  - Patient visits the pharmacy to buy the medication
  - If recommended, patient registers for the surgery at a hospital
- **Payment**
  - Patient makes payment to the doctor
  - Patient makes payment to the pharmacy
  - Patient approaches the insurance company or concerned agency for reimbursement of medical expenses.
  - Alternatively, the medical agencies (like the physician) approach the insurance company for reimbursement

What we need to explore now is how many of these steps can be performed effectively using mobile devices like cell phones or PDAs. I would not be overstating if I mentioned that most of the activities related to healthcare administration can be performed using various handheld devices. Table 1 shows how.

ADDITIONAL THOUGHTS

In addition to the normal healthcare administration process, there are many more things that can be achieved using handheld technologies. Some of them are:

- *Embedding important patient data on the smart cards* of mobile phones or PDAs. In cases of emergency like accidents, war, and so forth, quite often the patient is unable to provide any basic information about himself that is important for the doctor to initiate treatment. All such information, starting from the blood group of the patient, his address, contact details of his immediate relatives, his allergies, a brief medical

Table 1.

Activity	How to perform it using handheld device/s – there could be multiple options
Consultation • Patient visits the doctor • Patient provides details of their ailment to the doctor • Doctor carries out physical examination for symptoms • Doctor may suggest further diagnostic tests	• A personal visit by the patient to the doctor may not be required in many cases. Patient can talk to the doctor using cell phones and provide details of their ailment. • If needed, patient can also have a video conference with the doctor using cell phones with cameras & multi-media capabilities or smart phones – this would enable the doctor to observe the physical symptoms of the patient's ailment. • If a meeting with the doctor is absolutely necessary, a patient can access the doctor's calendar through their cell phone or PDA and take an appointment. Similarly doctors too can maintain and access their calendar through their cell phones or PDAs. • Having understood the ailment as well as symptoms, doctors can send a recommendation or prescription for diagnostics to the patient. Such a prescription can be created quickly using a pre-loaded form on the handheld device and then sent to the handheld device of patient as well as to the relevant laboratories.
Diagnosis • Patient visits various laboratories for diagnostic tests • Patient visits the laboratories again to collect the diagnostic test reports • Patient visits the doctor again with the diagnostic test reports	• As discussed at length in the previous section, most of the diagnostics can now be done through handheld devices. A patient in this case would have 3 choices – 1. If he is suffering from chronic ailment, he may have his own diagnostic device/s that he can use for the test and then send the result to the doctor, through a mobile device. 2. Patient can visit a laboratory to get the tests done. The laboratory, in turn, can flash the results to the handhelds of the doctor as well as the patient as soon as they are ready, saving the patient of a second visit. 3. With so many small handheld devices available, there is a good possibility of a doctor having them. Thus, patient can just pay one visit to the doctor and complete consultation, physical examination of symptoms as well as diagnostics – all at a time. In all the cases, the patient and doctor can achieve a lot with more convenience using handheld devices.
Treatment • The doctor prescribes medication to the patient or recommends a surgery • Patient visits the pharmacy to buy the medication • If recommended, patient registers for the surgery at a hospital	• After studying diagnostic reports, doctor can prescribe medication and therapies to the patient. This too can be done using their respective handheld devices. • A doctor can also take help from a lot of references (medical guides, reference, encyclopedia, etc.) preloaded on the handheld or accessible on the net through the handheld. • The prescription of medication and therapies duly signed by doctor can be transmitted to the patient's handheld device. • Patient can visit any pharmacy with the preloaded prescription from the doctor on the handheld device and buy the medicines by providing the prescription to the pharmacy. • If a surgery is recommended, the patient can take an appointment with the concerned doctor and surgery by accessing their calendar through the handheld. • In the surgery, if needed, the doctor/s can get medical advice / expertise from various resources on the net using their handheld devices or even having real-time interaction with experts around the world.
Payment • Patient makes payment to the doctor • Patient makes payment to the pharmacy • Patient approaches the insurance company or concerned agency for reimbursement of medical expenses. • Alternatively, the medical agencies (like the physician) approach the insurance company for reimbursement.	• Patient can pay the doctor / pharmacy using preloaded cash or credit/debit cards on the handheld device. • Patient can similarly receive all the invoices on his handheld and forward them to the insurance company along with a filled in prescribed form for reimbursement. • In case, the doctor has to approach the insurance agencies, they too can forward the relevant forms and documents in digital format from their handheld devices.

history consisting of all major healthcare-related events in the patient's life, and so on can be embedded on the smart card so that it is accessible at the click of a button. This can enable much faster treatment of the patient.
- *GPS enabling of handheld devices.* This also helps in tracking a patient during emergencies. On the diagnostics front, there are already devices that can keep track of a patient's health on certain key parameters and notify the doctor if a negative trend is building on any of them. An integration of these devices with GPS-enabled handheld devices would add a lot more value to provide proactive healthcare to patients.

Thus, a large part of healthcare administration can actually be carried out using handheld devices. One of the primary enablers for this is that the latest models of handheld devices are turning out to be as capable as desktops and laptops of not so long ago—virtually any task that one could complete on a desktop or laptop 3-5 years ago can now be accomplished through a handheld device. In fact, almost all the technologies are already available. All that is required is their integration or convergence.

MOBILE TECHNOLOGIES IN HEALTHCARE INFORMATION SYSTEMS

Having discussed how mobile technologies facilitate healthcare administration, which is essentially interactions between a healthcare service provider—a doctor or a clinic—and a patient for providing healthcare services to the patient, let us now see how these technologies can be useful in healthcare information systems.

Healthcare administration results in the generation of a lot of data that needs to be maintained and managed for future reference. Such data includes, but is not limited to, EMRs (electronic medical records), diagnostic reports, and patients' medical history. Healthcare service providers like doctors, clinics, and hospitals also need to maintain a lot of additional information on schedules of doctors (for appointments) and resources (like operation theatre, etc.), information related to finances and inventories, references on the latest developments in the healthcare industry through subscription to various databases, and so on. Such information needs to be readily accessible, anytime, anywhere. Such information also needs to be available for research studies, study of trends and patterns, and so forth.

Healthcare information systems consist of creation, maintenance, provisioning, and retirement of all such information related to healthcare. Typically this would mean maintaining a healthcare information management system (HIMS) in a clinic or hospital. While this is true in today's world, Vanjara (2005) has suggested a geography-time-person agnostic solution. The solution consists of creating networked virtual healthcare communities that have access to relevant information anytime, anywhere.

In both the scenarios, information can be created, maintained, accessed, and retired through handheld devices. Natural interfaces have further removed one of the major constraints in using handheld devices through the use of keyboards or mice. In addition, there are thousands of software applications and utilities, many of them available as freeware on various handheld platforms that enable mobile healthcare information systems.

Healthcare information can be broadly classified into two types:

1. **Medical Information:** This consists of all information pertaining to a patient's health and includes EMRs (electronic medical records), test reports with images as applicable, history of ailments, diagnostics and treatment, patterns and profile of a patient's health, and so on.

2. **Non-Medical Information:** This consists mostly of all commercial information like receivables and payables, accounts, inventory, and so on. In case of hospitals, such information would be much vaster, including data around various administrative and support functions like laundry, kitchen, and so on.

Currently, only mid-sized to large clinics or hospitals maintain comprehensive real-time healthcare information management systems. This scenario has two constraints: (1) the smaller entities do not have any system for HIMS, and therefore data/information related to patients' interactions with them is never maintained; and (2) even the HIMS in mid-sized or large clinics and hospitals is unable to capture and maintain medical information of patients from different sources. Imagine test reports from four different places being compiled in a single central database of a hospital! Sounds quite difficult.

Mobile devices enable HIMS in both cases. In case of smaller entities, they do not have to make large investments in infrastructure to have a comprehensive HIMS. They can actually subscribe to such infrastructure and Web-based applications, and then access them to maintain their information using handheld devices. The HIMS of mid-sized and large clinics and hospitals can be easily fed with all data from multiple sources almost on a real-time basis using handheld devices. Even voice-based interactions/transactions can be captured using the voice recognition feature of mobile devices. The primary idea behind handheld devices is to capture as much data as possible in digital format right from the point of origin. Manipulation and transfer of data after that becomes quite easy and fast.

Though maintaining all types of information—medical or non-medical—is possible using handheld devices, medical information is a primary candidate because it addresses more issues around mobility than non-medical information, which is to a large extent organization or location specific.

ISSUES AND CHALLENGES

Having read and heard so much about the potential and capabilities of mobile/handheld devices, the first obvious question that comes to mind is, why then are they not used so commonly in these areas of healthcare? Growth of mobile technology advancements in the entertainment and consumer sectors has been much higher than in business. Given the fact that it is business that actually contributes to the economy, and in case of healthcare, also to the fitness and well-being of people, this is quite unfortunate.

However, there are some reasons for these as well—in the form of issues and challenges of mobile technologies:

- **Integration of Devices, Applications, and Platforms:** As of now, most of the devices, applications, and platforms work independently. Even mobile phones of two different manufacturers at times cannot communicate with each other. At most, they exchange some data via IR or Internet; however, there is absence of tight integration in terms of features, functionality, and so on. A robust, standard-based integration of multiple devices, applications, and platforms would go a long way in facilitating wider use of mobile devices in the business of healthcare.
- **Investments and Infrastructure:** A lot of investments need to be made in the non-telecom space, including creation of application infrastructure to popularize the use of mobile devices. This would also entail creation of a lot of enterprise-scale application software that can expose selected services to multiple clients. To work at a global scale, at each level or tier the data will have to be transparent to devices and technologies. This again

will require investment in technologies and standards.

- **Common Standards:** Handheld devices, software applications and platforms, and the healthcare industry also need to have common standards of communication amongst themselves, like the MICS—the Medical Implant Communications Service, HL7, and so on. Such standards would enable multiple devices to talk to multiple software/services on different platforms.
- **Privacy and Security:** While handheld device space is not as rampant with these issues today as is the personal computer space, it will not be long before it catches up. Some cases of viruses on mobile phones and PDAs or data pilferage or snooping have already been reported. Privacy and security become of paramount importance in the healthcare space due to two primary reasons: (1) we are constantly dealing with individuals' personal information, and (2) the impact of security breach can cost a human life. Imagine a software virus that keeps on manipulating prescriptions or for that matter diagnostic reports while spreading itself.
- **Laws:** Laws related to the conduct of various industries like healthcare are also widely different in various countries. These too need to be aligned at a broad level for a universal solution to work. While this is very difficult and far fetched, if all the countries in the world can sign charters and conventions on pollution control, IPRs, and many such issues, creation of a common legal framework at a higher level for the benefit of mankind—for instance in the area of healthcare—can certainly be made possible in the long term.
- **Language:** A global solution also has to address the need of multiplicity of languages. This, though, is the least of problems, as quite a few solutions—Microsoft Windows to small accounting packages like QuickBooks and MYOB—have already addressed this issue. Unicode-based solutions can also be considered to address this issue (http://www.unicode.org
- **Need for Social Interaction:** Most of the technological advancements have resulted into lesser human-to-human interactions. This is creating a large gap in the social needs of people. Quite often, at a sub-conscious level, people express resistance to new technologies due to their adverse impact on the fulfillment of their social need.

CONCLUSION

This chapter reviews the healthcare and mobile technologies landscape as it exists today. It then suggests integration of both these technologies—already happening to some extent—for provisioning of superior healthcare services in a flexible manner from anywhere. It discusses how various mobile healthcare gadgets and communication devices can be networked and connected to healthcare administration and information systems. Having proposed a healthcare solution that is a few quantum jumps from where it exists today, the chapter also discusses some issues and challenges in making this happen. It also makes a few suggestions on how some of these issues and challenges can be taken care of.

REFERENCES

Apple Computer, Inc. (2005). *iPod. Which iPod are you?* Retrieved July 27, 2005, from http://www.apple.com/ipod/

Cellular Network. (2005). Retrieved August 3, 2005, from http://www.mywiseowl.com/articles/Cellular_network

Computerworld. (2005). *Combo Wi-Fi cell phone*. Retrieved August 3, 2005, from http://www.computerworld.com/mobiletopics/mobile/story/0,10801,93501,00.html?nas=AM-93501

Dixon, P. (2002). *Everything you wanted to know about the future of health care industry. Keynote speech*. Retrieved March 10, 2005, from http://www.globalchange.com/futuremedicine.htm

EIT.IHK. (2005). *Man machine interface*. Retrieved March 12, 2005, from http://www.eit.ihk-edu.dk/subjects/mmi/index.php?e=0

Free Medical Journals. (n.d.). Retrieved from http://www.freemedicaljournals.com/htm/special.htm

Global Change. (2005). *The future of technology. Recent news of a new kind of world*. Retrieved April 30, 2005, from http://www.globalchange.com/futuretechnology.htm

Google. (2005). *Google directory–health > news and media > publications > journals*. Retrieved July 27, 2005, from http://www.google.com/Top/Health/News_and_Media/Publications/Journals/

Handango. (2005). *Handheld software for healthcare*. Retrieved April 10, 2005, from http://www.handango.com/SoftwareCate gory.jsp?optionId=1_2_2&jid=5232151DE62C1F9 XXX94X1785XD761AX&special=&platformId=2&bySection=1&siteId=1&txtSearch=medical&sectionId=3314&topSectionId=3314&catalog=30&title=Medical

Health-Cares. (2005). *What causes a common cold?* Retrieved July 27, 2005, from http://respiratory-lung.health-cares.net/cold-causes.php

HealthDay. (January 21). *High-tech revolutionizes home health care*. Retrieved March 13, 2005, from http://www.azcentral.com/health/wellness/articles/0121hightechhealth-ON.html

Howard, V. (2004). Nanotechnology and medicine. *Ecologist Asia, 12*(1). Retrieved July 27, 2005, from http://www.sanctuaryasia.com/features/detailfeaturescategory.php?id=666& catid=45

HP iPAQ 6315. (2005). *Product description*. Retrieved July 27, 2005, from http://h10010.www1.hp.com/wwpc/us/en/sm/WF06b/215348-64929-215381-314903-f60-430120-430121-405360.html

Light Blue Optics (2004). About us. Retrieved July 27, 2005, from http://www.lightblue optics.com/index.html

Majeed, M. (2005, March 10). *Recent advances in medicine (speech at Trivandrum Medical College).* Retrieved April 15, 2005, from http://www.drmajeed.com/speeches.htm# Recent

Majumdar, A. (2004). *Ayurveda: The ancient Indian science of healing.* Delhi: Macmillan.

Medgadget. (2005a). *Amplichip.* Retrieved March 12, 2005, from http://www.medgadget. com/archives/2005/01/amplichip_1.html

Medgadget. (2005b). *Fantastic voyage: Smart Pill to expand testing.* Retrieved March 12, 2005, from http://www.medgadget.com/archives/2005/04/fastastic_voyag.html

Microsoft. (2005). *Product description of Windows Mobile and MS Office Mobile.* Retrieved July 27, 2005, from http://www.microsoft.com/windowsmobile/default.mspx and http://office.microsoft.com/en-us/assistance/HA011017991033.aspx

Microsoft Streets & Trips 2005 GPS. (2005). Retrieved April 30, 2005, from http://www.tigerdirect.com/applications/searchtools/item-details.asp?EdpNo= 1016599&Sku=M17-5667&SRCCODE= MSN&CMP=OTC-MSN

PDA Buzz. (n.d.). Retrieved from http://www.pdabuzz.com/

PDA Handyman. (n.d.). Retrieved from http://www.pdahandyman.com/

PDA Live. (n.d.). Retrieved from http://www.pdalive.com/

PMED (2005). *Patient monitors, defibrillators, telemetry, stress test, vascular*. Retrieved March 7, 2005, from http://www.pemed.com/pmonitor/pmon.htm

Ratib, O. (2004). iPod helps radiologists manage medical images. *RSNA News*, (December). Retrieved March 7, 2005, from http://www.rsna.org/publications/rsnanews/dec04/ipod-1.html

Shaheen, M. (2003). *How does WiFi work?* Retrieved August 3, 2005, from http://mason.gmu.edu/~mshaheen/page1.html

Unicode. (2005). *Unicode home page*. Retrieved March 10, 2005, from http://www.unicode.org/

Vanjara, K. (2005). The future of supply chain management: Shifting from logistics driven to a customer driven model. In Y. Lan & B. Unhelkar (Eds.), *Global integrated supply chain systems*. Hershey, PA: Idea Group Publishing (submitted for publication).

WEALTHY. (2005). *Description*. Retrieved March 12, 2005, from http://engadget.com/entry/1234000690040296/#comments

Chapter 3.2
Use of Telemedicine Systems and Devices for Patient Monitoring

Dionisia Damigou
National and Kapodistrian University of Athens, Greece

Fotini Kalogirou
National and Kapodistrian University of Athens, Greece

Georgios Zarras
National and Kapodistrian University of Athens, Greece

ABSTRACT

Today's health standards demand a high quality and efficiency as a major characteristic of every health service provided to the public, even in cases where patients have to be treated from a distance. The combination of medicine and information technology (telecommunications) led to the introduction of the term telemedicine. Telemedicine services are used in assisting remote patients. Interaction and feedback through patient monitoring systems and devices allow the health providers interfere when necessary, so medical maintenance can be guaranteed. This chapter deals with the different kinds of such systems and devices. The contribution of old and new telecommunication technologies is currently being discussed. The individual needs of every remote patient are taken into account, thus, several devices and systems are used for telemonitoring. This chapter indicates characteristics and features of the various kinds of patient monitoring systems and devices.

INTRODUCTION

Due to the technological revolution and the entry of informatics in our everyday lives, the expansion of telecommunications became a reality. Challenging telecommunication applications such as cellular telephony, communications through the intervention of satellites, the Internet, and so forth, support the extensive exchange of information. Several scientific areas have become related to telecommunication applications, including the field of health in which the term telemedicine was first introduced.

Telemedicine involves the transfer of medical information for use in diagnosis, treatment, and education over distances and brings medical services directly to the point of need.

A monitoring system adapts to the needs of each and every patient, care unit, and hospital equipped with hardware and software designed to give maximum functionality, flexibility, and responsiveness. Concerning the monitoring devices, severall types are available, including pulse oximeters, spirometers, glucose monitors, and so forth (Demiris, 2004).

TELEMEDICINE SYSTEMS FOR PATIENT MONITORING

There was a big demand for devices for patient monitoring, especially for people who are far away from health providers. The population is growing old, and the high health standards of nowadays claim that patients should be cared for and rehabilitated at their homes. Several telemonitoring systems were developed. The primary aim of such systems is to maintain the autonomy, independence, and quality of life for the frail elderly, disabled persons, remote patients, and their informal family careers by the application of telematic technology. For example, home health tele-assistance systems could provide a large range of services that would permit the user or patient to remain in his or her normal environment:

- Emergency alarm systems (tele-alarms),
- Post-hospital treatment monitoring,
- Social assistance (24-hour tele-assistance) and so forth (Linkous, 2003).

Remote medical services can also be used in other areas as medical care is provided to prisons, on board commercial aircraft and ships, in the military, and even at the South Pole.

TELEMEDICINE SYSTEMS AND TELEHOMECARE

The most common utilization of telemedicine for patients' monitoring is telehomecare. With more technologies moving into home care, and more and sicker patients being treated outside of the hospital environment, the home-care approach to health care is here to stay. Telehomecare is viewed as a method that uses telecommunication and videoconferencing technologies to enable a health-care provider at the clinical site to communicate with patients at their homes. It is one of the brightest examples of the new frontier of health care. These telehomecare solutions are low cost compared to the classic way of monitoring (Linkous, 2003).

TELEMEDICINE SYSTEMS AND INTERACTION

The telemedicine systems permit monitoring devices to interact with doctors' displaying systems. Such an interaction is called a virtual visit. These devices collect information and signs from the patient, and the resulting data can then be transferred through telephone lines (regular, ISDN, DSL, or T1), the Internet, or the wires of a LAN (local area network) or WAN (wide area network) to the doctor or other health scientists (Kyriacou, Voskarides, Pattichis, Istepanian, Pattichis, & Schizas, 2002). Recent technologies such as Bluetooth might also be used (Roke Manor Research Limited, http://www.roke.co.uk). The diffused data, in many cases, has the potential of updating the doctor's displaying devices in real time. Whenever there is a need (urgent case), the medical personnel might immediately interfere and get in contact with the patient.

There are several types of telemonitoring system applications; their purposes differ in proportion to the concerning situation. Examples of telemonitoring systems are provided in the following sections.

Systems for Monitoring Heart Function and Circulation

a. **Systems for Monitoring Cardiac-Rehabilitation Patients:** The first applications that were evaluated concerned systems that monitor cardiac-rehabilitation patients unable to return to a hospital-based program. Devices for these patients could detect arrhythmia and were found to be more effective than ambulatory electrocardiography. The most recent systems are much more complicated and sophisticated.

b. **Systems for Hypertensive Patients:** Other disease-management applications support hypertensive patients, enabling them to control their blood pressure, and are found to be efficient at the evaluation process.

c. **Systems for Chronic Heart Failure:** Studies for home monitoring of chronic heart failure revealed that there is a need for further investigation (Demiris, 2004).

Systems for Pulmonary Function

a. **Systems for Asthma Patients:** Telemonitoring systems for asthma patients provide assistance in daily routines; the systems alert health-care providers when necessary.

b. **Systems for Lung-Transplant Patients:** Systems for the monitoring of pulmonary function in lung-transplant recipients via the Internet are feasible and accurate.

Systems for the Management of Insulin-Dependent Diabetes

Distributed computer-based systems for the management of insulin-dependent diabetes were developed utilizing Internet technology and monitoring devices to support the normal activities of physicians and diabetic patients by providing a set of automated services enabling data collection, transmission, and analysis, and decision support (Demiris, 2004).

Systems to Assist Post-Transplant Patients

Other Web-based telemonitoring systems were developed to assist post-transplant patients; for example, for lung-transplant recipients, regular spirometry monitoring can be used for the early detection of acute infection and rejection of the allograft. A Web-based telemonitoring system, for these cases, provides direct transmission of home spirometry to the hospital (Demiris, 2004).

Systems for Emergency Response

Telemonitoring systems have also been introduced to provide emergency response to disaster situations. The goal is to offer quality health-care services to persons who are victims of a disaster. Portable telemedicine instrumentation packages can provide a compact, integrated suite of tools such as data-acquisition devices for ear, nose, throat, and skin imaging, lung sound auscultation, and so forth (Demiris, 2004).

Similar system projects were developed in many countries for several purposes.

- **United Kingdom:** In Oxford, a remote physiological-monitoring network was established to evaluate cardiorespiratory function during sleep in a number of infants in their homes for research purposes. The center designed and developed portable monitors for the continuous measurement of vital signs, which allowed downloading data from monitor memory to the hospital. This information was then further analyzed.

- **Israel:** A program was developed to gain access to cardiac patients. In the process of a typical patient call, a health professional collects descriptive information about the patient's condition while simultaneously receiving and recording the 12-lead electrocardiograph (ECG) transmitted by the client-managed portable ECG device. The ECG results are displayed and analyzed, and compared with previous ECGs using proprietary transtelephonic ECG-management software.
- **The Netherlands:** One of the areas that seem to be exceptionally well developed in the Netherlands is the use of handheld computers for electronic case management. After data are collected, they can be immediately transferred, using dial-up networking capacity, from the point of care to the central home-care database. Records are regularly updated and are available to different health professionals within the continuum of care.
- **Germany:** A telemonitoring project was designed in Germany and included extensive home monitoring of the vital parameters of infants at risk for sudden infant death syndrome (SIDS). The project monitored breathing movements, ECG, heart rate, and oxygen saturation. All sensors were noninvasive and integrated with the babies' pajamas. The parents were trained in assessing the physiological status of the children and in emergency intervention measures. Information was also transmitted to the test and research telemedical laboratory for further analysis.
- **A Four-Country Project (Greece, Great Britain, France, Germany):** This project was related to patients with renal failure requiring home hemodialysis. The goal of the project was to develop, apply, and evaluate telematics monitoring and consultation services for enabling the supervision of each hemodialysis session and possible intervention (Demiris, 2004).

TELEMEDICINE DEVICES FOR PATIENT MONITORING

Portable Monitoring Devices

There are several commercially available portable monitoring devices that are approved by the FDA, including pulse oximeters, blood-pressure monitors, weight scales, and glucose monitors. In some cases, data are stored in the device and retrieved at a later point or are displayed on a monitor at the completion of the test session (Demiris, 2004).

Wearable Sensors

There are three kinds of wearable sensors: physical, chemical, and biological. These sensors produce a signal in response to an event, which is then transferred to a circuit and becomes digitised. A physical sensor measures physical parameters such as temperature or pressure, whereas a biological or chemical sensor involves a receptor that binds with an analyte. The resulting digital data can be stored and/or displayed.

The concept of wearable sensors is based on the incorporation of sensors into watches, items of clothing, and eye glasses. Thus, one could argue that wearable sensors can function as noninvasive, in vitro diagnostic tools as they are capable of analyzing, among others, human sweat, tears, stress, strain, and pH increases.

There exists a knee-wearable sensor that is also known as the intelligent knee sleeve, and it was first designed for football players. It monitors knee strain or injury. Originally, this device was strapped to the knees, and its sleeve provided feedback to users by emitting an audio tone. It can be a useful application for home-care patients with mobility impairments or for those at the rehabilitation phase.

Another example of a wearable sensor is a small, portable detector in the form of a wristwatch that provides test results for cystic fibrosis in minutes, rather than the 24 hours that is the typical response time for a laboratory. Another wristwatch device uses an electric field to push pilocarpine nitrate into the skin, thereby dilating the pores. Sweat is absorbed and stored in a duct in the watch. The sample is analyzed by a sensor, and the levels of sodium, chloride, and potassium ions are recorded. Other devices in the form of wristwatches include glucose meters that measure glucose in the interstitial fluid as a low electric current pulls glucose through the skin, and a blood oxygen monitor.

In the last few years, the Smart Shirt was introduced that incorporates technology into the design of clothing to monitor the wearer's heart rate, ECG, respiration, temperature, and vital functions, alerting the wearer or physician if there is a problem. The Smart Shirt also can be used to monitor the vital signs of military personnel, chronically ill patients, firemen, and frail elderly persons living alone (Demiris, 2004).

Also, for military purposes, the U.S. military is developing innovative applications for advanced sensors and smart materials. Devices resembling wristwatches will be worn by all soldiers as part of the combat uniform. These devices will monitor the soldier's vital signs continuously by monitoring parameters such as noninvasive blood pressure, pulse oximetry, and medical imaging (Garshnek & Burkle, 1999).

Robots

It is impressive that robots have started claiming doctors' duties. At present, these robots are navigated by an operator using a joystick at a control station and can perform rounds within the hospital. A camera and microphone are mounted above the computer screen, allowing the operator to see and hear. A similar camera and microphone at the control station transmit the operator's face and voice. It is surprising that a number of doctors perform rounds on their patients with these robots, particularly from their homes. The full potential of robots is beginning to be explored. Future generations of control stations are expected to be portable and operational from virtually anywhere in the world (Norris, 2004).

ADVANTAGES OF TELEMEDICINE

The categories of telemedicine are the following:

- Teleconsultation
- Telediagnosis
- Telecare
- Remote clinical sessions and tele-education
- Remote data access
- Teleradiology
- Home care
- Telemonitoring (Mantas & Hasman, 2002)

Telemedicine's objective is to provide users with an integrated health-information service through an expert system, which gives access to existing information related to health, social care, and other general issues that is now distributed in dispersed databases. After achieving this objective, we expect the following advantages.

1. **Advantages in Monitoring the Patient:** The use of telemedicine is followed by an improvement of the quality of the monitoring of patients and an increase in the number of

patients being monitored. Moreover, tools are provided for medical decision support and guided monitoring work is allowed.

2. **Advantages in the Management of Treatment and Training:** Telemedicine improves the communication between doctor and patient, promotes patients' self-management and training, and decreases the response time in the treatment.
3. **Advantages in Telecare:** Telecare reduces the number of visits to hospitals, facilitates patients in finding information regarding their illnesses, allows the patients autonomy, and decreases short-term as well as long-term complications (Mantas, Aguilera, del Pozo Guerrero, Arredondo Waldmeyer, & Martínez Fernández, 2000).

In general, telemedicine offers the following:

- Reduction of health-care costs
- Access to health services in previously unserved or underserved areas
- Easy cooperation between health professionals
- Improved quality of care

CONCLUSION

The development of information-systems technology has led us to an increased number of telemonitoring applications to help with patient health care. These applications enable health professionals to carry out home health visits (virtual visits). Such applications and systems complete the management of data collection and reinforce data analysis. Telemedicine services can be shared among patients and several regional hospitals and other specialized health centers. Thus, patients are allowed to ask for advice, and, as a response, health professionals may interfere when necessary. It appears that the use of telemonitoring systems can cut the cost of medical care for rural and urban areas.

Interesting research and implementations have been developed to evaluate systems delivering assistance to different scientific health fields, contributing to cardiology, neurology, surgery, orthopedics, pediatrics, and so forth. A lot of them were found to be accurate and efficient enough, although the future promises more potent and sophisticated systems for patient monitoring.

However, there are still barriers to the wider adoption of telemedicine, affecting both healthcare professionals and their patients. The general public is not mature enough to get involved in such procedures due to the fact that it is not well acquainted with the subject. In the future, a lot of these difficulties will be overcome, and we expect that the public will accept the challenges of telemedicine and telemonitoring.

REFERENCES

Demiris, G. (2004). *Electronic home healthcare: Concepts and challenges.* Columbia: University of Missouri.

Gantenbein, R. (1992). Telehealth technology: Wyoming's efforts to facilitate research and service in a rural state. A*merican Stroke Association Conference,* Ft. Lauderdale, FL. Center for Rural Health Research and Education.

Garshnek, V., & Burkle, F. (1999). Applications of telemedicine and telecommunications to disaster medicine. *Journal of American Medical Informatics Association, 6*, 26-37.

Kyriacou, E., Voskarides, S., Pattichis, C., Istepanian, R., Pattichis, M., & Schizas, C. (2002). Wireless telemedicine systems: A brief overview. *IEEE Antennas & Propagation Magazine, 44*(2), 143-153.

Linkous, J. (2003). *Advances in telemedicine technology*. Helsinki, Finland: American Telemedicine Association (ATA).

Mantas, J., Aguilera, E. G., del Pozo Guerrero, F., Arredondo Waldmeyer, M. T., & Martínez Fernández, A. (2000). *Health informatics. Textbook in health informatics.* Athens: IOS Press.

Mantas, J., & Hasman, A. (2002). *Textbook in health informatics.* Athens: IOS Press.

Norris, K. (2004). *Medicine and technology: Robot makes the rounds.* Detroit, MI: Detroit Free Press.

KEY TERMS

Home-Care Technology: Any technology used to implement home-care telemedicine services. It can also incorporate any device or instrument for patient monitoring, therapy, or environmental control (Mantas & Hasman, 2002).

Telecare: The use of telecommunication systems to provide remote assistance in therapy to patients (Mantas & Hasman, 2002).

Teleconsultation: Remote access to a specialist's knowledge. This type of service is seen as a particular case of cooperative work or cooperative diagnosis (Mantas & Hasman, 2002).

Telediagnosis: Diagnosis of a patient by a remote physician. This kind of service does not operate directly between the patient and remote doctor (Mantas & Hasman, 2002).

Telehomecare: Uses telecommunication and videoconferencing technologies to enable a healthcare provider at the clinical site to communicate with patients at their homes (Demiris, 2004).

Telemedicine: The use of electronic information and communication technologies to provide and support health care when distance separates the participant. It emphasizes applications that link clinician to patient or one clinician to another (Gantenbein, 1992).

Telemonitoring: Remote monitoring of patients' physiological value. This kind of service is used with chronic and/or high-risk patients (Mantas & Hasman, 2002).

This work was previously published in Handbook of Research on Informatics in Healthcare and Biomedicine, edited by A. A. Lazakidou, pp. 221-227, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.3
Mobility in Healthcare for Remote Intensive Care Unit Clinical Management

Carolyn McGregor
University of Western Sydney, Australia

ABSTRACT

This chapter reviews current research directions in healthcare mobility and assesses its impact on the provision of remote intensive care unit (ICU) clinical management. Intensive care units boast a range of state of the art medical monitoring devices to monitor a patient's physiological parameters. They also have devices such as ventilators to offer mechanical life support. Computing and IT support within ICUs has focused on monitoring the patients and delivering corresponding alarms to care providers. However many intensive care unit admissions are via intra and inter health care facility transfer, requiring receiving care providers to have access to patient information prior to the patient's arrival. This indicates that opportunities exist for mobile gadgets, such as personal digital assistants (PDAs) to substantially increase the efficiency and effectiveness of processes surrounding healthcare in the ICUs. The challenge is to transcend the use of these mobile devices beyond the current usage for personal information management and static medical applications; also to overcome the challenges of screen size and memory limitations. Finally, the deployment of mobile-enabled solutions within the healthcare domain is hindered by privacy, cost and security considerations and a lack of standards. These are some of the significant topics discussed in this chapter.

INTRODUCTION

Intensive care units (ICUs) worldwide offer support for patients in need of critical care. The research, development, and adoption of new information technologies (ITs) and information systems (ISs) within ICUs, and particularly neonatal intensive care units (NICUs) to support patient and care provider mobility, is currently lagging behind other industries and other areas of healthcare (McGregor, Heath, & Wei, 2005a; McGregor, Kneale, & Tracy, 2005b; Wu, Wang, & Lin, 2005). In order to understand and improve upon this lag, we need to understand the current healthcare scenario within the context of intensive care.

To start with, mobile clinical management solutions within the context of intensive care units need to consider not only the mobility of the patient, but equally importantly, the mobility of the care provider. When an incident requiring critical care occurs, patients may already be located in the care provider's ICU. However, the patient may also be located elsewhere in the care provider's hospital, in another hospital, in their home, or in another location outside the hospital of the care provider. Patients may also be in transit between any of these locations via an ambulance, helicopter, or inter-hospital transport.

Care providers can be understood in this chapter as any physician, clinician, or nursing specialist responsible for some aspect of the clinical management of the ICU patient. In daily routines, physicians, clinicians, nurses, and other staff of the hospital have to be reached and updated of new incidents and information while they are commuting in their work environments (Kafeza, Chiu, Cheung, & Kafeza, 2004). However, similar to the patients, care providers may also be located within the ICU, their office, elsewhere within their hospital, in their home, or in another location outside their hospital (e.g., attending an off-site meeting or conference).

Ammenwerth, Buchauer, Bludau, and Haux (2000) report that one of the major clinical management issues that mobile technologies can help with within the hospital is communication and reachability of care providers. This clinical management issue has the additional challenge of determining to whom the message should be sent (Kafeza et al., 2004). Both of these issues are particularly relevant within the ICU setting.

When critical care clinical management is required, the sooner the patient/care provider(s) information exchange can commence, the faster the clinical management can commence. In the case where the patient and care provider(s) are not located together within the ICU, critical care can still commence, provided there is adequate clinical management support to facilitate clinical decision making and execution.

Mobile healthcare systems (MHSs) have been defined by Wu et al. (2005) as the use of IS/IT to exchange healthcare information and services via mobile devices anytime and anywhere, providing patients and care providers with easy access to resources whether stationary or moving.

Recent research directions for computing and IT support within ICUs has focused on the delivery of alarms/alerts to care providers (Catley & Frize, 2003; Catley, Frize, Walker, & St. Germain, 2003; Shabot, LoBue, & Chen, 2000; Sukuvaara, Makivirta, Kari, & Koski, 1989; van der Kouwe & Burgess, 2003). However these approaches do not enable mobility in healthcare and neither do they exploit the substantial benefits possible by proper application of mobility. Furthermore, many intensive care unit admissions are via intra- and inter-healthcare facility transfer, requiring receiving care providers to have access to patient information, prior to the patient's arrival and often while the care provider is also in transit. These are some interesting challenges in terms of communication and reachability of care providers.

Recent surveys show that between 25-35% of physicians, as distinct from care providers in general, use personal digital assistants (PDAs) (Carroll & Christakis, 2004; Fontelo, Kim, & Locatis, 2003). However, Carroll et al. (2003) further note that these PDAs are mainly for personal information management and static medical applications. Opportunities exist for PDAs and similar handheld devices to enhance and effectively deliver services within the ICU clinical management sector. However, PDA screen size and memory are seen as crucial factors in the development of PDA applications. In addition, deployment of mobile-enabled solutions within the healthcare domain is impacted by privacy, confidentiality, cost, and security considerations in addition to a current lack of standards.

This chapter reviews current research directions in healthcare mobility and assesses its impact on the provision of remote intensive care unit clinical management. A background to intensive care unit clinical management is first introduced. Recent computing and IT-related research to support ICUs is then presented. Hardware and associated research to support ICU clinical management mobility is then described. A comparison of recent ICU research within the context of its ability to support mobility is presented. Issues impacting the implementation of mobile ICU clinical management solutions are then detailed. Finally, the conclusion and future directions are presented.

BACKGROUND TO ICU CLINICAL MANAGEMENT

Clinical management systems are designed to assist care providers in diagnosis and treatment using existing, already established methods of diagnosis and accepted treatments (Gross-Portney & Watkins, 2000). Hence, mobility in clinical management must support mobility in relation to diagnosis and treatment. Tasks such as medication monitoring, emergency hospitalization of patients, laboratory examination results, ordering and shipment of drugs, and exchange of information relating the patient clinical management occur frequently (Kafeza et al., 2004). Within the ICU context, clinical management systems must respond actively and very timely to the patient's needs, which can be life critical.

One of the most prominent objectives within the modern hospital is the need for accurate, safe, and continuous communications among departments and highly specialized medical staff. In addition, the need for flexible communications to enable communication with other hospitals is also dominant. As a result there has been a great demand among the care providers for a mobile alert management system that is robust, efficient, cost effective, simple, and user friendly (Kafeza et al., 2004).

Intensive care units (ICUs) worldwide offer support for patients in need of critical care. They boast a range of state-of-the-art medical monitoring devices to monitor a patient's physiological parameters such as blood oxygen, blood pressure, and heart rate. Other devices such as ventilators offer mechanical life support.

Broadly, there are three types of intensive care, namely, adult, pediatric, and neonatal. While the age of the patient is the differentiator between adult and pediatric ICUs, the clinical management differs greatly from these ICUs to the neonatal ICUs (NICUs)—where gestational age greatly impacts clinical management.

Approximately 18% of babies born in New South Wales (NSW), Australia, require special care or neonatal intensive care admission (NSW Health Department, 1994). Premature babies can be up to 17 weeks early and may only weigh 450 grams; they can spend three or four months in intensive care and have dozens of specific diseases before discharge. In addition, 15% of neonatal intensive care admissions are transferred after delivery from smaller remote hospitals without intensive care facilities. Similar conditions apply elsewhere within Australia and internationally, where small remote hospitals are spread throughout a given country supported by centrally located referral hospitals with NICUs.

Remote hospitals have equipment to provide limited NICU support within 'special care nurseries'; but without the ability for a neonatologist to receive information from this equipment, the baby must be moved to a referral hospital with neonatologist support. Given the critical requirement to maintain a consistent environment, moving a baby at this time can be life threatening. Critically ill, term and pre-term babies that have to be transferred have higher mortality rates and much higher rates of long-term disability than similar babies born in hospitals with intensive care facilities (McGregor, Bryan, Curry, & Tracy, 2002). A major limitation is that the attending care provider at the remote hospital must contact a neonatal specialist (neonatologist) via telephone, or in some instances the provider—who may or may not be located at the NICU at that time—must describe via e-mail (Deodhar, 2002) the baby's symptoms and, where possible, relay any physiological information verbally, or narratively in the context of an e-mail. The consulting neonatologist must then make decisions based on this verbal or textual exchange.

It is very common for critically ill babies to have significantly abnormal variation in the measured parameters minute by minute, and not all these variations are made available to the consulting neonatologist. Frequent transient falls in blood pressure and blood oxygen content, often with swings into the high range, may be of critical importance in survival and quality of survival, free of significant disability (Lister, Bryan, & Tracy, 2000).

Hence the neonatologists located at referral hospitals require the ability to obtain information from the monitors attached to babies. Similarly, a neonatologist need not be located at a PC within the hospital to view patient data, but should be free to view this information through any device offering a secure Internet/intranet connection. These scenarios open up opportunities for application of mobility in ICU management, as discussed in this chapter.

RECENT COMPUTING AND IT RESEARCH TO SUPPORT ICU CLINICAL MANAGEMENT

Much of the recent computing and IT related research to support intensive care units (ICUs) has focused on clinical alerts (Catley & Frize, 2003; Catley et al., 2003; Shabot et al., 2000; Sukuvaara et al., 1989; van der Kouwe & Burgess, 2003). The information made available to these systems is limited to a small set of physiological data and/or clinical data from patients located within their ICUs. In addition, care provider access to these systems is limited to the receipt of alerts, with minimal content via e-mail and in some cases pagers.

An integrated XML-based healthcare framework for NICU clinical alerts is described by Catley and Frize (2003) and Catley et al. (2003). The alerting is based on the individual device alarms, and predicts mortality and ventilation requirement probability and estimated length of stay. In the implementation as presented, only e-mail alerts and Java GUI alerts are generated. They indicate that future directions for this research are a WML-based alert using Java Servlet technology running on a Web server connected to a WAP Gateway. The WAP Gateway would transmit alerts via WML to mobile devices.

Shabot et al. (2000) describe a software system which extracts clinical information from clinical information systems on a continuous basis and sends it through event detection algorithms. Alerts for detected events are forwarded through a commercial paging system to designated care providers and pharmacists.

InCare was a rule-based alarming system prototype to support clinical management within ICUs (Sukuvaara et al., 1989) that detected four important patient pathological conditions, which develop gradually during postoperative recovery of cardiac patients, namely: (1) hyperdynamic state, (2) hypoventilation, (3) hypovolemia and left ventricular failure. User interface to the alarms was via a PC.

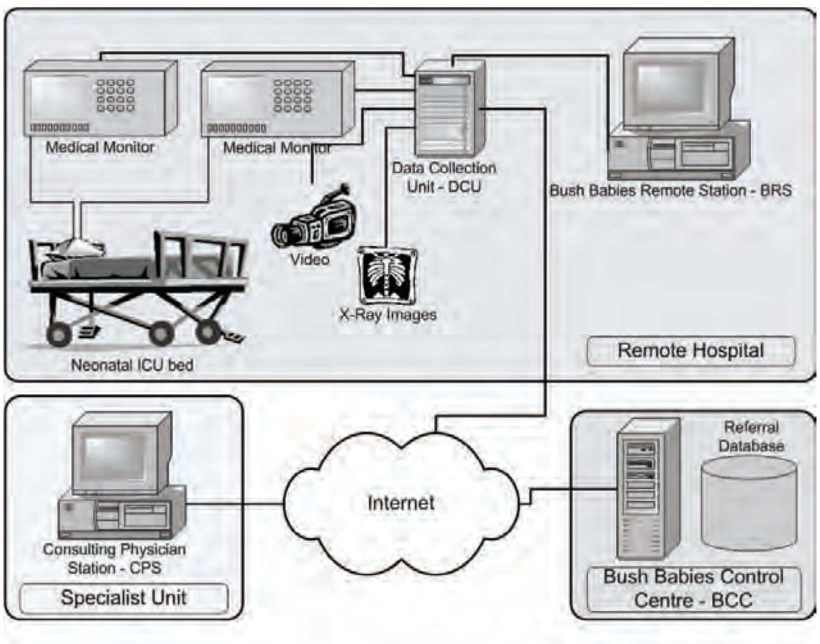
Van der Kouwe and Burgess (2003) present an architecture for continuous electrophysiological monitoring within a neurointensive care unit. While this information is available in real time through Web-based interface, access is limited to PCs via the hospital's secure intranet.

Shin, Huh, Lee, and Kim (2003) have developed Web-based real time for checking temperature and humidity within infant incubators in NICUs. However the access to this information was limited to the NICU central monitoring station PC, with the information delivered via the hospital intranet.

The Bush Babies Broadband research project (McGregor, Kneale, & Tracy, 2005b), supported by the Telstra Broadband Fund, aims to significantly improve the quality of treatment for babies born in regional and remote areas by providing the first on-demand virtual neonatal intensive care unit architecture in Australia. The Bush Babies architecture is shown in Figure 1. Real-time data collected from medical monitors and ventilators attached to the baby, audiovisual streams, and static physiological data such as x-ray images are transmitted from the data collection unit (DCU) to the consulting neonatologist to gain a better picture of the patient's condition than is currently available. The Bush Babies Remote Station enables the remote care provider to initiate a bush babies session. The BRS uses a centralised Bush Babies Control Centre database to select a neonatologist from a NICU where space for the baby would be available if transport was required. A limitation of that research is that the only consulting physician station (CPS) device that neonatologists can use to view this patient condition information is via the screen of a PC or laptop. Given that consulting neonatologists are not always located in their offices within their NICU when there services are required, alternate and more portable technologies to deliver the information need to be investigated.

In association with the Bush Babies project, McGregor et al. (2005a) and McGregor, Purdy, and Kneale (2005c) propose a Web service-based framework for the transmission of XML-encoded physiological data output by medical monitoring and life support devices. That research, together with the previously mentioned Bush Babies, are portions of the "e-Baby" research collaboration (McGregor et al., 2002, 2005a, 2005b, 2005c) that

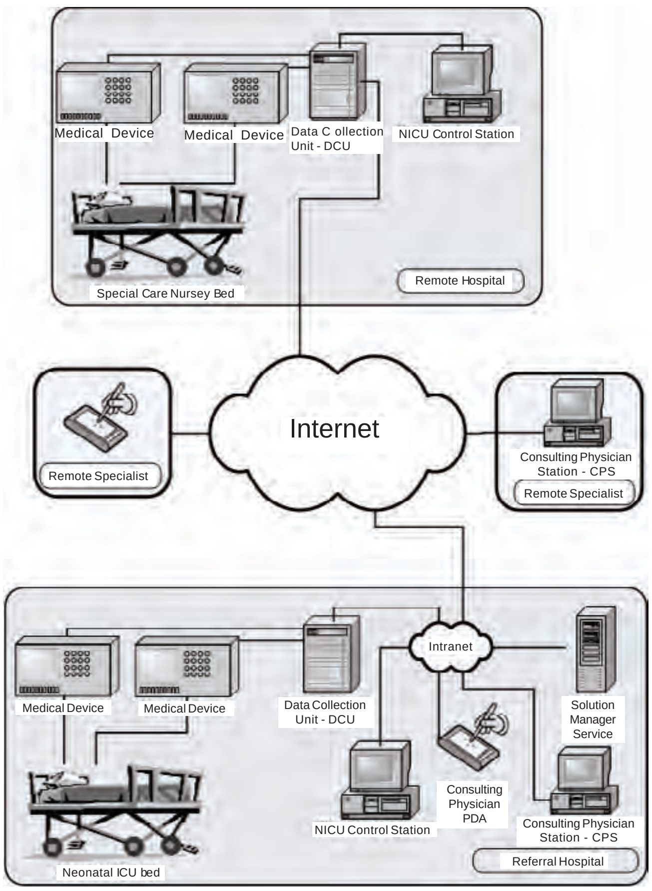
Figure 1. Bush Babies architecture (McGregor et al., 2004)



is researching new approaches to the application of computing and information technology to support mobility in healthcare through local and remote neonatal intensive care. The high level e-Baby architecture is shown in Figure 2.

The Web service-based framework for the transmission of XML-encoded physiological data forms part of the solution manager service (SMS). The SMS is situated in the referral hospital NICU and receives and stores data collected by the data collection units via the *physiological log Web service* for near-real-time analysis and trend detection. The consulting physician station is used by the neonatologists to access the physiological data located in SMS via a set of *analyse Web services* or via a direct link to the data as it is streamed through the *physiological log Web service.* That research enables patient and care provider mobility, as a result of the Web service-based data transmission. However, the prototype as presented only tests transmission of data from patients located either within the NICUs or remote hospital and care providers located within the hospital with the NICU. In addition, care provider access is only available through the PC-based CPS, though future research indicates the forwarding of data to consulting physician PDAs.

Figure 2. e-Baby architecture (McGregor et al., 2005a)



HARDWARE TO SUPPORT ICU CLINICAL MANAGEMENT MOBILITY

The variety of mobile devices available for use to support clinical management within ICUs include: PDAs, laptops, notebooks, GPSs, and smartphones. Recent surveys show that approximately 25% or more of physicians use PDAs, although mainly for personal information management and static medical applications. A PDA provides many advantages. It starts quickly at the push of a button. It is convenient to carry around, fitting easily into a shirt pocket or handbag. Some devices can function for weeks of regular use after a quick battery recharge. However, PDA screen size and memory are seen as crucial factors in the development of PDA applications (Fontelo et al., 2003).

Carroll and Christakis (2004) recently surveyed pediatricians in relation to their use of PDAs and found that 35% currently use PDAs at work; the most common uses were for drug reference (80%), personal scheduling (67%), and medical calculations (61%).

A plethora of software packages are available for PDAs to support clinical management in the broad areas of medical handheld software collections, medical publishers, university medical handheld resources, document readers, access to medical literature, pharmacopoeias, specialty specific, and patient tracking. Patient tracker (www.handheldmed.com) gives the user the option to enter patient records, including demographics, laboratory results, medication,/allergy lists, test results, and radiology reports. Wardwatch (www.torlesse.com) was designed to aid medical staff in ward rounds. Medical Pocket Chart (www.gemedicalsystems.com) is an electronic medical record keeper (Fischer, Mehta, Wax, & Lapinsky, 2003).

Providing care providers with the ability to perform research at the bedside via PDA access to PubMed and clinical trial Internet sites was described by Fontelo et al. (2003).

Carroll (2001, 2002) and Carroll, Tarczy-Hornoch, O'Reilly, and Christakis (2004) describe the implementation of a PDA-based patient record and charting system for an NICU; however, the charting component relies on the care providers completing patient flowsheets manually into the PDA, rather then having the physiological data streaming to the PDA.

A PDA-based approach for managing patient data is defined by Lapinsky et al. (2001). Patient data was entered into the Memopad using a customized template. This data was transferred between care providers using the PDAs' infrared ability. Daily paper notes were generated via infrared link to a HP LaserJet printer, as the hospital policy required paper records.

NICU Notes (Schulman, 2003) enables care providers to collect data at the point of care and utilizes synchronization to move the patient data from the PDA via an ODBC DSN to a secure Microsoft Access application.

A critical PDA issue as defined by Carroll (2002) is the asynchronous nature of hot syncing. Information on the PC and PDA only match immediately after hot syncing. The sheer volume of data being passed during the hot sync process caused the hot sync process to fail intermittently, resulting in incomplete or duplicated information.

COMPARISON OF RECENT RESEARCH

Having discussed the routine usage of mobility in healthcare thus far, we now focus our attention on comparing the previously presented computing and IT research to support ICU clinical management—with an aim to assess the ability of that research to support mobility.

The comparison is broadly categorized into the areas of broad functionality, patient and care provider mobility, and finally architecture mobility. Within the context of broad functionality, three areas were considered: the clinical management

Table 1. Broad functionality

Research Identifier	**Clinical Management Function(s) Supported**	**Decision Time**	**Decision Quality**
(Catley & Frize, 2003; Catley et al., 2003)	Clinical alerts predicting mortality, ventilation requirements, and length of stay	Non-critical	Not available
(Shabot et al., 2000)	Clinical event detection, laboratory results, and medication alerts	Non-critical and life threatening	High
(Sukuvaara et al., 1989)	Alarms relating to postoperative recovery of cardiac patients	Life threatening	High
(van der Kouwe & Burgess, 2003)	Samples electro physiological data	Life threatening	High
(Shin et al., 2003)	Infant humidicrib temperature and humidity monitoring	Life threatening	High
(McGregor et al., 2005b)	Video, image, and physiological data stream monitoring	Life threatening	High
(McGregor et al., 2005a, 2005c)	Physiological data stream monitoring	Life threatening	High
(Carroll, 2002; Carroll et al., 2004)	Patient record charting system	Non-critical	Medium
(Lapinsky et al., 2001)	Managing patient data	Non-critical	Medium
(Schulman, 2003)	Managing patient data	Non-critical	Medium

function(s) that was (were) being supported, the time sensitivity of the decision, and the decision quality sensitivity. Time sensitivity was considered important, as Panniers (1999) has stated that only low and medium urgency decisions are suitable for computerization into a decision support system to support clinical management. The comparison based on broad functionality is presented in Table 1.

Secondly, the degree of patient and care provider mobility was assessed, together with the extent of information that was available to the care provider and is summarized in Table 2.

Finally, the extent to which the proposed architectures incorporate mobility is summarized in Table 3.

Contrary to Panniers' (1999) time sensitivity observations, the research indicates that clinical management systems are being developed to support life-threatening conditions by alerting care providers quickly of the development of the situation.

When data is being forwarded directly in a time series stream from medical devices, data quality

Table 2. Degree of patient and care provider mobility

Research Identifier	Extent of Patient Mobility	Extent of Care Provider Mobility	Extent of Information Available to (from) Care Provider
(Catley & Frize, 2003; Catley et al., 2003)	Located within NICU	Accessing PC within hospital (wireless access proposed for future research)	Text-based notification that physiological data values have exceeded threshold
(Shabot et al., 2000)	Located within ICU	Alerts sent to pager	Limited text-based alerts, based on type of alert
(Sukuvaara et al., 1989)	Located within ICU	Accessing PC within ICU	Physiological data streams
(van der Kouwe & Burgess, 2003)	Located within ICU	Accessing PC within hospital	Physiological data streams
(Shin et al., 2003)	Located within NICU	PC located at NICU control station	Data stream
(McGregor et al., 2005b)	Located within local NICU or remote hospital	Accessing PC within hospital	Video, image, and physiological data stream
(McGregor et al., 2005a, 2005c)	Located within local NICU or remote hospital	Accessing PC within hospital	Physiological data stream
(Carroll, 2002; Carroll et al., 2004)	Located within NICU	Located within NICU	(Clinical charts updated by care provider)
(Lapinsky et al., 2001)	Located within ICU	Located within ICU	(Clinical charts updated by care provider)
(Schulman, 2003)	Located within NICU	Located within NICU	(Clinical charts updated by care provider)

Table 3. Extent to which architectures incorporate mobility

Research Identifier	User Interface Device(s)	Networks	S oftware/Middleware
(Catley & Frize, 2003; Catley et al., 2003)	Web-based	Wired hospital intranet	XML-based messaging
(Shabot et al., 2000)	Pager	PageNet network	Pager messaging
(Sukuvaara et al., 1989)	PC GUI	Wired hospital intranet	Not available
(van der Kouwe & Burgess, 2003)	Web-based	Wired hospital secure intranet	Not available
(Shin et al., 2003)	Web-based	Wired hospital secure intranet	HTML document
(McGregor et al., 2005b)	Web based	Wired hospital secure intranet	Simple Medical Data Protocol (SMDP) document
(McGregor et al., 2005a, 2005c)	Web-based	Wired hospital secure intranet	XML document
(Carroll, 2002; Carroll et al., 2004)	PDA	Hotsync to PC	PDA HotSync
(Lapinsky et al., 2001)	PDA	Between PDAs and printer via infrared	PDA infrared
(Schulman, 2003)	PDA	Hotsync	PDA Hotsync

is high. However, when PDAs are used to collect data from care providers, errors still occur.

None of the research reviewed catered to patients located elsewhere within the care provider's hospital, and the only out-of-hospital location that was supported was another special care nursery within McGregor et al.'s (2005b) research.

Only Shabot et al. (2000) enabled care provider communication from outside the hospital, and this was via a pager. Hence, information available to care providers outside the hospital was limited.

While most user interfaces for the delivery of information were Web enabled, the task of delivery to devices other than PCs and notebooks has not been adequately addressed.

Standards for user interfaces and communications have not been developed as part of these research efforts, nor does the research contain references to other standards efforts. This indicates that such standards do not exist within the context of ICU clinical management.

ISSUES IMPACTING IMPLEMENTATION

In addition to considering PDA usage, there are several factors that to date are still issues impacting successful research, development, and implementation of mobile clinical management solutions within the ICU setting. These factors are indeed common to all clinical management solutions and include wireless network interference in ICUs, security, privacy, confidentiality, cost, and a lack of communications standards.

While the adoption of 802.11b wireless networks is increasing, Fontelo et al. (2003) state that security issues may preclude current deployment of wireless devices for medical data access and utilize infrared access stations within their research. This link was restricted to a 15-degree arc on each side of the centre and a maximum distance of 8 feet. They also found that a transient, rapid disruption of the infrared (IR) beam, such as a person walking between the IR point and the PDA, did not disconnect an established link. In addition, the use of devices via wireless networks within ICUs currently interferes with many of the devices used within ICUs for critical care.

Fischer et al. (2003) state that patient confidentiality and costs are the main implementation issues for handheld systems used within clinical management and that consumer interest may be the limiting factor to successful implementation. In addition, the lack of standards and limited bandwidth for data transfer may also impact increased implementation.

However, they conclude that *a growing body of literature supports the use of handheld devices in a variety of medical settings and with the rapid advances in this technology, the mobile computer may well become an essential medical tool.*

While the deployment of mobile clinical management solutions within ICUs offers the potential of improved patient care and service quality and increasing care efficiency, most applications within the broad healthcare context have failed or not been implemented as predicted, with 30% of failures attributed to non-technical factors (Wu et al., 2005).

For mobile IS/IT solutions to significantly impact ICU clinical management and result in a paradigm shift in the approaches to healthcare within this context, several standards need to be developed to support the ICU clinical management functions. These functions include the exchange of information relating to the patient clinical management, medication monitoring, ordering and shipment of drugs, and the examination of laboratory results.

While all monitoring and life support devices used within ICUs to supply physiological data have the ability to output the device readings usually via a serial port, the data formats vary greatly from device to device. As such, efforts to make device data from devices attached to local or remote patients accessible for mobile viewing are hampered by the myriad of formats required for transmission. McGregor et al. (2005a, 2005c) propose a Web service-based framework for the transmission of physiological data output by such devices, proposing an XML format for such data transmission; however, it has not to date resulted in mainstream standards adoption for the transmission of physiological data.

To enable ICU clinical management mobility, standards for physiological data transmission, medication monitoring, laboratory examination results, and the ordering and shipment of drugs are still required.

CONCLUSION AND FUTURE DIRECTION

While the paradigm shift to mobile clinical management for ICUs offers the potential to significantly improve the speed, efficiency, and effectiveness of critical care within ICUs, there are currently several factors impacting its mainstream adoption.

This chapter has presented a review of current healthcare mobility within the context of ICU clinical management. A comparison of recent ICU computing and IT-related research indicates that the issue of both patient and care provider mobility has not been considered a priority within these research efforts.

Several issues continue to impact successful implementation of mobile clinical management solutions for ICUs; these include issues relating to the use of wireless networks, in addition to security, privacy, cost, and a lack of communications for data exchange and user interface standards.

Current computing and IT-related research to design and develop the next-generation IT/IS solutions for ICUs is not adequately incorporating the issue of mobility of patient and care provider.

Traditionally, clinical research is used as the catalyst for providing evidenced-based recommendations for change to clinical management practices. To gain care provider acceptance, changes to clinical management practices proposed through the introduction of computing and IT approaches should consider traditional clinical research approaches to validate findings and gain support and acceptance.

ACKNOWLEDGMENT

The work presented in this chapter is associated with a larger, long-term research effort known as "e-Baby" (McGregor et al., 2002, 2005a, 2005b, 2005c) that is researching new approaches to the application of computing and information technology to support mobility in healthcare through local and remote neonatal intensive care. The researchers wish to thank the staff of the Neonatal Intensive Care Unit, Nepean Hospital, Sydney West Area Health Services, NSW, for their continued support and collaboration in the e-Baby research.

REFERENCES

Ammenworth, E., Buchauer, A., Bludau, B., & Haux, R. (2000). Mobile information and communication tools in the hospital. *International Journal of Medical Informatics, 57*(1), 21-40.

Carroll, A. E., Tarczy-Hornoch, P., O'Reilly, E., & Christakis, D. A. (2004). The effect of point-of-care personal digital assistant use on resident documentation discrepancies. *Pediatrics, 113*(3), 450-454.

Carroll, A. E., & Christakis, D. A. (2004). Pediatricians' use of and attitudes about personal digital assistants. *Pediatrics, 113*(2), 238-242.

Carroll, S. S., & Tarczy-Hornoch, P. (2001, November 6). Development of a personal digital assistant (PDA) based client/server NICU patient data and charting system. *Proceedings of the AMIA Symposium*, Washington, DC (pp. 100-104).

Carroll, S. S., & Tarczy-Hornoch, P. (2002, November 12). The implementation of a personal digital assistant (PDA) based patient record and charting system: Lessons learned. *Proceedings of the AMIA Symposium*, San Antonio, TX (pp. 111-115).

Catley, C., & Frize, M. (2003, April 9-12). A prototype XML-based implementation of an integrated 'intelligent' neonatal intensive care unit. *Proceedings of the 4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine*, Birmingham, UK (pp. 322-325).

Catley, C., Frize, M., Walker, C. R., & St. Germain, L. (2003, September 17-21). Integrating clinical alerts into an XML-based health care framework for the neonatal intensive care unit. *Proceedings of the 25th Annual IEEE International Conference of the Engineering in Medicine and Biology Society*, Cancun, Mexico (Vol. 2, pp. 1276-1279).

Deodhar, J. (2002). Telemedicine by e-mail: Experience in neonatal care at a primary care facility in rural India. *Journal of Telemedicine and Telecare*, *8*(2), 20-21.

Fischer, S. T., Mehta, S., Wax, R., & Lapinsky, S. E. (2003). Handheld computing in medicine. *JAMIA*, *10*(2), 139-149.

Fontelo, A. M., Kim, G., & Locatis, C. (2003). The PDA as a portal to knowledge sources in a wireless setting. *Telemedicine Journal and E-Health: The Official Journal of the American Telemedicine Association*, *9*(2), 141-147.

Gross-Portney, L., & Watkins, M. (2000). *Foundations of clinical research, applications to practice* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Kafeza, E., Chiu, D. K. W., Cheung, S. C., & Kafeza, M. (2004). Alerts in mobile healthcare applications: Requirements and pilot study. *IEEE Transactions on Information Technology in Biomedicine*, *8*(2), 173-181.

Lapinsky, S. E., Weshler, J., Mehta, S., Varkul, M., Hallett, D., & Stewart, T. E. (2001). Handheld computers in critical care. *Critical Care*, *5*(4), 227-231.

Lister, R., Bryan, G., & Tracy, M. (2000, July 3-5). The e-Babies Project: Integrated data monitoring and decision making in neonatal intensive care. *Proceedings of the European Conference of Information Systems*, Vienna, Austria (Vol. 2, pp. 1289-1293).

McGregor, C., Bryan, G., Curry, J., & Tracy, M. (2002, January 7-10). The e-Baby data warehouse: A case study. *Proceedings of the 35th Annual Hawaii International Conference on System Sciences* (HICSS), Big Island (pp. 3018-3024).

McGregor, C., Heath, J., & Wei, M. (2005a, March 29-April 1). A Web service based framework for the transmission of physiological data for local and remote neonatal intensive care. *Proceedings of the IEEE International Conference on E-Technology, E-Commerce and E-Service*, Hong Kong (pp. 496-501).

McGregor, C., Kneale, B., & Tracy, M. (2005b, July 4-7). Bush Babies broadband: On-demand virtual neonatal intensive care unit support for regional Australia. *Proceedings of the 3rd International Conference on Information Technology and Applications* (ICITA), Sydney, Australia (Vol. 2, pp. 113-117).

McGregor, C., Purdy, M., & Kneale, B. (2005c). Compression of XML physiological data streams to support neonatal intensive care unit Web services. *Proceedings of the IEEE International Conference on E-Technology, E-Commerce and E-Service,* Hong Kong (pp. 486-489).

NSW Health Department. (1994). *NSW midwives data collection*.

Panniers, T. L. (1999, November 6-10). Selecting appropriate problems for decision support system in the neonatal intensive care unit. *Proceedings* *of the AMIA Annual Symposium,* Washington, DC (p. 1133).

Schulman, J. (2003, November 8-12). NICU Notes: A Palm OS(R) and Windows(R) database software product and process to facilitate patient care in the newborn intensive care unit. *Proceedings of the AMIA Symposium,* Washington, DC (p. 999).

Shabot, M. M., LoBue, M., & Chen, J. (2000, January 4-7). Wireless clinical alerts for critical medication, laboratory and physiologic data. *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences,* Maui, HI (Vol. 1, p. 6).

Shin, D. I., Huh, S. J., Lee, T. S., & Kim, I. Y. (2003). Web-based remote monitoring of infant incubators in the ICU. *International Journal of Medical Informatics, 71,* 151-156.

Sukuvaara, T., Makivirta, A., Kari, A., & Koski, E. (1989, September 19-22). An intelligent intensive care alarming system. *Proceedings of the Conference on Computers in Cardiology* (pp. 225-228).

van der Kouwe, A. J. W., & Burgess, R. C. (2003). Neurointensive care unit system for continuous electrophysiological monitoring with remote Web-based review. *IEEE Transactions on Information Technology in Biomedicine, 7*(2), 130-140.

Wu, J.-H., Wang, S.-C., & Lin, L.-M. (2005). What drives mobile health care? An empirical evaluation of technology acceptance. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS '05).*

This work was previously published in Handbook of Research in Mobile Business, edited by B. Unhelkar, pp. 83-95, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.4
Utilizing Mobile Phones as Patient Terminal in Managing Chronic Diseases

Alexander Kollmann
ARC Seibersdorf Research GmbH, Austria

Peter Kastner
ARC Seibersdorf Research GmbH, Austria

Guenter Schreier
ARC Seibersdorf Research GmbH, Austria

ABSTRACT

Mobile information and communication technologies are advancing rapidly and provide great opportunities for home monitoring applications in particular for outpatients and patients suffering from chronic diseases. Because of the ubiquitous availability of mobile phones, these devices can be considered as patient terminals of choice to provide a telemedical interaction between patients and caregivers. The most challenging part still is the patient terminal, that is, to offer the user a method to enter measured data into a system as well as to receive feedback in a comfortable way. The objective of this chapter is to present and compare solutions for mobile-phone-based patient terminals as developed by us and other authors.

INTRODUCTION

"Telemonitoring is defined as the use of audio, video, and other telecommunications and electronic information processing technologies to monitor patient status at a distance" (Field, 1996, p. 271). This concept may be particularly suitable in the management of chronic diseases where a close partnership as well as collaboration between patient and healthcare provider are essential. New paradigms such as prevention and patient empowerment promote the development of novel care approaches in which outpatient monitoring is a basic aspect.

Rapid advancements of information and communication technologies and the increasing availability of mobile phones open new perspectives in using these devices for tele-monitoring applications to deliver healthcare to people geographically remote from physicians or medical centers. The possibility to use the mobile phone for standard voice communication as well as for the transmission of a variety of multimedia information like text, audio, images, and videos makes it the communication interface of choice for patient-centered tele-monitoring applications.

The basic idea is to track patients' personal health status using the mobile phone as a patient terminal and to send the data to a remote monitoring centre. An automated monitoring process checks the values and gives feedback in order to guide the patient through the self-managing process and to turn the doctor's or other caregiver's attention to the patient when necessary by means of notifications and alerts. The most challenging part in this scenario still is the patient terminal, that is, to offer the user an easy method to enter measured data into a system as well as to get feedback of the current health status in a comfortable way.

This chapter will focus on the usage of mobile phones in the management of chronic diseases and gives an overview of available technologies. Furthermore, it will present and compare already implemented mobile-phone-based home monitoring concepts as developed by us and other authors.

BRIEF HISTORICAL OUTLINE

Basically, tele-monitoring combines topics from the fields of medicine, information and communication technology, and computer science. Particularly, information and communication technologies have undergone rapid advancements over the past decades, driven by the needs of modern information society. Communication devices such as mobile phones or personal digital assistants (PDAs) became smaller and more powerful, and advanced from single-purpose stand-alone devices to multipurpose networked devices that make them usable for tele-monitoring applications indeed.

However, reviewing the literature, the exact date when tele-monitoring was first mentioned in healthcare is still unknown (Brown, 1995). Starting with the first words transmitted by telephone in 1876 by Alexander Graham Bell, communication technology was ready to be used to facilitate healthcare services. For example, William Einthoven, the father of electrocardiography (ECG), transmitted ECG signals over wired telephone lines in 1906 (Barold, 2003).

In the 1930s, when the telephone became standard equipment in households, it also became the mainstay of medical communication and remained a major element until today. Wireless communication technologies were invented at the same time. Around the time of World War I, radio communication was established in a wider area, and, around 1930, it was used in remote areas such as Alaska and Australia to transfer medical information (Zundel, 1996).

Besides pioneering efforts of a few physicians using off-the-shelf commercial equipment to overcome time and distance barriers, current tele-monitoring concepts originated from developments in the manned space-flight program introduced by the National Aeronautics and Space Administration (NASA) in the early 1960s (Brown, 1995). The main intention was to monitor physiological parameters like heart rate, body temperature, ECG, and oxygen and carbon-dioxide concentration of astronauts in space and transmit the data to earth in order to establish an understanding of the health and well-being of the astronauts while they were in orbit.

Nowadays, tele-monitoring is adjudicated an important role in health systems since for a number of indications the cost effectiveness and the medical benefits have been approved (Meystre, 2005). Moreover, there is still a driving force to improve outpatient care by shifting tasks from hospitals to patients' homes, particularly in the management of chronic conditions, which will be discussed in detail in this chapter.

MANAGING CHRONIC DISEASES

Chronic diseases are cited in literature as diseases that have one or more of the following characteristics: They are permanent, leave residual disability, are caused by nonreversible pathological alteration, require special training of the patient for rehabilitation, or may be expected to require a long period of supervision, observation, or care (National Library of Medicine, 2005).

Chronic diseases such as heart failure, hypertension, cancer, diabetes, and asthma were the major cause of death (59% of the 57 million deaths annually) and global burden of disease (46%) in 2004 (World Health Organization [WHO], 2005).

It is also obvious that chronic diseases tend to become more common with age. WHO reported that in 1998, 88% of the population in developed countries over 65 years old suffered from at least one chronic health condition. Since populations are aging worldwide and chronic diseases have a significant impact on healthcare systems, new strategies in prevention, early detection of illness pattern, and long-term treatment are needed.

To address the quality and effectiveness of healthcare services for chronic-disease management, the chronic care model was developed at the MacColl Institute for Healthcare Innovation in 1998 (Wagner, 1998). The model describes the transformation of healthcare from a system that is essentially reactive, responding mainly when a person is sick, to one that is proactive and focused on keeping a person as healthy as possible. Besides strategic and lasting changes in health systems, the model suggests enforcing patients' central role in their care and self-management (Epping-Jordan, Pruitt, Bengoa, & Wagner, 2004).

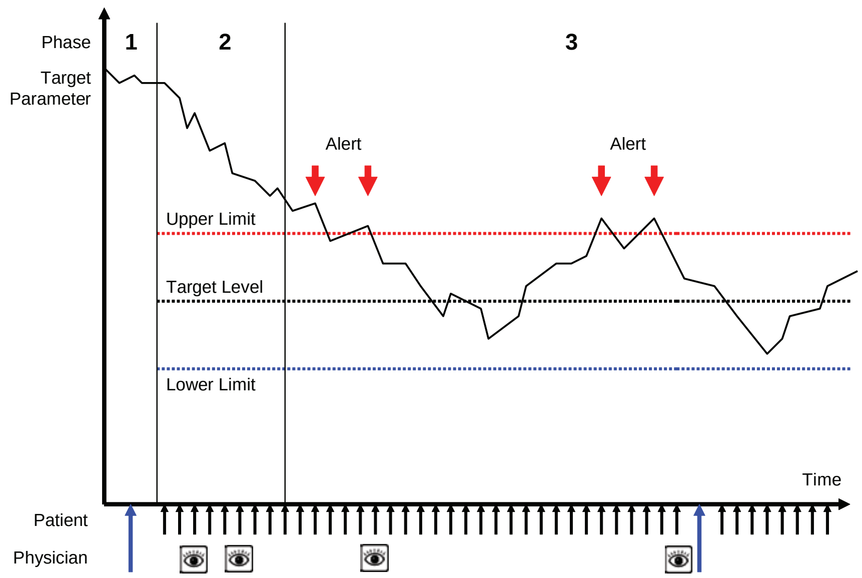
Effective self-management support means more than just telling patients what to do. Because of the fact that neither the chronic condition nor its consequences are static, the process of self-management might be complex and patients are often overstrained. Thus, a close partnership between patients and caregivers is essential in order to guide the patient through the self-management process and to the best possible health status.

It is widely recognized that the monitoring of health-related parameters like blood pressure, blood glucose, or well-being in regular intervals is the central element in the strategy of an effective self-management of chronic conditions. Basically, monitoring is defined as periodic measurements that guide the management of a chronic or recurrent condition. It can have an impact on the improvement of therapy adherence, the better selection of treatments based on individual response, or the better titration of medication (Glasziou, Irwig, & Mant, 2005).

Depending on the target parameter and the phase of treatment, monitoring can be done by clinicians, patients, or both. The strategy of monitoring can be divided mainly into three phases shown in Figure 1. Blue arrows indicate visits at the physician's office. The eye icon indicates tele-monitoring activity. Short, black arrows show the patient's self-monitoring activity.

- **Pretreatment.** Monitoring before treatment is mostly performed at the physician's office by the physician. The main goal is to access the patient's current health status and to verify the need for a medical intervention. Beside standard measurements like the assessment of blood pressure, more complex diagnostic tests like blood tests or urine tests are performed to confirm decisions. At that stage, a person has to be accustomed to his or her new role as a patient in the patient-

Figure 1. Three phases of monitoring (pretreatment, adoption, long-term treatment)



physician relation. Moreover, objectives of treatment, monitoring parameters, target levels, and their upper and lower limits should be well defined.

- **Adoption.** After establishing the target level, application of medication will be started in order to reach the objectives of treatment. Monitoring disease-related parameters, for example, blood pressure in case of hypertension, will show the effect of treatment in a representative way. In case of the necessity to titrate medication, monitoring has to be done very carefully and in shorter intervals to avoid abnormalities or the worsening of the health status.
- **Long-term treatment.** Although the target level has been reached periodically, monitoring is essential because the course of a chronic condition is rarely static. Overstepping or undershooting the predefined target level should be observed carefully. Additionally, alerts could be generated in order to turn the patient's and physician's attention to this special situation. Periods of problems can easily be identified via graphical representation. If noticed problems cannot be solved via tele-medical intervention, the patient has to be ordered to the physician's office for further examination. However, the main objective of tele-monitoring is to detect illness patterns at the earliest possible stage in order to avoid emergency situations and hospitalization.

Today, patients are asked to track their key measures like blood pressure, heart rate, diabetes-relevant data, well-being, or side effects of medication by daily taking notes on a piece of paper, called a health-data diary. The captured data are expected to show trends in the illness patterns and to help the doctor to guide the patient to the best possible health status.

However, patients' motivation for using the conventional method in self-management is often poor. This is hardly surprising since patients are often confronted with complex documents and instructions. On the other hand, paper-based diaries lack proper data representation, feedback, and timely delivery of data. Therefore, an easy-to-use and patient-centered data-acquisition system is essential to guide the patient through data capturing and the complex process of self-management.

USING INFORMATION AND COMMUNICATION TECHNOLOGIES IN HOME MONITORING ENVIRONMENT

The effective management of chronic diseases requires a close partnership between the patient and healthcare provider, which can be supported by contemporary information and communication technologies (Celler, Lovell, & Basilakis, 2003). The timely delivery of data is indispensable to detect an aggravation in illness patterns and to ensure appropriate medical decisions at the earliest possible stage.

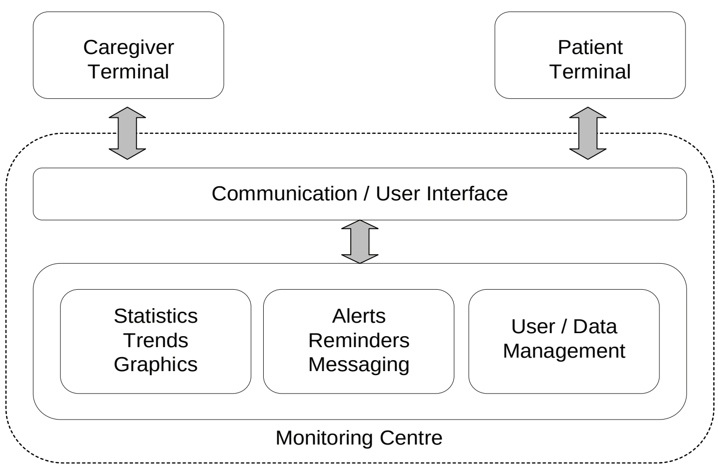
The basic idea is to track the patient's personal health status using a patient terminal, and to transmit the data to a remote monitoring centre where the data are processed and trends, statistics, and graphical representations are generated. An automated monitoring process checks the values and gives feedback. When necessary, the doctor's or other caregivers' attention will be turned to the patient by the means of notifications or alerts.

In our research, we developed a home monitoring platform for regular and home-based measurement and transmission of health parameters like blood pressure, body weight, symptoms, and medication. The system, shown in Figure 2, has been built using mostly standard components and state-of-the-art Internet technology, and it comprises the following.

Monitoring Centre

A 24/7-accessible server system receives, stores, and processes the data. The results are forwarded to the caregivers in a standardized and easily comprehensible format. Central components are user and data management to ensure the security, integrity, and traceability of data. Role-based, hierarchical user management guarantees that only authorized users are able to view, edit, or enter data in the system.

Figure 2. Main components of a home monitoring platform



Patient Terminal

The patient terminal provides the user with an adequate communication interface for entering data into the system as well as to receive feedback or medical advice. Although the availability of PCs (personal computers) with Internet access is already high in developed countries, the use of computers still presents a barrier to the adoption of Web-based solutions, especially by the elderly and technically unskilled people. Hence, different solutions have been developed during the last years to give such users an easy-to-use and intuitive way for entering their health data comfortably using manual, automatic, or semi-automatic methods in conjunction with standard medical measurement devices.

Caregiver Terminal

The Web-based caregiver terminal gives authorized people access to assigned patients and their data via PC or mobile Internet-enabled devices like PDAs or tablet PCs. Specially designed GUI (graphical user interface) components give the physician a quick overview and support straightforward navigation. Transmitted values as well as statistics, trends, and graphical representations can be accessed easily.

Statistics, Trends, and Graphics

An automated process analyses incoming data and generates statistics, trends, and graphical representation. Additionally, the process is able to check for alert conditions (limits given by the responsible physician) or to correlate the values with predefined patterns to identify abnormalities in the course of treatment.

Messaging System

In case of detected aberrance, notification messages will be sent to the physician or the patient via SMS (short-message service) or e-mail immediately in order to initiate further action. Additionally, it turned out that reminder messages are useful for patients and caregivers. The possibility to send interactive messages in several ways enforces communication between patients and caregivers.

Communication and User Interface

Besides the possibility to access or enter data via a Web interface, interfaces to several wireless communication technologies such as WAP (wireless application protocol), SMS, and MMS (multimedia messaging service) are provided in order to send or receive data.

GOING WIRELESS

A current trend in information and communication technologies is the convergence of wireless communication and computer networks as well as a moving from stand–alone, single-purpose devices to multipurpose network devices. Up to now, several studies have demonstrated that tele-monitoring using information and communication technologies on different integration levels - reaching from a simple telephone call to wearable or implantable sensor systems - can effectively assist patients in the management of care. Although the use of older approaches (telephone, fax) is still common, latest innovations in computer and network technologies can be considered for tele-monitoring applications to support the patient with a method to enter and transmit data in a comfortable way.

The Telephone as Communication Interface

Telephone care services represent the oldest method to deliver medical advice to the patients' homes. Nowadays, standard telephone lines are available in almost every household and even technically unskilled people are able to handle the telephone. For monitoring purposes, patients are contacted by a call centre to hand over their key measurements in predefined follow-up intervals. Specially trained personnel check the values against the patient's history and predefined limits and give feedback in order to assist the patient in self-management.

The usage of telephone interventions in the management of outpatients suffering from chronic heart failure was demonstrated in the DIAL(randomised trial of telephone intervention in chronic heart failure) trial (GESICA, 2005): 1,518 patients with stable chronic heart failure and optimal treatment were enrolled in 51 centers in Argentina. The DIAL trial intervention strategy was based on frequent telephone follow-ups provided by nurses trained in heart failure. The purpose of interventions was mainly to educate and monitor the patients as well as to increase the patients' adherence to diet and drug treatments. The results indicate that patients in the usual-care group were more likely to be admitted to the hospital for reasons of worsening and they were also more likely to die than patients in the tele-monitoring group who received telephone intervention. Moreover, patients in the intervention group showed a better quality-of-life score than patients randomized to the usual-care group.

Besides the possibility to use plain old telephone systems (POTS) for voice communication, they can also be used to access the Internet and network-based data services. Medical measurement devices like blood-pressure meters, blood-glucose meters, or scales equipped with modems can be connected to the POTS directly or by the means of a home terminal.

The Trans-European Network Home-Care Management Systems study (TEN-HMS) was the first large-scale, randomized, prospective clinical trial to decide if home-based tele-monitoring services for heart-failure patients are able to reduce hospitalizations and to improve patient well-being while reducing the overall costs of care (Cleland, Louis, Rigby, Janssens, Balk, & TEN-HMS, 2005). In total, 426 patients were randomized to the control group, nurse telephone-support group, or tele-monitoring group. The results indicate that patients randomized to the tele-monitoring group faced a reduced number of days spent in the hospital (minus 26%). Furthermore, it led to an overall cost saving compared to the nurse telephone-support group (minus 10%). Tele-monitoring also significantly improved survival rates relative to the usual-care group and led to high levels of patient satisfaction.

Mobile Phone as Communication Interface

Recent statistics indicate that the number of mobile phones throughout the world exceeded 1.5 billion in 2004 (CellularOnline, n.d.). Moreover, it is estimated that within a few years, about 70% of cell phones in the developed countries will have Internet access. Thus, mobile phones have become potential devices for serving as patient terminals in tele-monitoring applications. Mobile phones as well as communication technologies have undergone incredible changes and advancements during the last years. The amazing employment of mobile phones started in the 1980s when first-generation (1G) cellular systems were introduced. This technology was based on analog circuit-switched technology. Low data rates prevented this technology from being used for data transfer. Up to now, most of the 1G networks have been replaced by second-generation (2G) wireless networks, which are based on digital circuit-switched technologies. Several standards have been developed in different parts of the world: the Global System for Mobile Communication (GSM) technology in Europe, code division multiple access (CDMA) technology in the USA, and personal digital communication (PDC) in Japan. Second-generation wireless networks are digital and expand the range of applications to more advanced voice services and data capabilities such as fax and SMS at a data rate up to 9.6 kbps, which still makes it mostly impractical for extensive Web browsing and multimedia applications.

Through the years, several advanced techniques based on 2G networks were introduced such as general packet radio service (GPRS) and enhanced data rates for global evolution (EDGE). These technologies, also known as 2.5G networks, make it possible to use several time slots simultaneously when sending or receiving data, resulting in a significantly increased data rate (171 kbps for GPRS, 384 kbps for EDGE). The data packages are sent over the network using an IP (Internet protocol) backbone so that mobile users can assess services on the Internet.

The Universal Mobile Telecommunications System (UMTS) presents the third generation (3G) of wireless communication technology. The broadband, packet-switched transmission concept supports the transmission of text, digitized voice, video, and multimedia at data rates up to 2 Mbps. At the moment, users in real networks can expect performances up to 384 kbps for downloading and at least 64 kbps for uploading data. Third-generation systems are expected to have the following features: fixed- and variable-rate bit traffic, bandwidth on demand, asymmetric data rates in the forward and reverse links, multimedia mail storage and forwarding, the capability to determine the geographic position of mobile units and report it to both the network and the mobile terminal, and international interoperability and roaming (Tachakra, Wang, Istepanian, & Song, 2003).

In the course of rapid advances in communication technology and increased data rates, new multimedia services appeared that are the basis for further developments of mobile-phone-based applications for healthcare.

Short-Message Service

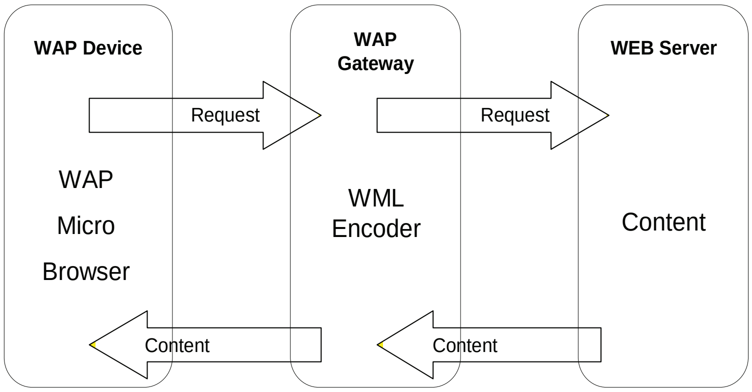
SMS allows sending and receiving text messages of up to 160 characters in length to and from mobile phones as defined within the GSM digital mobile-phone standard (Buckingham, 2000a). The text can comprise words or numbers or an alphanumeric combination. SMS is a store and forward service. This means that SMS messages are not sent directly from sender to recipient. Each mobile telephone network that supports SMS has at least one messaging centre to handle and manage the short messages.

To receive the text message on the server side, a mobile phone is connected to the server via standard interfaces such as RS232. Using the Hayes command set (Attention (AT) commands), the arrived messages can be accessed and stored in the database for further processing.

There are several reports on the use of SMS in medical application areas. SMS has been basically used for patient and appointment reminders. Outpatient clinics that are using SMS-based appointment reminder systems are seeing a reduction in missed appointments or "did not attends" (DNAs). It has been demonstrated that outpatient clinics that deployed SMS patient reminder systems saw DNA rates fall by as much as 30% even though less than 20% of patients chose to use the service (Research and Markets, 2005).

Besides the usage of SMS as a reminder service, SMS has also been applied in several medical application areas to transmit data from patients' homes to a monitoring centre. In 2004, Ferrer-Roca, Cárdenas, Diaz-Cardama, and Pulido presented the use of SMS for diabetes management. Twenty-three diabetic patients (18 to 75 years old) were asked to transmit data such as blood-glucose levels and body weight to a central server. The server automatically answered via SMS with a

Figure 3. WAP programming architecture. The application is stored on the content server and the requested pages are generated dynamically



prerecorded acknowledgement, specific help, or warning messages when data were out of range. During an 8-month study period, an average of 33 SMS messages was sent per patient and month. Unfortunately, no medical benefit was reported. However, they conclude that SMS may provide a simple, fast, efficient, and low-cost adjunct to the management of diabetes at a distance.

On the other hand, medical benefit in asthma monitoring could be demonstrated by Ostojic, Cvoriscec, Ostojic, Reznikoff, Stipic-Markovic, and Tudjman (2005). Sixteen patients (24.6 +/- 6.5 years old) were asked to transfer peak expiratory flow (PEF) measurements at least three times a day during a 16-week study period. Patients randomized to the tele-monitoring group received a weekly adjustment by an asthma specialist based on the values sent to the monitoring centre. They reported that asthma overall was better controlled in the intervention group, according to their findings of reduced PEF variability.

Wireless Application Protocol

WAP is an open global standard for communication between mobile phones or other mobile devices and the Internet (Buckingham, 2000b). WAP-based technology enables the design of advanced, interactive, and online mobile services, such as mobile banking, Internet-based news services, or even tele-monitoring applications.

The WAP standard is based on Internet standards like the hypertext markup language (HTML), extensible markup language (XML), and transmission control protocol/Internet protocol (TCP/IP). Basically, it consists of the wireless markup language (WML) specification, which is a markup language derived from HTML. However, WML is strongly based on XML, so it is much stricter than HTML. WML is used to create Web pages including text, images, user input, and navigation mechanisms that can be displayed on a WAP micro browser.

The WAP architecture consists of a WAP device, a WAP gateway, and a content server (Figure 3). The handheld WAP device communicates with the content server, which stores information and responds to user requests. The gateway in between translates and passes information between the device and the server. To access an application stored at the content server, a connection to the WAP gateway is initialized. Thereafter, the WAP request is converted to HTTP (hypertext transfer protocol) and forwarded to the content server.

Upon handling the request, the requested content is returned to the gateway, transformed into WAP, and sent back to the device to be displayed via the micro browser. The new version of WAP, WAP 2.0, is a reengineering of WAP using a cut-down version of the extensible hypertext markup language (XHTML) with end-to-end HTTP connection. This means that using WAP 2.0, the WAP gateway becomes dispensable.

A couple of authors (Hung & Zhang, 2003; Salvador et al., 2005) have already demonstrated tele-monitoring applications using WAP technology after WAP browsers became a standard feature of mobile phones. It is mentionable that besides demonstrating principal functionality in several pilot trials, only few clinical trials using WAP for patient monitoring have been mentioned in the literature yet.

For example, Italian researchers have shown that aftercare and patient communication might be improved if patients are asked to fill in daily questionnaires using WAP on their mobile phones (Bielli, Carminati, La Capra, Lina, Brunelli, & Tamburini, 2004). They developed the Wireless Health Outcomes Monitoring System (WHOMS), which allows structured questionnaires to be sent to the patient by the medical management team. Each day, an SMS message was sent informing patients of the survey and giving a link to a WAP site accessible through a standard GPRS connection. Users were asked to rate symptoms such as pain, lack of energy, and difficulty sleeping. The collected data were viewable by the doctors in a graphical format that highlighted the patients' states of health.

Although 42% of the patients failed to fill in the questionnaires mainly because of neophobia and unfamiliarity with the technology, the researchers concluded that health-outcome monitoring using mobile phones can be the method of choice for future developments in quality-of-life assessments. They address the need to develop more user-friendly communication terminals supported by upcoming technologies for mobile phones to increase the adoption rate.

In the course of our research, a lot of effort has been done to make WAP technology applicable for the broad usage in home monitoring applications and the management of chronic diseases.

Figure 4. Graphical user interface generated by WML guides the patient through the data-acquisition process

The aim of the Cardio-Memory study (Scherr, Zweiker, Kollman, Kastner, Schreier, & Fruhwald, in press) was to evaluate whether WAP technology would be an acceptable, feasible, and reliable option to provide tele-monitoring for patients with chronic heart failure or hypertension. In the course of a clinical pilot trial, 20 patients (mean age 50 +/- 14 years) were enrolled. Each participant was equipped with a mobile phone, an automatic blood-pressure device, and a digital weight scale. Patients were asked to measure their blood pressure, heart rate, and weight every day. After accessing the system with the micro browser, the menu promptly routed the patients through subsequent entry templates generated by WML syntax (Figure 4). Eventually, all data were sent to the central database at the remote monitoring centre for further processing.

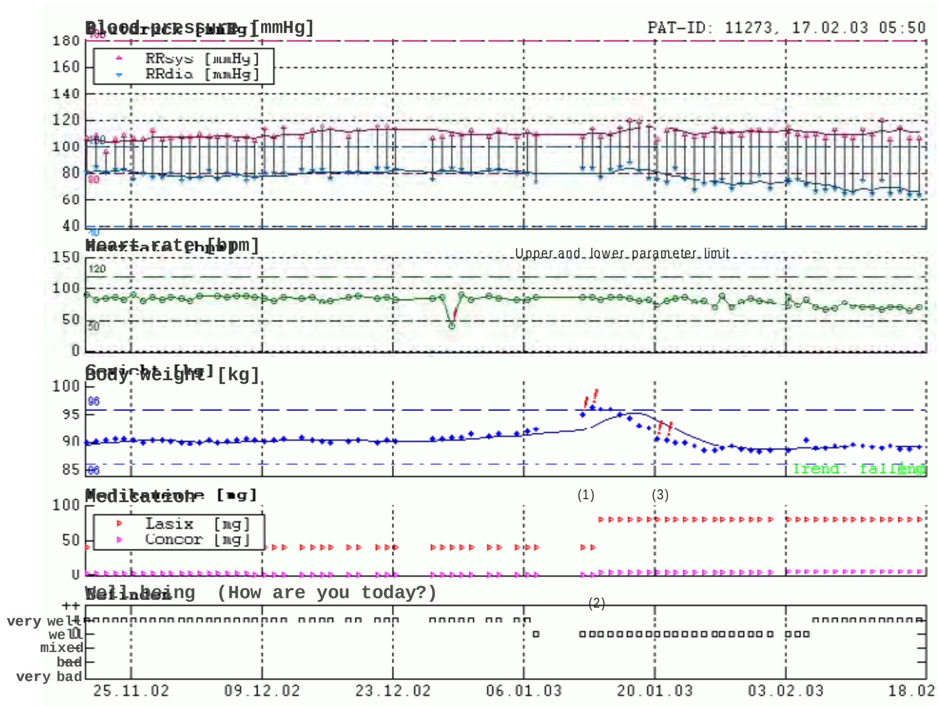
Authorized physicians could access the data via a secure Web site at any time. The Web site provided the data of each patient in numerical format and graphical trend charts, including patient-specific upper and lower parameter limits. Furthermore, it allowed physicians to set automatic reminders: These computer-generated SMS messages reminded patients to take their medications, weigh themselves, measure blood pressure and heart rate, and to transfer the respective data to the secure server at the monitoring centre. Furthermore, the physician could individually configure an automatic warning system for each patient. If the patient's values exceeded individually predefined vital parameter limits, physicians were notified immediately by a computer-generated SMS or e-mail warning. If an intervention (e.g., adjustment of medication dose) was indicated (for example, because a patient exceeded a vital parameter limit), the physician was able to contact the patient directly via mobile phone to confirm the parameters and ask the patient to make an adjustment in medication.

During the 3-month monitoring period, there were 2,040 data-transfer sessions (mean 102 +/- 43 per patient). Only 1 out of 20 patients dropped out after 25 days due to severely impaired vision. The Cardio-Memory system was evaluated through a questionnaire at the end of the study. Eighteen out of 20 patients rated the software as easy to operate, which allows us to conclude that overall patients' acceptance with the system was high. The entire process (measurements and data transfer) took approximately 3 minutes to be completed. Moreover, patients felt that computer-generated reminders about missing data increased their compliance with the treatment regime.

An example of a successful tele-medical intervention of a patient suffering from chronic heart failure is shown in Figure 5. Cardiac decompensation caused a weight gain of more than 2 kg within 3 subsequent measurements and led to a patient alert (1). The physician contacted the patient and increased the diuretic dose (2). Subsequently, the patient lost weight and remained in stable conditions until the end of the observation period (3).

Following the promising results of the Cardio-Memory trail, an advanced study was set up in October 2003 in order to evaluate WAP technology in the management of chronic heart failure (Scherr et al., 2005). In a randomised, prospective, multi-center study, 240 patients who had been admitted to the hospital because of heart failures will be randomised to either pharmacological treatment (control group) or to pharmacological treatment plus tele-medical care (telegroup). Telegroup patients are provided with a mobile telephone, a digital weight scale, and a fully automated blood-pressure device. During the follow-up period of 6 months, patients are asked to send their self-measurements using a WAP application on the mobile phone on a daily basis. Up to now, 65 patients (45 male, 20 female; 64 ± 11 years old) from six centers have been randomised, and 44 patients have completed the study so far. Three patients from the telegroup dropped out due to being not able to handle the monitoring system equipment, in particular, the mobile phone.

Figure 5. Trend chart of a patient monitored with Cardio-Memory



Intermediate results indicate that WAP-based tele-medical surveillance of patients with a recent episode of acute heart failure significantly contributes to an improvement of functional status and may be a promising tool to improve heart-failure therapy and reduce emergency situations and hospitalizations.

Java-Based Software Application Running on Mobile Phones

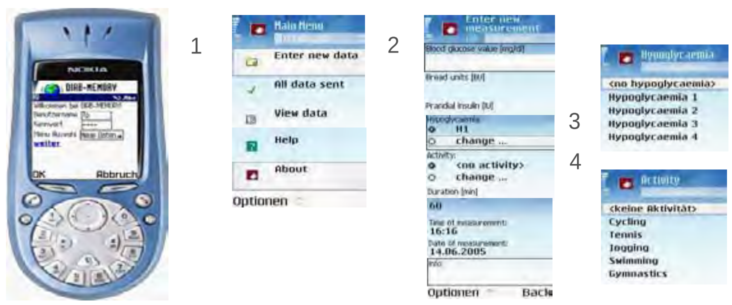
Java 2 Platform, Micro Edition (J2ME), the small footprint version of the Java technology, is optimized to run on memory-constrained devices like mobile phones, PDAs, and TV-set-top boxes. Each device category receives its own profile that includes a set of category-specific application program interfaces (APIs) and a configuration that consists of a minimum set of APIs and a Java virtual machine. In particular, mobile phones support the mobile information device profile (MIDP), which includes APIs covering the user interface, networking, persistent storage (a record-oriented database), security, and messaging (Sun Microsystems, 2002).

The usage of a J2ME-based software application running on mobile phones has been evaluated in a pilot trial concerning the management of diabetes patients (Riedl, Kastner, Kollmann, Schreier, & Ludvik, 2005). The software has been designed to support the user in entering diabetes-related data like blood-glucose level, insulin dose, bread units, well-being, and activities with remote synchronization to the database at the central monitoring centre. The graphical user interface is shown in Figure 6.

After having logged in, the data is entered via the well-designed graphical user interface and stored directly in a local database on the mobile phone. From time to time, the user can initiate a synchronization process via HTTP request and

Figure 6. A Java-based software application running on mobile phones supports the user in home-based data acquisition of diabetes-related key measurements

Flow chart



1 … User authentication
2 … Data entry (Blood glucose, bread units, insulin dose, …)
3 … Hypoglycaemia classification if occurred [1 to 4]
4 … Physical activity if performed

GPRS data transfer. After the uploading of the data to the remote monitoring centre, the user is able to access several graphical representations, statistics, and trends on his or her data through a secure Web interface.

The software has been evaluated in a clinical pilot trial. Ten diabetes mellitus type 1 patients (36.1 ± 11.4 years) were asked to collect diabetes-related data (blood glucose, bread units, insulin dose, well-being, activity, etc.) at least 3 times and up to 10 times a day by using J2ME software application running on a NOKIA 7650 mobile phone. In case of less than 3 transmitted data records a day, an automatic reminder SMS message was sent to the patient. The total number of received messages was 10,053 (1,257 +/- 351 per patient). Average blood-glucose level during the initial and final 14 days was 138.5 +/- 63.7 vs. 137.1 +/- 69.2 mg/dl, and HbA1c values were 7.7 +/- 0.4% vs. 7.6 +/- 0.4%, respectively. Although no tele-medical intervention by the physician was yet established, corresponding parameters decreased due to an advanced adherence to therapy. Additionally, ubiquitous availability and the appealing user interface were well perceived by the patients resulting in a high success rate for data transmission and patient acceptance.

Semiautomatic Methods

Running a J2ME-based software application on mobile phones supports also the possibility to access integrated mobile-phone features like Bluetooth. Bluetooth wireless technology is a short-range radio technology that makes it possible to transmit data over short distances between mobile phones, computers, and other devices. Supporting standard measurement devices like blood-pressure meters or glucose meters with a Bluetooth communication module gives them the possibility to transfer data to the mobile phone in an automated way. The software running on the mobile phone handles data exchange and triggers further data transmission to a central monitoring centre (Husemann & Nidd, 2005).

Multimedia Messaging Service

MMS, an extension to the SMS, defines a way to send and receive wireless messages that may include images, audio, and video clips in addition to text. When the technology will be fully developed, the transmission of streaming videos will be supported as well. A common current application of MMS messaging is picture messaging - the use of camera phones to take photos for immediate delivery to a mobile recipient.

MMS has been used in our research as a unique approach to the human-computer interface challenge based on digital-camera-enabled mobile phones. Values shown on the display of the measurement device were taken by a camera-enabled mobile phone. Special Symbian-based (Symbian, 2005) software, running on the mobile phone, helps the user to handle this data-acquisition process. All necessary configurations are made by the program automatically. If an online connection to the data carrier service (GPRS, UMTS) at a given location and time is not available, the mobile phone repeatedly tries to send the data until the message has been sent successfully. Data acquisition and transmission can be achieved with only two keystrokes.

At the monitoring centre, the MMS arrives as a photo attached to an e-mail. The subject of this e-mail contains the IMEI (international mobile station equipment identity) of the mobile phone. Thus, the photo can unambiguously be assigned to a patient. Other important parameters like capturing date and time are also stored within the e-mail.

Special software, running on the Web server, fetches the e-mail available on the mail server, extracts the photo, and moves it to the local file system. Subsequently, the incoming photo is registered to the database and assigned to the corresponding patient.

Thereafter, a special character-extraction algorithm starts to process the photo in order to extract the numerical values. In case of successful character extraction, the values are stored into the database and the report-processing unit is started.

Figure 7. "MoniCam" principle: A special software application helps the user to take a photo of the display and send it to a central monitoring centre via MMS, where the values are extracted and stored in the database automatically

Graphs and trends are generated and limit checks are performed. Finally, all information is made accessible via the Web interface to authorized users. In case of unsuccessful character extraction or failed plausibility check, the user may receive a message via SMS to repeat the measurement.

This new method has been evaluated in a feasibility study on five users with four different blood-pressure devices and two different camera-enabled mobile phones (Schreier, Kollmann, Kramer, Messmer, Hochgatterer, & Kastner, 2004). The results indicate that the rate of correct value extraction varied considerably with respect to the type of the measurement device but was comparable for the two different types of mobile phone. For two types of blood pressure meters the method was capable of determining the correct values in well above 90% of the cases. The MMS-based transmissions of the taken photos succeeded in all cases.

REQUIREMENTS ON MOBILE-PHONE-BASED TELEMONITORING APPLICATION

Because of the ubiquitous availability of mobile phones, these devices may be considered to be the patient terminal of choice to provide an interface to home monitoring systems. It has to be taken into account, however, that elder patients are often very unskilled and not familiar with the handling of mobile phones. To avoid rejection of this concept as well as to motivate the patients in using such devices, consequentially, an easy-to-learn and easy-to-use system for data acquisition is essential. Based on our experience, the ideal human-computer interface would have the following properties.

High Usability

Usability tells us how well the users can use the system productively, efficiently, and pleasantly to reach the goals in a certain environment. Using mobile phones as home monitoring terminals is limited in terms of input interaction (small buttons to enter data as well as to navigate) and output interaction (small display size and poor resolution). To overcome these limitations, an appealing user interface, well-structured graphical design (metaphoric), and intuitive navigation are essential. The number of buttons to be used as the necessary number of button clicks for navigation and entering data should be reduced to a minimum. Moreover, the data-acquisition process should be easy to learn, intuitive, and well structured to decrease barriers against technical devices and improve users' satisfaction in using the system.

Low Cost

The main driving force for introducing new innovations into healthcare is to reduce costs. Hence, patient terminals for home monitoring applications should ideally be based on off-the-shelf technology available without significant set-up expenses or extra costs.

Error Resistant

User actions as well as data entries should be checked for plausibility on the earliest possible stage. For example, a data-entry plausibility check on the side of the patient will exclude errors at origin. A major aspect in handling errors is to provide the user with well-defined error messages that are also easy to interpret and to understand. Well-structured error handling can guide the user through difficulties without confusing him or her.

Off-Line Data Acquisition

It is a well-known fact that wireless networks especially sometimes lack availability (e.g., in buildings). To provide the user with the possibility to enter data at any time and any location, intermediate local data storage is desirable. When the network is available again, stored data can be synchronized with the central database.

High Flexibility and Adaptability

Patient terminals should provide flexibility in terms of adding or removing parameters to the user interface, corresponding to changing conditions in the treatment of chronic diseases. For meeting the needs of an adjusted, personalized patient terminal, the system should support the possibility to interrogate different parameters depending on the time of day or the patient's history. The user interface for daily use should be configurable to different patients and conditions such as displaying the appropriate drug name to enter the correct dosage. Ideally, user interactions or setup procedures are no longer necessary.

High Security Level

Handling data about the health status of patients, in general, requires a high standard of security. In most countries, end-to-end encryption is compulsory to meet security directives and laws. As a consequence, most standard communication technologies and protocols already provide some sort of data protection to guarantee that data cannot be accessed, stored, or manipulated while they are transmitted via Internet or wireless networks.

Bidirectional Communication

The ideal patient terminal combines the possibility to provide the user with an interface not only for entering data but also for receiving feedback, reminders, or medical advice on the same device.

Device Independency

Currently, numerous models of mobile phones from various manufacturers are available on the market to be used as patient terminals. Therefore, a software application that provides a graphical user interface highly demands independency of device specifications like operating system, display resolution, mobile Internet browser, menu

Table 1. Overview of the properties and technologies (+ means the method complies with the requirement, +/- means it complies only partly, - means does not comply)

	Technology				
Requirement	SMS	WAP	J2ME	MMS	Automatic
Overall usability	-	+/-	+	+	+/-
Low cost	+/-	+/-	+/-	-	-
Error resistant	-	+/-	+	+	+
Off-line data acquisition	+	-	+	+	+/-
High level of flexibility/ adaptability	-	+	+/-	-	-
High security level	-	+	+	+/-	+
Bidirectional communication	+	+/-	+/-	+	+/-
Device independency	+	-	+/-	+	-

navigation, and so forth to ensure high usability for a wide range of mobile phones.

Table 1 gives an overview of the requirements and the degree of compliance of the various technologies as discussed in this chapter and proposes a three-level classification based on our experience and the results of trials of other authors.

SMS

Using SMS technology for data acquisition is not very suitable due to the lack of usability and the fact that a graphical user interface cannot be provided. Basic skills in handling mobile phones as well as knowledge in using the keypad are essential to enter values in a predefined template. SMS for data transmission can be considered when only a single value has to be transmitted (e.g., PEF value in case of monitoring asthma patients; Ostojic et al., 2005). On the other hand, SMS is very suitable for sending messages, automatically generated reminders, or medical advices from the healthcare provider to the patient as an additional way of communication.

WAP

Several studies have been performed utilizing a WAP browser to provide the user with a user interface to enter data. Although a graphical user interface is provided, basic skills in handling a mobile phone and using the numeric keypad are important. Moreover, WAP lacks the possibility of client-side data-entry plausibility checks. The data have to be sent to the server before the values can be processed or checked for plausibility. In case of an error, the user is prompted to the page where the error occurred and asked to reenter the data.

On the other hand, WAP technology provides high flexibility and adaptability because WML scripts are server based and can be generated dynamically depending on user settings and requirements. This fact makes this method seem quite suitable for providing dynamical questionnaires such as quality-of-life assessments or to track medication intake, which is quite different from patient to patient. WAP technology may also be useful were a stepwise data input is required.

A disadvantage of WAP-based systems is that an online connection has to be established during the whole data-entering process. We experienced that this fact can lead to problems. For example, when the data connection is lost, users are confused by the corresponding error message. Thus, incomplete record sets are common and some patients were not able or willing to use the method for daily data acquisition.

Java

Java-based software is less affected by temporary lack of network availability because the data can be stored locally on the mobile phone and the data-transmission process may be postponed until network connectivity is available. Additionally, implementation of checking the plausibility of the entered data is also feasible, resulting in a lower error rate.

Moreover, Java technology allows one to improve usability through the design of user-friendly GUIs by using metaphoric elements. However, updating software applications running on the mobile phones is difficult and requires some additional user experience, although over-the-air application downloading (OAD) simplifies the way applications are delivered to customers.

In addition to standard TCP and HTTP, the Java environment supports the secure hypertext transfer protocol (HTTPS) using standard secure socket layer (SSL) to enable encrypted, secure connections. J2ME is particularly suitable for monitoring applications where frequent data acquisition is necessary (e.g., diabetes) and provides a fast and efficient method to enter, store, and transmit data.

Irrespective of which technology will be used, the most challenging part in developing mobile-phone-based applications for home monitoring is to guarantee interoperability. For example, WAP technology is strongly based on standards but every mobile-phone manufacturer speaks a slightly different WAP language. According to our experience, J2ME applications also change their looks depending on the type of the mobile phone.

MMS

Using a camera-enabled phone for data acquisition by taking a photo of the display provides a really easy and intuitive method for health-data acquisition at home. On the other hand, there is a huge effort on the server side to provide an image-processing algorithm to extract and interpret values correctly. This method lacks also in terms of flexibility and adaptability. Furthermore, costs for the MMS picture transfer are an obstacle for daily usage.

Automatic

Some measurement devices are already equipped with short-range wireless communication technologies like Bluetooth or infrared to transmit data to the mobile phone, which operates as a hub and a gateway to relay those data to a central monitoring centre. Because there is no mass market for such devices yet, they are usually considerably more expensive. Another weak point of this concept, according to our experience, is that it is sometimes not straightforward to establish communication between the mobile phone and the measurement device. Depending on the model, it may be necessary to navigate deep into the menu for setting up the connection. This makes such methods often unusable for technically unskilled people. However, once the system is set up correctly, data acquisition is fairly automated and error resistant. Using the mobile phone as a hub provides a bidirectional communication between patients and healthcare providers; hence, the mobile phone can receive SMS. Therefore, in the long run, automated systems will be the method of choice for elderly and unskilled or handicapped patients to transmit their self-measurements to the monitoring centre.

DEVELOPING MOBILE APPLICATIONS FOR TELEMONITORING APPLICATIONS

Because of small display sizes, limited resolution, and restricted possibilities for user interaction, navigation design and the implementation of software applications for mobile devices is quite different from software developments on other platforms like stationary desktop PCs.

Mobile software applications can be divided into two groups:

1. Highly goal driven.
2. Entertainment focused.

Highly goal-driven services aim at providing fast replies to specific problems, whereas entertainment-focused services enable the user to pass the time, for example, by offering gossip, games, or sports results (Ramsay & Nielsen, 2001). Mobile applications for home monitoring are definitely goal-driven, that is, to provide the users with a method to enter data in the most suitable way. Therefore, a user-centered design process is necessary. This process involves a number of important phases.

Analysis

The initial phase of software development is the most critical one. In this stage, the requirements of the patient terminal should be clearly defined by determining the user-group characteristics and the monitoring scenario. Typical attributes of the user group are age, expertise, experience level, and physical limitations. The definition of the monitoring scenario comprises the parameter to be monitored, the way feedback is given, and the demand in terms of flexibility.

Thereafter, the use case of the specific monitoring scenario should be clearly described in order to define objectives and features that require close cooperation between technician and physician. Once the general use case has been defined, the technology has to be selected that best supports the requirements of the respective application. Table 1 can serve as a guide to get an overview of properties of available technologies. Features as well as limitations of the selected technology should be carefully balanced during the design process.

Design

According to our experience, the typical software application for home monitoring applications comprises the following four stages:

1. **Identification:** The user is asked to log onto the system with a unique user name and password combination to facilitate authentication and access control.
2. **Data acquisition:** Menus and input templates guide the user through the data-acquisition process.
3. **Transmission:** Entered data are either stored locally or are transmitted to a central database in a monitoring centre.
4. **Feedback:** To indicate that data storage or transmission has been executed successfully or the presence of special situations, feedback should be given to the user immediately after data transfer in a representative form.

The development of a consistent, easy-to-use application does not require coding at this stage. On the contrary, ideally, a conceptual framework is established, representing the application and its workflow on a metaphoric level (e.g., a storyboard).

Evaluation

The concept designed in Step 2 should be evaluated by both types of users: patients and physicians. Usability considerations should be made; for example, unnecessary button clicks or confusing workflows should be avoided. There is also the need to identify sources of errors. If significant changes are necessary, an iteration starting with the design step is required.

Implementation

There are several software-development kits (SDKs) available to develop software for mobile devices. However, irrespective of which SDK or programming language is adopted, developing software applications for mobile phones requires substantial experience in programming efficient and reliable software.

Testing

The developed software can be tested and debugged on emulators that are provided by the mobile-phone manufacturers. However, according to our experience, emulators are often error prone and sometimes special features are not supported. Hence, it is essential to test developed software on target devices themselves.

FUTURE TRENDS

Current trends in chronic-disease management enforce shifting tasks from the clinic to patients' homes. This means that self-management and collaboration between patients and caregivers will become more and more important and may benefit from upcoming technologies.

Technical Advancements for Mobile-Phone-Based Patient Terminals

Phones enabled with radio-frequency identification (RFID) and near-field communication (NFC) may soon be used as patient terminals to provide an intuitive and easy-to-use way for health-data acquisition at home.

RFID tags are able to uniquely identify an object, animal, or person, or to store data. They have been introduced in the industry as an alternative to the bar code. Passive RFID tags are powered by the magnetic field generated by the reader. The tag's antenna picks up the magnetic energy, and the tag communicates with the reader in order to retrieve or transmit data.

In a current research project, we developed a scenario where RFID tags are used in a home monitoring environment. Objects to be tracked or identified are equipped with RFID tags. For example, tagging medication boxes with RFID tags provides an easy and intuitive method for the patient to indicate which medication has been taken simply by touching the box with an RFID-reader-enabled phone. Special software running on the mobile phone fetches the information from the tag, adds a time stamp, and initializes a transmission to the monitoring centre automatically. Hence, no cumbersome user interaction and configuration is needed.

Electronic Data Capture in Clinical Trials

The combination of mobile and Web-based technologies will improve clinical-trial efficiencies through increased data accuracy, higher data yield per patient, and real-time access to trial data (Stokes & Paty, 2002) by using mobile phones for timely and patient-centered data acquisition. Patient diaries as well as consequent monitoring of health parameters of interest can add meaningful information about the safety and efficacy of a treatment and can save time and money. Additionally, quality-of-life data will also play a central role in future clinical trials, which can be accessed easily by mobile-phone-based software solutions.

LESSONS LEARNED

Self-management and cooperation between patients and healthcare providers are the basic aspects and strategies of efficient chronic-disease management. An important element in efficient self-management is the monitoring of health-related data reliably. Because of the ubiquitous availability of mobile phones, these devices can be used as patient terminals so as to provide the patient with a method to enter data easily as well as to receive feedback or medical advice from remote healthcare professionals.

The most critical part in this respect is the user interface. Software specifically developed to the needs of the respective patient group is required to guide them through the data-acquisition process. Mobile data services and transmission protocols like SMS, MMS, WAP, and HTTP can be used to exchange data and information between patients and their caregivers. These methods have already been evaluated in several clinical trials and feasibility studies, and medical benefit could be demonstrated as well. However, using mobile phones as patient terminals is limited due to small display, poor resolution, and small buttons for user interaction.

Up to now, there is no method that fulfils all criteria of an ideal patient terminal in terms of high usability, adaptability, flexibility, and low cost. Every method for entering data implies specific advantages and disadvantages. Hence, when designing a mobile-phone-based home monitoring system, the patient terminal that best fits into a particular monitoring application has to be chosen on an individual basis, depending on the requirements, the user group, and the medical demand.

REFERENCES

Barold, S. S. (2003). Willem Einthoven and the birth of clinical electrocardiography a hundred years ago. *Cardiac Electrophysiology Review, 7*, 99-104.

Bielli, E., Carminati, F., La Capra, S., Lina, M., Brunelli, C., & Tamburini, M. (2004). A wireless health outcomes monitoring system (WHOMS): Development and field testing with cancer patients using mobile phones. *BioMed Central Medical Informatics and Decision Making, 4*, 7.

Brown, N. (1995). *A brief history of telemedicine*. Retrieved September 25, 2005, from http://tie.telemed.org/articles/article.asp?path=articles&article=tmhistory_nb_tie95.xml

Buckingham, S. (2000a). *What is SMS?* Retrieved September 25, 2005, from http://www.gsmworld.com/technology/sms/intro.shtml

Buckingham, S. (2000b). *What is WAP?* Retrieved September 25, 2005, from http://www.gsmworld.com/technology/wap/intro.shtml

Celler, B. G., Lovell, N. H., & Basilakis, J. (2003). Using information technology to improve the management of chronic disease. *Medical Journal of Australia, 179*, 242-246.

CellularOnline. (n.d.). *Stats snapshot*. Retrieved September 25, 2005, from http://www.cellular.co.za/stats/stats-main.htm

Cleland, J. G. F., Louis, A. A., Rigby, A. S., Janssens, U., Balk, A. H. M. M., & TEN-HMS. (2005). Noninvasive home telemonitoring for patients with heart failure at high risk of recurrent admission and death: The Trans-European Network-Home-Care Management System (TEN-HMS) study. *Journal of the American College of Cardiology, 45*, 1654-1664.

Epping-Jordan, J. E., Pruitt, S. D., Bengoa, R., & Wagner, E. H. (2004). Improving the quality of healthcare for chronic conditions. *Quality and Safety in Healthcare, 13*, 299-305.

Ferrer-Roca, O., Cárdenas, A., Diaz-Cardama, A., & Pulido, P. (2004). Mobile phone text messaging in the management of diabetes. *Journal of Telemedicine Telecare, 10*, 282-285.

Field, M. J. (Ed.). (1996). *Telemedicine: A guide to assessing telecommunications for healthcare*. Washington, DC: National Academy Press.

GESICA Investigators. (2005). Randomised trial of telephone intervention in chronic heart failure: DIAL trial. *British Medical Journal, 331*, 425-427.

Glasziou, P., Irwig, L., & Mant, D. (2005). Monitoring in chronic disease: A rational approach. *British Medical Journal, 330*, 644-648.

Hung, K., & Zhang, Y. T. (2003). Implementation of a WAP-based telemedicine system for patient monitoring. *IEEE Transactions on Information Technology in Biomedicine, 7*, 101-107.

Husemann, D., & Nidd, M. (2005). Pervasive patient monitoring: Take two at bedtime. *ERCIM News, 60*, 70-71.

Meystre, S. (2005). The current state of telemonitoring: A comment on the literature. *Telemedicine Journal and E-Health: The official journal of the American Telemedicine Association, 11*, 63-69.

National Library of Medicine. (2005). *MeSH descriptor data* (C23.550.291.500).Rockville Pike: Bethesda. National Library of medicine.

Ostojic, V., Cvoriscec, B., Ostojic, S. B., Reznikoff, D., Stipic-Markovic, A., & Tudjman, Z. (2005). Improving asthma control through telemedicine: A study of short-message service. *Telemedicine*

Journal and E-Health: The official journal of the American Telemedicine Association, 11, 28-35.

Ramsay, M., & Nielsen, J. (2001). *WAP usability*. Fremont, CA: Nielsen Norman Group.

Research and Markets. (2005). *Mobile and wireless services for outpatients*. Retrieved September 25, 2005, from http://www.researchandmarkets.com/reports/297065/297065.htm

Riedl, M. , Kastner, P., Kollmann, A., Schreier, G., & Ludvik, B. (2005). *Diab-Memory: A smart phone based data service for intensified insulin therapy in patients with type 1 diabetes mellitus. A pilot study*. Proceedings of American Diabetes Association's 65th Annual Scientific Sessions, San Diego, CA.

Salvador, C. H., Carrasco, M. P., Gonzalez de Mingo, M. A., Muñoz Carrero, A., Márquez Montes, J., Martin, L.S., et al. (2005). Airmed-cardio: A GSM and Internet services-based system for out-of-hospital follow-up of cardiac patients. *IEEE Transactions on Information Technology in Biomedicine, 9*, 73-85.

Scherr, D., Fruhwald, F. M., Zweiker, R., Kastner, P., Schreier, G., & Klein, W. (in press). Mobile phone based surveillance of cardiac patients at home. *Journal of Telemedicine and Telecare*.

Scherr, D., Kollmann, A., Hallas, A., III, Krappinger, H., Auer, J., Kastner, P., et al. (2005). Telemonitoring for heart failure patients following acute decompensation: First results on influence of the system on functional status and heart failure therapy. *European Heart Journal, 26*(Suppl.20).

Schreier, G., Kollmann, A., Kramer, M., Messmer, J., Hochgatterer, A., & Kastner, P. (2004). Mobile phone based user interface concept for health data acquisition at home. In J. Klaus (Ed.), *Computers helping people with special needs* (pp. 29-36). Springer.

Stokes, T., & Paty, J. (2002). Electronic diaries, Part 1: What is a subject diary, and how do regulations apply? *Applied Clinical Trials*. Retrieved September 25, 2005, from http://www.actmagazine.com/appliedclinicaltrials/article/articleDetail.jsp?id=83521

Sun Microsystems. (2002). *Java™ 2 Platform, micro edition*. Retrieved September 25, 2005, from http://java.sun.com/j2me/docs/j2me-ds.pdf

Symbian. (2005). *Symbian OS: The mobile operating system*. Retrieved September 25, 2005, from http://www.symbian.com/

Tachakra, S., Wang, X. H., Istepanian, R. S. H., & Song, Y. H. (2003). Mobile e-health: The unwired evolution of telemedicine. *Telemedicine Journal and E-Health: the official Journal of the American Telemedicine Association, 9*, 247-257.

Wagner, E. H. (1998). Chronic disease management: What will it take to improve care for chronic illness? *Effective Clinical Practice: ECP, 1*, 2-4.

World Health Organisation. (2005). *Facts related to chronic diseases*. Retrieved September 25, 2005, from http://www.who.int/dietphysicalactivity/publications/facts/chronic/en/

Zundel, K. M. (1996). Telemedicine: History, applications, and impact on librarianship. *Bulletin of the Medicine Library Association, 84*(1), 71-79.

APPENDIX: AUSTRIAN RESEARCH CENTERS GMBH-ARC

Austrian Reseach Center GmbH-ARC (ARC-sr) is Austria's largest centre for applied research and development, employing over 500 highly qualified specialists. The corporation's portfolio is strongly oriented toward national and European-wide research projects and development programs. ARC-sr comprises nine business divisions. The Division of Biomedical Engineering comprises research teams in the fields of e-health systems and smart biomedical systems, and cooperates with all major Austrian medical and technical universities in Vienna, Graz, and Innsbruck as well as a number of additional medical institutions and industrial partners within Austria and abroad.

The E-Health Systems Research Team focuses on e-health-related applications and research and development projects according to its mission: "to provide new connections between patients, physicians and other healthcare partners." Accordingly, we developed several home monitoring solutions utilizing modern information and communication technologies and a 24/7 monitoring service centre to establish these new connections between physicians and patients. The integration of different devices and monitoring indications as well as standardization are major objectives. The ultimate goal of our tele-, home, and health-monitoring concept is to empower physicians and patients with advanced data-acquisition methods for prevention, diagnosis, and therapy management.

This work was previously published in Web Mobile-Based Applications for Healthcare Management, edited by L. Al-Hakim, pp. 227-257, copyright 2007 by IRM Press (an imprint of IGI Global).

Chapter 3.5
Implementation of a Computerized System in an Oncology Unit

Taxiarchis Botsis
Athens Medical School, Greece

Konstantinos Syrigos
Athens Medical School, Greece

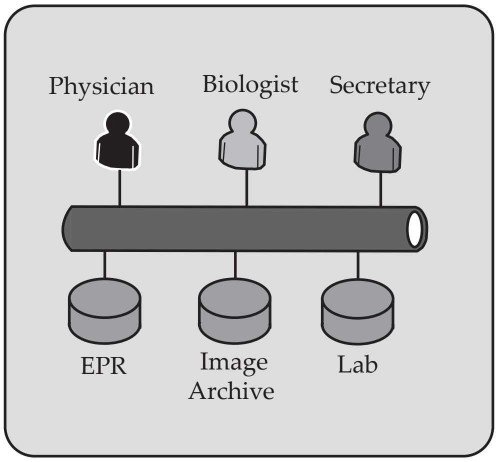
ABSTRACT

Information management is essential for health professionals in order to maintain a level of productivity for health care services management. This is significant when treating cancer patients. The main target of this study was to employ computers to enhance the daily practice of Oncology Unit (Sotiria Hospital, Athens, Greece). Accordingly, a computerized system was developed consisting of three modules: the EPR, the Image Archive, and the Lab Module. The EPR Module is a database application that stores clinical results, physician orders, and several administrative data. The Image Archive Module is used mainly for the reduction of images volume and the Lab Module stores information about the patient blood samples. These two modules interoperate through EPR Module under strict data security policies. Key physicians, biologists, and secretary personnel are involved in data entry and information management, while the system administrator is responsible for the system functioning. Improved health care, user satisfaction, and cost savings were the most important benefits gained with this system. The need of similar systems in oncology is crucial and could involve additional applications, such as quality of life (QoL) systems.

INTRODUCTION

Health Informatics is one of the fastest growing areas of information and communication technology (ICT) (Eysenbach, 2000). It is a multifaceted field concerned with electronic patient records, image processing, computer aided diagnosis, research support, database archival, and hospital

Figure 1. Unit's computerized system



management. Despite the remarkable barriers to adoption there are significant technical, legal, economical, professional, and cultural reasons for the evolution of such systems. In parallel, their use should promote and must certainly not be in conflict with the fundamental principles of medical ethics.

In the clinical work processes, the handling of huge amounts of information is a very important issue. Generally, health care information systems are used to guarantee quality and efficiency of the medical practice. The structured analysis and communication of clinical information are necessary in specific areas of health care (Asp & Petersen, 2003). This is extremely important in the case of caring for cancer patients. The cooperative care for cancer patients (declared as "Shared Care") requires complete, distributed, and summarized clinical records as cancer documentation (Blobel, 2000). Computer-based applications assist crossing specialty boundaries involving members of a multidisciplinary team in an oncology unit (Benghiat, Saunders, & Steele, 1999). Moreover, computer systems support collaboration between patients and health care providers in the area of symptom management (Goldsmith, McDermott & Safran, 2004).

One of the most widely discussed issues among health informatics professionals is the Electronic Patient Record (EPR), also referred to as Electronic Health Record (EHR). EPR is an indicator of the progress in health informatics domain and allows health providers, patients, and payers to interact more efficiently and in life-enhancing ways. It offers new methods of storing, manipulating, and communicating various types of medical information, and is thus characterized as more powerful and flexible compared to paper-based systems. EPR systems are not usually stand-alone, but enhance a variety of add-on components according to the specific requirements of each environment.

In the past, many patient record systems were developed to provide oncology staff with a key infrastructure requirement in information management that is essential to maintain efficient and effective health care (Chamorro, 2000).

In order to embody computers in daily clinical practise and provide qualitative health care to cancer patients, we have developed a computerized system. It consists of three modules: the EPR Module, the Image Archive Module, and the Lab Module. These modules are tightly integrated through the EPR Module that works as the main platform of the system. The system administrator is responsible for the secure system operation and for the users training as well. The system and the users involved are briefly presented in Figure 1.

SYSTEM MODULES

EPR Module

The core of this computerized system is the Electronic Patient Record Module (EPR Module). EPR Module is a relational database, which is accessed by the end users through a friendly interface. EPR data include clinical results (laboratory results, clinical outcome) and physician/patient care orders (orders, requisitions, consultations). It also includes administrative data such as admitting diagnosis, patient location, and follow-up appointments. All EPR data are linked to a unique Patient Record Number (PRN). Data are captured at all stages of patient care, from diagnosis to treatment and follow-up. The EPR Module is used further to produce reports, organize data, and provide immediate feedback.

Data are retrieved from the database using various forms. First of all, there is a welcome screen with a few command buttons that allow users to navigate to the other forms. An example is depicted in Figure 2. In this form, the user is able to navigate through a variety of patients' data (e.g., diagnosis, type of diagnosis) and view the other EPR forms (e.g., "Therapies" form) by clicking on the corresponding button. Additionally, there are search and printing options available.

The Unit secretary is responsible for data entry, under the guidance of the supervising doctor. The data source is the existing paper-based record, which is kept in the Oncology Unit. As shown in Figure 2, the data amount imported in each form is small enough and the included information is well organized. The main idea is that the EPR Module should not be a copy of the paper-based record but a summarized description of patient's status.

Figure 2. Part of EPR module. User is able to enter various data (personal data, diagnosis, type of diagnosis, symptoms, etc.) through a friendly interface.

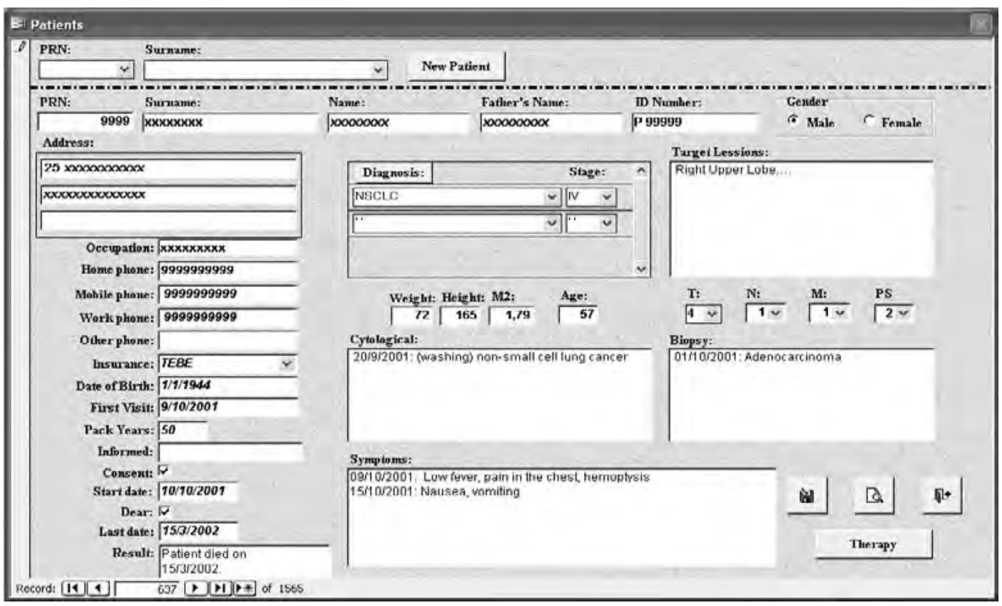


Image Archive Module

The need that led to the development of the Image Archive Module was the reduction in the volume of images, in order to free space in the Unit's archive. This module archives two categories of images, the digital and the film-based ones.

The first category consists of digital images (computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET)) that patients deliver to the Unit immediately after their acquisition. Because there is no need for further process, these images are stored in the archive at once. The film-based have to be digitized and then imported in the Archive Module. For this task a film digitizer (Kodak LS40) is used. The images of each patient are stored according to his PRN and can be accessed through the EPR Module.

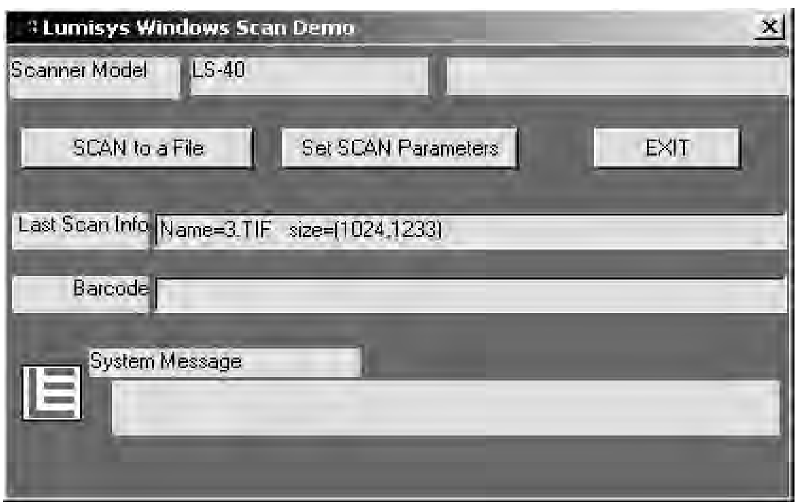
The software program that controls the digitization process is presented in Figure 3. The user sets the scan parameters (image analysis, image format, etc.), selects the archive file, and scans the films. In case of a problem during the scan process, the system produces an error message and the process is interrupted. A success message is produced only after the successful accomplishment of all the appropriate sequences required of an image digitization process. This program produces images in either DICOM or TIFF format. In our study, DICOM format was used. The DICOM images of each patient are further integrated into the EPR Module.

Lab Module

The Lab Module is a simple relational database that stores information about the patient blood samples that are kept in the laboratory. Additionally, it stores data about the patient chemotherapies (cycle, disease, etc.) that need to be kept along with each blood sample. As in the other two modules, all records are stored according to PRN and can be accessed through EPR. Lab Module is very useful, not only for the biologists, but also for the physicians who are in charge of research projects.

The user is able to perform simple and advanced searches for samples, depending on predefined or chosen criteria. Moreover, various reports are produced according to the given criteria.

Figure 3. Image digitizing. User sets the parameters and then scans the film-based images.



Users

Key physicians, biologists, and administrative personnel are involved in data entry and data management in all modules.

Specially trained administrative personnel extract data from paper-based records and subsequently insert them into the EPR Module. Data entry is performed on a daily basis; when a patient leaves the Oncology Unit, his personal record (paper and electronic) is updated the same day or the day after. Physicians verify the extracted data and control the whole process for its reliability. Regarding the Image Archive Module, secretaries perform image digitization using the Kodak LS40 film digitizer. They simply pass the films to the digitizer and save the image files in the archive. Finally, biologists are responsible for data entry in the Lab Module.

The system administrator who is a health informatics specialist is responsible for the flawless functioning of the entire system and is also confronted with the training of the staff.

Security Strategies

Every threat that violates the security of the health care data is carefully controlled using a set of security strategies. The reported strategies could be summarized as follows:

- All modules are password protected.
- There are various levels of authorization corresponding to each kind of user; for example, biologists are not made aware of patient names because all patient data are linked to the PRN.
- The user login and interventions to the system are recorded.

Apart from these strategies, additional actions are taken for other security aspects. The recorded data are backed up at the end of each week and the equipment is locked when it is not in use.

DISCUSSION

In the oncology field, computers may offer great help in data recording and in supporting the clinicians during the therapy process. The collected data are suitable not only for patient care, but also for quality assessment, research, and management. In order to achieve these goals, the best data management should be assured. This need is quite obvious in an Oncology Unit, which is a demanding environment with continuous information overflow and increased pressure for immediate decisions.

Apart from EPR systems, many other computer applications have been developed in the oncology field. A pilot study provided Internet lessons to oncology patients and family members to contribute in a timely fashion to the well being of the patients (Edgar, Greenberg, & Remmer, 2002). Furthermore, various programs attempted either to establish clinical screening programs for the quality of life of cancer patients (Carlson & Bultz, 2003), or to report their psychosocial functioning (Allenby, Matthews, Beresford & McLachlan, 2002).

There are many reasons for using computerized systems in oncology that mainly deal with the quality and efficiency of care. First of all, when using a computer system, voluminous amounts of easily retrievable and analyzed data can be stored. The access to structured and well-organized data offers the health care professionals a better insight into the patient's condition, which leads to improved treatments. Moreover, the data collected in a computerized system can be easily used for research purposes.

However, the paper-based data management is still a reality in clinical practice. Many hospitals retain paper-based records as the main source of patient data. This is also the case for Sotiria General Hospital, where the Oncology Unit is established. As a result and in order to assure the completeness and accuracy of the EPR Module, many data are entered from these paper-based records.

It is true that the modern radiological equipment yields images in digital form, but the time of film-based radiology has not yet come to an end. Additionally, although the application of PACS (Picture Archiving and Communication Systems) is spreading, it is not always possible to access them. Moreover, image procedures (CT, MRI, PET) do not take place at the same radiologist location for all patients. Usually, each patient carries his images when visiting a hospital, then delivers them to his physician and, finally, the images are stored at the clinic's archive.

The situation described above causes the enlargement of image volume and leads to storing and accessing problems. Even though the Oncology Unit had a well-organized film-based image archive, we faced these two problems in a very short period of time. Consequently, the Image Archive Module was developed for two purposes: the reduction of images volume and the ease of access to stored images. Because all images are stored in the Image Archive Module, there is no reason to keep them in both formats (digital and film-based). After their storage, the patient receives the original images and keeps them in his personal record. The

laboratory is one of the first areas in health care where computers were introduced. Computers support the entire process, from sample collection to final report generation and validation. Apart from this process, oncology laboratories perform additional tasks. The most important task is the storage of patient samples at low temperatures (-80 ^{o}C or -20 ^{o}C) for future analysis. The numerous samples have to be well-organized in order to avoid changing or losing them. The Lab Module maintains data about the patient blood samples that are stored in the laboratory. The basic advantage of this module is that it offers the possibility for immediate access to specific samples, by using the search facilities of the module. Moreover, the user is able to categorize samples according to patients' disease or other specific research-targeted criteria.

At this point, it should be mentioned that all users are able to access the modules through the local network. Even though each module serves different purposes, they all work in one. More specifically, through the EPR Module, the user is able to access each patient's images and blood sample information.

Prior to this system development, most of the hospital employees did not have the opportunity to involve computers in their daily practice. Considering this, we used existing technologies to help staff members and offer them the possibility to work and interact with a system based on the Unit's specific needs. Naturally, at the beginning several staff members faced difficulties in using the system, because they were not familiar with computers. This problem was resolved after the daily training and the systematic use of the system modules.

Moreover, user satisfaction is an important aspect and is highly correlated to the system layout (Sittig, Kuerman & Fiskio, 1999). The screen design was based on simple and user-friendly methods governed by the principle that users had to type as less as possible. In order to measure user satisfaction, an evaluation method should be followed, such as the distribution of a well-tested questionnaire. Nevertheless, we decided not to perform a thorough analysis of users' satisfaction, but to briefly evaluate it through the daily practice and the reported problems.

The widespread application of computers increases the accessibility of data, especially when various techniques are applied. Unauthorized use of data is not an illusory danger. Therefore, the access to these data must be regulated. The fact that only authorized users should have access to the data is a logical consequence of the right to privacy. The privacy sensitivity of data is strongly context dependent. Data on various diseases are often considered to be extremely sensitive and oncology data are categorized as such. In our case, data security and patient confidentiality policies were followed to assure the maximum level of system reliability.

In order to reduce the threats concerning data confidentiality, integrity, and availability, some basic methods were used. Regarding confidentiality, all modules are password-protected. The passwords are changed often by the system administrator and are available only to the involved users. Moreover, the equipment (computers, scanner) is under close surveillance and locked when not used, in order to assure physical security. Concerning integrity, data are regularly backed up at the end of the week. Doubtless, database modules enhance various security mechanisms for referential, logical, and entity integrity. For example, each patient receives a unique Patient Record Number (PRN) to avoid duplicate values.

The patients and their families are not actively involved in the system processes, but their role as a data source remains crucial. It is important to note that they are informed that their data are stored in the Unit's system and that it is possible to request a copy for their own use. The data stored in the system are not used for research purposes or published without the patients' permission. It is important to note that extremely sensitive patient data (e.g., DNA), are not stored in the Unit's system in any format. Any sensitive data used for research purposes (research studies, clinical trials) are collected and stored elsewhere as specified per protocol.

Altogether, oncology is often practiced in a quite demanding environment, but not only in the health sector. Doubtless, there are many hospital departments that could benefit from the experience gained using the system described here. For this purpose, the Oncology Unit has already presented its experience to the local staff at Sotiria Hospital and to the public at Med-e-tel Conference (Luxembourg, 5-7 April 2006). The Unit's future plans include close collaboration with organizations or departments eager to share their vision and exchange information with the Unit.

Regarding images, it is obvious that as cost-effective storage media capacity increases, so will the temptation to store all images digitally. Moreover, the extraction of the appropriate data not only improves health care but also saves time in physicians' daily practice. As a result, the Oncology Unit invests the saved working hours in other activities without spending extra funds. Apparently, the economic impact of the described computerized approach is very important.

CONCLUSION

Despite the remarkable barriers to adoption, the use of computers is growing. A variety of health care applications are developed in order to cover the demands of the new era. It is obvious that the Health Informatics field is growing very quickly and some health care areas, such as oncology, may gain important benefits from the implementation of appropriate computerized and digitized systems.

Using the current version of the developed system as an example, we intend to manage patient data in a more beneficial way. Consequently, the collected data could be used in other applications as well. For example, the digitized images could be processed and studied further. Additionally, a quality of life measurement system with immediate feedback of results to clinicians could be developed, as in the Velikova, Brown, Smith, and Selby (2002) study.

Generally, the field of oncology demands new computerized ways of data management. It is clear that computerized systems may open up unexpected paths to new directions in future practice and research.

ACKNOWLEDGMENT

We would like to thank Valsamo Anagnostou, MD, for her assistance in article revision and the helpful comments of the anonymous reviewers to improve the quality of the original version of this article.

REFERENCES

Allenby, A., Matthews, J., Beresford, J., & McLachlan, S.A. (2002).The application of computer touch-screen technology in screening for psychosocial distress in an ambulatory oncology setting. *European Journal of Cancer Care*, *11*(4), 245-53.

Asp, L., & Petersen, J. (2003). A conceptual model for documentation of clinical information in the EHR. In *Proceedings of the MIE 2003*. IOS Press.

Benghiat, A., Saunders, V., & Steele, W.V. (1999). Computerizing the cancer centre. *Clinical Oncology*, *11*(1), 33-39.

Blobel, B. (2000). Onconet: A secure infrastructure to improve cancer patients' care. *European Journal of Medical Research*, *18; 5*(8), 360-368.

Carlson, L.E., & Bultz, B.D. (2003). Cancer distress screening. Needs, models, and methods. *Journal of Psychosomatic Research*, *55*(5), 403-409.

Chamorro, T. (2001). Computer-based patient record systems. *Seminars in Oncology Nursing, 17*(1), 24-33.

Edgar, L., Greenberg, A. & Remmer. J. (2002). Providing Internet lessons to oncology patients and family members: A shared project. *Psycooncology, 11*(5), 43946.

Eysenbach, G. (2000). Recent advances: Consumer health informatics. *BMJ, 320*, 1713-1716.

Goldsmith, D., McDermott, D., & Safran, C. (2004). Improving cancer related symptom management with collaborative health care. *Medinfo, 11*, 217-221.

Sittig, D.F., Kuerman, G.J., & Fiskio, J. (1999). Evaluating physician satisfaction regarding user interactions with an electronic medical records system. In *Proceedings of the AMIA Symposium* (pp. 400-404).

Velikova, G., Brown, J.M., Smith, A.B., & Selby, P.J. (2002). Computer-based quality of life questionnaires may contribute to doctor-patient interactions in oncology. *British Journal of Cancer, 7, 86*(1), 51-59.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 2, Issue 3, edited by J. Tan, pp. 31-38, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 3.6
Artificial Intelligence Techniques in Medicine and Healthcare

Rezaul Begg
Victoria University, Melbourne, Australia

INTRODUCTION

Now-a-days, researchers are increasingly looking into new and innovative techniques with the help of information technology to overcome the rapid surge in healthcare costs facing the community. Research undertaken in the past has shown that artificial intelligence (AI) tools and techniques can aid in the diagnosis of disease states and assessment of treatment outcomes. This has been demonstrated in a number of areas, including: help with medical decision support system, classification of heart disease from electrocardiogram (ECG) waveforms, identification of epileptic seizure from electroencephalogram (EEG) signals, ophthalmology to detect glaucoma disease, abnormality in movement pattern (gait) recognition for rehabilitation and potential falls risk minimization, assisting functional electrical stimulation (FES) control in rehabilitation setting of spinal cord injured patients, and clustering of medical images (Begg et al., 2003; Garrett et al., 2003; Masulli et al., 1998; Papadourokis et al., 1998; Silva & Silva, 1998).

Recent developments in information technology and AI tools, particularly in neural networks, fuzzy logic and support vector machines, have provided the necessary support to develop highly efficient automated diagnostic systems. Despite plenty of future challenges, these new advances in AI tools hold much promise for future developments in AI-based approaches in solving medical and health-related problems. This article is organized as follows: Following an overview of major AI techniques, a brief review of some of the applications of AI in healthcare is provided. Future challenges and directions in automated diagnostics are discussed in the summary and conclusion sections.

ARTIFICIAL INTELLIGENCE TECHNIQUES

There have been a number of artificial intelligence (AI) tools developed over the past decade or so (cf., Haykin, 1999; Keckman, 2002). Many of these have found their applications in medical and health-related areas. Commonly applied AI techniques can be listed as:

- Neural networks
- Fuzzy logic
- Support vector machines
- Genetic algorithms
- Hybrid systems

In the following, we give a brief overview of neural networks, fuzzy logic and the relatively new support vector machines.

Neural Networks

Artificial neural networks work much like the human brain and have the ability to learn from training data and store knowledge in the network. In the learning phase, it maps relation between inputs and the corresponding expected outputs. During the learning phase, knowledge is acquired and stored in the network in the form of synaptic weights and biases. This knowledge is used to make future predictions in response to new data or inputs during the testing phase. Usually, the network has one input and one output layer, and one or more hidden layers depending on the complexity of the problem. Learning can be supervised; that is, the network is provided with both the inputs and their desired outputs during the leaning process, or it can be unsupervised or self-organizing learning. There are a number of learning algorithms available (Haykin, 1999), and among them back-propagation learning algorithm is the most widely used. In this method, an error signal based on the difference between network-generated output (g_i) and desired output (d_i) is propagated in the backward direction to adjust the synaptic weights according to the error signal. During the learning process, the aim is to minimize an objective function such as the mean-squared error (E),

$$E = \frac{1}{n}\sum_{i=1}^{n}(d_i - g_i)^2$$

Neural networks are frequently used as diagnostics, and therefore it is important to have good generalization ability, that is, good performance in predicting results in response to unseen data. One limitation of neural networks is the possibility of being stuck in local minima during training rather than converging to the global minimum. To overcome this the network is usually trained several times with random initial weights to avoid converging to the local minima. Neural networks have found the majority of their applications in pattern recognition, time-series prediction, signal processing and financial forecasting.

Fuzzy Logic

Fuzzy sets were introduced by Zadeh (1965), and they deal with imprecise and uncertain information or data. Naturally, this has been found suitable for many medical and health-related problems, as it relates to the way humans think. Since the early work of Zadeh, there has been an exponential rise in the number of scientific papers applying fuzzy sets in biology, medicine and psychology areas (Teodorescu et al., 1998).

Support Vector Machines

Support vector machines are a relatively new machine learning tool and have emerged as a powerful technique for learning from data and solving classification and regression problems. This has been particularly effective for binary classification applications. SVMs originate from Vapnik's statistical learning theory (Vapnik, 1995). SVMs perform by nonlinearly mapping the input data into a high dimensional feature space (by means of a kernel function) and then constructing a linear optimal separating hyperplane by maximizing the margin between the two classes in the feature space.

For m training data with input-output pairs $(y_1,\boldsymbol{x}_1),\ldots,(y_m,\boldsymbol{x}_m)$ where each input data $\boldsymbol{x}_i \in \Re^N$ belongs to a class $y_i \in \{-1,+1\}_{i=1,\ldots,m}$, the decision function for a new data ($\boldsymbol{x}_i$) can be given by the sign of the following function (Gunn, 1998):

$$f(x) = sign(\sum_{i \in SVs} \alpha_i y_i K(x_i, x) + b)$$

where, a_i is a nonnegative Lagrange multiplier corresponding to $\boldsymbol{x}_i$, $K(.)$ is a kernel function and b is the bias. The Lagrange multipliers are obtained as the solution of a convex quadratic programming problem. The data points $\mathbf{x}_i$s corresponding to a_i>0 are called support vectors. Such $\boldsymbol{x}_i$s are the only data points in the training set relevant to classification since the decision surface is expressed in terms of these points alone (support vectors, SV). For linearly separable problems, the number of SVs and the hyperplane are determined by a subset of the training set only. For nonlinearly separable problems, a_i in SVs are constrained by an upper bound C, which is regarded as a regularization parameter. This parameter makes a trade-off between margin maximization and minimization of classification errors in the training data set (Gunn, 1998).

Hybrid Systems

Recently, researchers have started looking into ways of combining various AI tools in order to maximize performance of the AI system. The main idea behind this is to offset limitation of one system by cascading with another AI tool. As a result, hybrid systems like Neuro-Fuzzy (neural networks and fuzzy logic), Neuro–SVM (neural networks and support vector machines) and Fuzzy-SVM (fuzzy logic and support vector machines) systems have evolved. Hybrid systems have been applied in many applications, including some biomedical areas (Teodorescu et al., 1998).

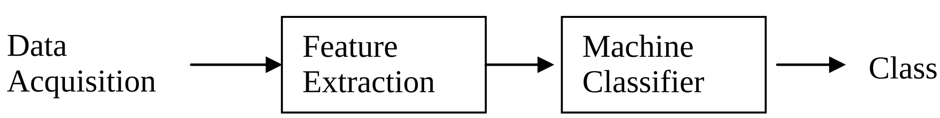
APPLICATIONS IN HEALTHCARE AND MEDICINE

In addition to applications in medical diagnostic systems, AI techniques have been applied in many biomedical signal-processing tasks, including analysis of ECG, EEG and human movement data (Nazeran & Behbehani, 2001). Neural network models have played a dominant role in a majority of these AI-related applications in health and medicine. Many of these applications are for pattern recognition or classification. A typical classification application usually has a number of steps or procedures as shown by the flow diagram (see Figure 1). This involves feature extraction from the input data before feeding these features to the classifier for designing and developing automated classification models, and finally testing the models for generalization.

Medical Decision Support Systems

Medical decision support systems (MDSS) are designed to construct a knowledge database by way of receiving a list of symptoms as input features and their corresponding disease type(s) as the output. Such a developed symptom-to-disease mapping system then facilitates the diagnostic process by generating new responses due to a new set of symptoms. Neural networks have been used to aid MDSS. Silva and Silva (1998) developed such a neural network-based MDSS system for

Figure 1. Stages of a typical pattern recognition task



a relatively small set of 20 randomly selected diseases and reported encouraging results. Disease symptoms in this study were represented as sinusoidal damped waveforms. Hybridization has been shown to improve diagnostic accuracy. For example, Dumitrache et al. (1998) reported an improvement in accuracy (by 28%) for medical diagnosis in cardiac disease using a hybrid decision-making system compared to a classical expert system.

Cardiology

Several studies have applied neural networks in the diagnosis of cardiovascular disease, primarily in the detection and classification of at-risk people from their ECG waveforms (Nazeran & Behbehani, 2001). Celler and Chazal (1998) have applied neural networks to classify normal and abnormal (pathological) ECG waveforms: 500 ECG recordings (155 normal and 345 abnormal) were used to extract features from the QRS complex for training and testing the classifier. The abnormal ECG recordings had six different disease conditions. The classifier was able to recognize these waveforms with 70.9% accuracy.

Electroencephalography

AI tools, including neural networks, fuzzy clustering and SVMs, have been shown to be useful for analyzing electrical activity of the brain, the electroencephalogram (EEG) signals. Features extracted from EEG recordings of the brain have been used with AI tools for improving communication between humans and computers and also for effective diagnosis of brain states and epileptic seizures (Garrett et al., 2003; Geva & Kerem, 1998; Nazeran & Behbehani, 2001).

Opthalmology

Neural networks have been shown to be an effective diagnostic tool to identify glaucoma disease. Glaucoma is more prevalent in older age and can cause loss of vision. Papadourokis et al. (1998) applied backpropagation neural network to classify normal patients and patients with glaucomatic optic nerve damage from perimeter examination. Several neural network models were tested using 715 cases, including 518 glaucoma cases, and they reported 90% recognition accuracy with two hidden layer networks and training with 80% of the input data. In an effort to compare effectiveness of different AI techniques in recognizing glaucoma diagnosis, Chan et al. (2002) used standard automated perimetry data to compare classification performance of several classifiers including multiplayer perceptron and support vector machines (SVM). In-depth analysis of performance of these classifiers was carried out using areas under the receiver operating characteristic (ROC) curves and also sensitivity (true positive rates) and specificity (1 - false positive rates) measures. Machine classifiers were found to perform superiorly in the classification tasks, whereas SVM showed significantly improved performance compared to a multiplayer perceptron. A self-organizing fuzzy structure has also been developed and applied to predict the onset of hypoglycemia for diabetic patients (Hastings et al., 1998).

Gait Analysis and Rehabilitation

Gait is the systematic analysis of human walking. Various instrumentations are available to analyze different aspects of gait (cf. Begg et al., 1989). Among its many applications, gait analysis is being increasingly used to diagnose abnormality in lower limb functions, and also to assess the progress of improvement as a result of treatments and interventions. Recently, neural networks and fuzzy logic techniques have been applied for gait pattern recognition and clustering gait types. Barton and Lees (1997) classified gait patterns based on hip-knee angle diagrams and Holzreiter and Kohle (1993) used neural network models to identify normal and pathological gait patterns from measurements of foot-to-ground reaction forces using force platforms. Wu et al. (1998) applied a back propagation neural network to classify gait patterns of patients with ankle arthrodesis and normal subjects, and reported a superior classification by the neural networks compared to statistical technique (linear discriminant analysis) (98.7% vs. 91.5%). Gait analysis is being increasingly used in rehabilitation settings, and also combining with AI techniques to improve gait control and functionality. Tong and Grant (1998) applied neural networks to optimize sensor sets for control of FES system in people with spinal cord injury and showed improvements in accuracy as a result of neural network aided control. Fuzzy logic has also been recently applied with great success in: clustering children gait with and without neurological disorder (O'Malley, 1997) and also detection of gait events such as the foot contact and take-off during walking in the analysis of paraplegic gait (Skelly & Chizeck, 2001).

Support vector machine (SVM) has recently been applied to classify young and elderly gait patterns (Begg et al., 2003). Gait changes with aging, with potential risks of loss of balance and falls. Recognizing gait patterns with potential falls risks would help to detect at-risk people so that rehabilitation programs could be undertaken to minimize the risk of falls. AI techniques such as neural networks and SVMs have demonstrated their potentials for detecting gait degeneration due to aging and appear to have potential applications in falls prevention in the elderly.

SUMMARY AND FUTURE TRENDS

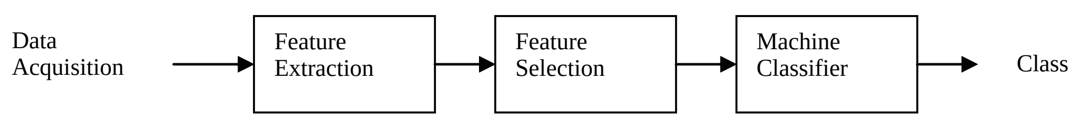
There are plenty of future challenges for AI to be routinely used in medicine and health. The use of automated medical decision support system in routine use in the clinic would make a significant impact on our healthcare system.

One possibility is that healthcare in the future will be built on knowledge networks (Erlandson & Holmer, 2003). Applications of telemedicine and informatics in healthcare can help to provide support to patients in remote areas, and to share expert knowledge or limited resources. Furthermore, effective networking between informatics and biomedical engineering can also help to complement each other's knowledge and to fight in partnership the various challenges faced by the medical and healthcare systems.

One major aim of a classification task is to improve its recognition accuracy or generalization performance. Selecting features that offer the most discriminatory information between the classes or medical groups could help to improve classification accuracy. At the same time, removing redundant features might also improve the medical recognition task. It has been demonstrated in a number of studies that selecting a small number of good features in fact improves the recognition accuracy (Begg et al., 2003; Chan et al., 2002; Yom-Tov & Inbar, 2002), in addition to the added advantages of classifier simplicity and fast processing time. Therefore, a pattern recognition task may be improved by decomposing it into the following stages (see Figure 2).

Recently hybrid systems are emerging, combining various AI tools with improved performance in medical diagnosis and rehabilitation.

Figure 2. Stages of a pattern recognition task involving "feature selection" sub-task



Application of hybrid systems in medical diagnosis has already provided increased efficiency in diagnosis of cardiac disease by as much as 28% compared to classical expert system. This and other applications provide much hope and encouragement for more future applications based on hybridization of AI systems in helping to solve problems in medicine and healthcare.

CONCLUSION

Artificial intelligence, particularly neural networks, fuzzy logic and the recently introduced support vector machines, played a key role over the years for many important developments in medicine and healthcare. Despite such developments there are many future challenges and currently only a few AI-based systems are routinely used in the clinical setting. Continued developments in AI fields are providing much impetus that is needed to tackle the many problems of the healthcare system.

REFERENCES

Barton, J.G., & Lees, A. (1997). An application of neural networks for distinguishing gait patterns on the basis of hip-knee joint angle diagrams. *Gait & Posture, 5,* 28-33.

Begg, R.K., Kamruzzaman, J., & Zayegh, A. (2003). An application of support vector machines for recognizing young-old gait patterns. *Proceedings of the World Congress on Medical Physics and Biomedical Engineering (WC2003*), Sydney, Australia.

Begg, R.K., Wytch, R., & Major, R.E. (1989). Instrumentation used in clinical gait analyses: A review. *Journal of Medical Engineering & Technology, 5,* 115-121.

Celler, B.G., & Chazal, P. (1998). Low computational cost classifiers for ECG diagnosis using neural networks. *Proceedings of the International Conference of Engineering in Medicine & Biology Society (EMBC 1998)* (pp. 1337-1340).

Chan, K., Lee, T.W., Sample, P.A., Goldbaum, M.H., Weinreb, R.N., & Sejnowski, T.J. (2002). Comparison of machine learning and traditional classifiers in glaucoma diagnosis. *IEEE Transaction on Biomedical Engineering, 49*(9), 963-974.

Dumitrache, I., Mihu, I.R., & Rosu, K. (1998). Hybrid decision making system for medical diagnosis. In E.C. Ifeachor, A. Sperduti & A. Starita (Eds.), *Neural networks and expert systems in medicine and healthcare* (pp. 279-285). World Scientific Publishing.

Erlandson, B.E., & Holmer, N.G. (2003). Collaborative multimedia medical technologies in the management of healthcare. *World Congress on Medical Physics and Biomedical Engineering (WC2003*), Sydney, Australia.

Garrett, D., Peterson, D.A., Anderson, C.W., & Thaur, M.H. (2003). Comparison of linear, non-linear and feature selection methods for EEG signal classification. *IEEE Transactions on Neural Systems & Rehabilitation Engineering, 11,* 141-147.

Geva & Kerem. (1998). Brain state identification and forecasting of acute pathology using unsupervised fuzzy clustering of EEG temporal patterns. In T. Teodorrescu, A. Kandel & L.C. Jain (Eds.), *Fuzzy and neuro-fuzzy systems in medicine* (pp. 57-93). CRC Press.

Gunn, S. (1998). *Support vector machines for classification and regression.* ISIS Technical Report. University of Southampton, UK.

Hastings, G, Ghevondian, N., & Nguyen, H. (1998). A self-organizing fuzzy estimator for hypoglycaemia monitoring in diabetic patients.

Proceedings of the International Conference of Engineering in Medicine & Biology Society (EMBC 1998) (pp. 1371-1374).

Haykin, S. (1999). *Neural networks: A comprehensive foundation* (2nd ed). Englewood Cliffs, NJ: Prentice Hall.

Holzreiter, S.H., & Kohle, M.E. (1993). Assessment of gait pattern using neural networks. *Journal of Biomechanics, 26*, 645-651.

Kecman, V. (2002). *Learning and soft computing: support vector machines, neural networks and fuzzy logic models.* NJ: IEEE MIT Press.

Masulli, F., Schenone, A., & Massone, A.M. (1998). Application of the probabilistic clustering to the segmentation of multimodal medical images. In E.C. Ifeachor, A. Sperduti & A. Starita (Eds.), *Neural networks and expert systems in medicine and healthcare* (pp. 241-249). World Scientific Publishing.

Nazeran, H., & Behbehani, K. (2001). Neural networks in processing and analysis of biomedical signals. In M. Akay (Ed.), *Nonlinear biomedical signal processing: Fuzzy logic, neural networks and new algorithms* (pp. 69-97).

O'Malley, M.J., Abel, M.F., Damiano, D.L., & Vaughan, C.L. (1997). Fuzzy clustering of children with cerebral palsy based on temporal-distance gait parameters. *IEEE Transactions on Rehabilitation Engineering, 5*, 300-309.

Papadourakis, G.M., Gross, H.G., & Alexakis, I. (1998). Detection of glaucomatic nerve damage using artificial neural networks. In E.C. Ifeachor, A. Sperduti & A. Starita (Eds.), *Neural networks and expert systems in medicine and healthcare* (pp. 140-150). World Scientific Publishing.

Silva, R., & Silva, A.C.R. (1998). Medical diagnosis as a neural networks pattern classification problem. In E.C. Ifeachor, A. Sperduti & A. Starita (Eds.), *Neural networks and expert systems in medicine and healthcare* (pp. 25-33). World Scientific Publishing.

Skelly, M.M., & Chizeck, H.J. (2001). Real-time gait event detection for paraplegic FES walking. *IEEE Transactions on Neural Systems & Rehabilitation Engineering, 9,* 59-68.

Teodorrescu, T., Kandel, A., & Jain, L.C. (1998). *Fuzzy and neuro-fuzzy systems in medicine.* In Teodorrescu, Kandel & Jain (Eds.). CRC Press.

Vapnik, V.N. (1995). *The nature of statistical learning theory.* New York: Springer.

Wu, W.L., Su, F.C., & Chou, C.K. (1998). Potential of the back propagation neural networks in the assessment of gait patterns in ankle arthrodesis. In E.C. Ifeachor, A. Sperduti & A. Starita (Eds.), *Neural networks and expert systems in medicine and healthcare* (pp. 92-100). World Scientific Publishing.

Yom-Tov, E., & Inbar, G.F. (2002). Feature selection for the classification of movements from single movement-related potentials. *IEEE Transactions on Neural Systems & Rehabilitation Engineering, 10,* 170-177.

Zadeh, L. (1965). Fuzzy sets. *Information and Control, 8,* 338-353.

KEY TERMS

ECG: The electrical activity of the heart recorded from the body surface using electrodes as electrical potential is known as electrocardiogram or ECG.

EEG: The electrical activity of the brain recorded from the scalp as electrical potential is known as electroencephalogram or EEG.

Fuzzy Logic: The concept of fuzzy logic is that many classes in the natural environment are fuzzy rather than crisp. It deals with imprecise and uncertain data.

Gait Analysis: Analysis of human walking patterns. It is used to analyze abnormality in lower limb problems and assess treatment or intervention outcomes.

Hybrid Systems: Integration of two or more artificial intelligence tools to improve efficiency or system performance.

Neural Networks: Neural networks resemble the human brain and able to store knowledge during training and use this for decision making during testing phase.

Support Vector Machines: Introduced by Vapnik and capable of learning from data for solving classification and regression problems.

This work was previously published in Encyclopedia of Information Science and Technology, Vol. 1, edited by Mehdi Khosrow-Pour, pp. 157-162, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.7
PDA Usability for Telemedicine Support*

Shirley Ann Becker
Florida Institute of Technology, USA

INTRODUCTION

Telemedicine is broadly defined as the use of information and communications technology to provide medical information and services (Perednia & Allen, 1995). Telemedicine offers an unprecedented means of bringing healthcare to anyone regardless of geographic remoteness. It promotes the use of ICT for healthcare when physical distance separates the provider from the patient (Institute of Medicine, 1996). In addition, it provides for real-time feedback, thus eliminating the waiting time associated with a traditional healthcare visit.

Telemedicine has been pursued for over three decades as researchers, healthcare providers, and clinicians search for a way to reach patients living in remote and isolated areas (Norris, 2001). Early implementation of telemedicine made use of the telephone in order for healthcare providers and patients to interact. Over time, fax machines were introduced along with interactive multimedia, thus supporting teleconferencing among participants. Unfortunately, many of the early telemedicine projects did not survive because of high costs and insurmountable barriers associated with the use of technology.

Telemedicine has been resurrected during the last decade as a means to help rural healthcare facilities. Advances in information and communications technology have initiated partnerships between rural healthcare facilities and larger ones. The Internet in particular has changed the way in which medical consultations can be provided (Coiera, 1997). Personal computers (PCs) and supporting peripherals, acting as clients, can be linked to medical databases residing virtually in any geographic space. Multimedia data types, video, audio, text, imaging, and graphics promote the rapid diagnosis and treatment of casualties and diseases.

Innovations in ICT offer unprecedented healthcare opportunities in remote regions throughout the world. Mobile devices using wireless connectivity are growing in popularity as thin clients that can be linked to centralized or distributed medical-data sources. These devices provide for local data storage of medical data, which can be retrieved and sent back to a centralized source when Internet access becomes available. Those working in nomadic environments are connected to data sources that in the past were inaccessible due to a lack of telephone and cable lines. For the military, paramedics, social workers, and other healthcare providers in the field, ICT advances have removed technology barriers that made mobility difficult if not impossible.

Personal digital assistants (PDAs)[1] are mobile devices that continue to grow in popularity. PDAs are typically considered more usable for multimedia data than smaller wireless devices (e.g., cell phones) because of larger screens, fully functional keyboards, and operating systems that support many desktop features. Over the past several years, PDAs have become far less costly than personal-computing technology. They are portable, lightweight, and mobile when compared to desktop computers. Yet, they offer similar functionality scaled back to accommodate the differences in user-interface designs, data transmission speed, memory, processing power, data storage capacity, and battery life.

BACKGROUND

Computing experts predicted that PDAs would supplant the personal computer as ubiquitous technology (Chen, 1999; Weiser as cited in Kim & Albers, 2001). Though this has not yet happened, PDA usage continues to grow with advances in operating systems, database technology, and add-on features such as digital cameras. They are being used in sales, field engineering, education, healthcare, and other areas that require mobility. In the medical field, for example, they are being used to record and track patient data (Du Bois & McCright, 2000). This mobility is made possible by enterprise servers pushing data onto these devices without user intervention. Enterprise servers are also capable of pulling data from a localized (PDA) database such that centralized data sources are readily updated.

A PDA synchronizes with laptops and desktop computers, making data sharing transparent. This is made possible by a user interface and functionality that are compatible in terms of computing capabilities and input and output devices (Myers, 2001). Compatibility is a major issue in telemedicine given that medical and patient data gathered or stored on a PDA is typically sent to a centralized data source. Nomadic use of PDAs mandates this type of data integration whether it is real-time or batched data when wireless connectivity is temporarily inaccessible (Huston & Huston, 2000). In addition, telemedicine data sharing is typically asymmetric in that the enterprise server transmits a larger volume of medical data to the PDA. In turn, the PDA transmits only a small volume of patient data to the server (Murthy & Krishnamurthy, 2004).

Though PDAs hold great promise in promoting healthcare in remote regions, the usability of these devices continues to be an issue. There are physical constraints that typically do not apply to a laptop or desktop computer (Table 1 describes these constraints). The user interface of a PDA is modeled after a desktop environment with little consideration for physical and environmental differences (Sacher & Loudon, 2002). Yet, these differences are significant in terms of usability given the small screen and keyboard sizes and limited screen resources in terms of memory and power reduction (Brewster, 2002).

There has been important research on PDA usability, primarily in the effective use of its limited screen area. Early research focused primarily on the display of contextual information in order to minimize waste of the screen space while maximizing content (Kamba, Elson, Harpold, Stamper, & Sukariya as cited in Buchanan, Farrant, Jones, Thimbleby, Marsden, & Pazzani, 2001). More recent efforts are taking into account not only screen size, but navigation, download time, scrolling, and input mechanisms (Kaikkonen & Roto, 2003).

Table 1. User-interface design constraints for PDA devices (Paelke, Reimann, & Rosenbach, 2003)

Limited resolution	Typical resolution of a PDA is low (240*320 pixels). This impacts the visibility of content, objects and images.
Small display size	The small screen size of a PDA limits the number of objects and the amount of text on a screen page. This limitation impacts design layout in terms of font size, white space, links, text, images, and graphics, among others.
Navigational structure	Navigation is impacted by the increased number of screen pages required to accommodate text and objects that on a desktop or laptop would fit on one screen page. Design choices include a long page with a flat navigation hierarchy versus the design of multiple short pages with a deeper navigational hierarchy.
Limited use of color	A PDA uses a gray scale or a color palette limited to several thousand color choices (compared to millions of color choices for desktop applications). Readability and comprehension may be impacted when color is used to relay information or color combinations are insufficient in contrast.
Limited processing power	Limited processing power impacts the quality of graphical displays and imaging. It also restricts the use of interactive real-time animation.
Mouse is replaced with stylus pen	A PDA does not use a mouse, which has become a standard peripheral in a desktop environment. As a result, there is a learning curve associated with the use of a stylus pen, which replaces mouse functionality.
Small keyboard size	The PDA keyboard size and layout impacts data entry. As a result, it is more difficult for users to entered lengthy and complex medical data in a real-time environment.

PDA USABILITY AND TELEMEDICINE

An important finding of usability research associated with mobile technology is the need for usability testing beyond a simulated environment. Waterson, Landay, and Matthews (2002), in their study of the usability of a PDA, found that usability testing should include both content and device design. Chittaro and Dal Cin (2002) studied the navigational structures of mobile user interfaces. Their research also identified the need for actual devices to be used in usability testing. Real-world constraints would take into account screen-size and page-design issues, date entry using a built-in keypad, wireless accessibility, data transmission speeds, visual glare, background noise, and battery power, among others.

Our initial findings also reflected the need for usability testing in the telemedical environment in which technology is used. We initiated research on the use of PDAs for monitoring diabetic patients living in remote regions of the United States (Becker, Sugumaran, & Pannu, 2004). Figure 1 illustrates one of our user-interface designs for the Viewsonic® PocketPC. This screen shows part of a foot form completed by a healthcare provider during a home visit. The data entered by the user is stored in a local database that can be transmitted wirelessly to an enterprise server.

The PocketPC is used in this research because of its low cost and its support of relational database technology. It has a built-in digital camera, which is important because of the physical distance between a patient and a healthcare facility. Images of foot sores are taken during a home visit and stored in the local database residing on the PDA. These images become part of the patient's history when transmitted to the enterprise server. Later, the images can be viewed by a clinician for the timely diagnosis and treatment of the sores.

Our research has shown the technical feasibility of using PDA technology to gather data in the

Figure 1. PDA used to gather data about a diabetic patient's foot health

field during a home visit. However, more research is needed to address usability issues uncovered during the use of the PDA during simulated home visits. A significant finding in the use of PDA technology is that usability is tightly integrated with the technological challenges associated with it. One such challenge is the heavy reliance on battery power when PDAs are deployed in the field. When the battery no longer holds a charge, critically stored relational data may be irretrievable due to pull technology used to transmit data from a local source to a central one.

As part of this research, the usability of multimedia data formats is being studied to improve information access in a nomadic environment. For rapid diagnosis and treatment of casualties, multimedia data formats may prove critical. In our work, images are being used to replace textual descriptions that would consume valuable screen space. Figure 1 illustrates this concept of using color codes to represent physical areas of the foot. As such, foot problems can be reported for each area by clicking on the list appearing on the right side of the screen. Audio capabilities are also being explored in providing helpful information that otherwise would be text based. Both of these multimedia capabilities are in the design phase and will be tested in future field studies.

FUTURE TRENDS

Table 2 identifies research opportunities associated with the use of PDAs in telemedicine. Much of what has been done in this area has focused on tracking patient histories. However, there are significant opportunities for real-time data retrieval and transmission using PDA technology. Clinicians could use a PDA, for example, to send prescriptions to pharmacies, receive lab reports, and review medical data for the diagnosis and treatment of patients. These devices could also be used to minimize human error associated with more traditional mechanisms of recording patient data.

There are infrastructure challenges associated with the use of telemedicine in terms of technology acceptance and utilization. Chau and Hu (2004) point out that although telemedicine is experiencing rapid growth, there are organizational issues pertaining to technology and management. It is critical that organizational support is available throughout the implementation stages of tele-

Table 2. Telemedicine research opportunities using PDA technology (Wachter, 2003)

Diagnosis and Treatment	Mobile decision support software would allow for data entry of patient symptoms with output providing a diagnosis and treatment plan.
Patient Tracking	Synchronizing a PDA with a hospital's centralized data source would allow vital signs and other information to be gathered in real-time at the point of care. A clinician would have the capability of obtaining lab reports and test results once they have been entered into the system.
Prescriptions	A PDA would be used by a clinician to send a patient prescription to a pharmacy. This would minimize human error associated with interpreting handwritten prescriptions. It would also provide a centralized tracking system in order to identify drug interactions when multiple prescriptions for a patient are filled.
Medical Information	Clinicians would have access to information on medical research, drug treatments, treatment protocols, and other supporting materials. According to Wachter (2003), a leading clinical PDA technology vendor has converted more than 260 medical texts into PDA formats thus supporting this effort.
Dictation	PDAs support multimedia data including audio, images, and text. As such, clinicians would have an opportunity to record multimedia patient data directly linked to patient history data in a centralized source.
Charge Capture	Data entry into a PDA that is transmitted to a centralized source would provide the means for efficient billing of medical charges to a patient.

medicine. Past experience in the use of ICT with no infrastructural support resulted in failure. The effective management of telemedicine systems and supporting technologies is needed to address barriers to ICT acceptance by healthcare personnel and patients. As such, there are research opportunities in the organizational acceptance and use of PDAs in a telemedical environment.

Security, safety, and social concerns have also been identified by Tarasewich (2003) as research challenges in the use of mobile technology. Though encryption and other security technologies can readily be used during the transmission of data, there remains the issue of security associated with lost or stolen PDAs. Given the memory, data storage, and other technological constraints of a PDA, research is needed on developing security mechanisms for localized data. Research is also needed on ensuring localized data remains private and is accessible only by authorized personnel.

CONCLUSION

The exponential growth of wireless and PDA technologies has brought unprecedented opportunities in providing managed healthcare. For the military and others working in nomadic environments, PDA technology offers the capability for rapid diagnosis and treatment of casualties. Regardless of location, healthcare personnel could be provided with real-time access to reference materials, patient lab reports, and patient history data.

Though there is great promise in the use of PDAs for providing telemedical services, there is research needed in the usability of these devices. Multimedia data formats offer alternative interfaces to accessing data, and research is needed to assess their impact on ease of use and understandability. In addition, technological constraints need to be studied in terms of their impact on device usability. Memory, data storage, transmission speeds, and battery life need to be considered as part of usability testing to assess the impact on rapid medical diagnosis and treatment.

There is a major challenge of moving from traditional medical services and resources to an environment that promotes PDA technology and telemedicine. The potential benefits are great in terms of ubiquitous helth care with no time or space constraints. However, widespread acceptance of PDA technology in a telemedical environment will only become achievable through the design of usable interfaces.

REFERENCES

Becker, S. A., Sugumaran, R., & Pannu, K. (2004). *The use of mobile technology for proactive healthcare in tribal communities.* 2004 National Conference on Digital Government Research (pp. 297-298).

Brewster, S. (2002). Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing*, *6*, 188-205.

Buchanan, G., Farrant, S., Jones, M., Thimbleby, H., Marsden, G., & Pazzani, M. (2001). Improving mobile Internet usability. *Proceedings of the Tenth International Conference on World Wide Web* (pp. 673-680).

Chau, P. Y. K., & Hu, J. H. P. (2004). Technology implementation for telemedicine programs. *Communications of the ACM*, *47*(2), 87-92.

Chen, A. (1999, September 27). Handhelds on deck. *InfoWeek*, 67-72.

Chittaro, L., & Dal Cin, P. (2002). Evaluating interface design choices on WAP phones: Navigation and selection. *Personal and Ubiquitous Computing*, *6*(4), 237-244.

Coiera, E. (1997). *Guide to medical informatics, the Internet and telemedicine*. London: Chapman and Hall.

Du Bois, G., & McCright, J. (2000, September 4). Doctors are on the move. *eWeek*, 31.

Huston, T. L., & Huston, J. L. (2000). Is telemedicine a practical reality? *Communications of the ACM*, *43*(6), 91-95.

Institute of Electrical and Electronics Engineers (IEEE). (1990). *IEEE standard computer dictionary: A compilation of IEEE standard computer glossaries.* New York: Institute of Electrical and Electronics Engineers.

Institute of Medicine. (1996). *Telemedicine: A guide to assessing telecommunications in health care*. Washington, DC: National Academy Press.

Kaikkonen, A., & Roto, V. (2003). Navigating in a mobile XHTML application. *Proceedings of the Conference on Human Factors in Computing Systems* (pp. 329-336).

Kamba, T., Elson, S. A., Harpold, T., Stamper, T., & Sukariya, P. (1996). Using small screen space more efficiently. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 383-390).

Kim, L., & Albers, M. J. (2001). Web design issues when searching for information in a small screen display. *Proceedings of the 19th Annual International Conference on Computer Documentation* (pp. 193-200).

Murthy, V. K., & Krishnamurthy, E. (2004). *Multimedia computing environment for telemedical applications.* Poster presentation at the IADIS International Conference on Web Based Communities, Lisbon, Portugal.

Myers, B. A. (2001). Using handhelds and PCs together. *Communications of the ACM*, *44*(11), 34-41.

Norris, A. C. (2001). *Essentials of telemedicine and telecare.* New York: John Wiley & Sons.

Paelke, V., Reimann, C., & Rosenbach, W. (2003). A visualization design repository for mobile devices. *Proceedings of the Second International Conference on Computer Graphics, Virtual Reality, Visualization and Interaction in Africa* (pp. 57-62).

Perednia, D. A., & Allen, A. (1995). Telemedicine technology and clinical applications. *Journal* *of the American Medical Association, 273*(6), 483-487.

Sacher, H., & Loudon, G. (2002). Uncovering the new wireless interaction paradigm. *Interactions, 9*(1), 17-23.

Tarasewich, P. (2003). Designing mobile commerce applications. *Communications of the ACM, 46*(12), 57-60.

Wachter, G. W. (2003). Making rounds with handheld technology. *Telehealth Practices Report, 8*(5), 7-9.

Waterson, S., Landay, J. A., & Matthews, T. (2002). In the lab and out in the wild: Remote Web usability testing for mobile devices. *Conference on Human Factors in Computing Systems* (pp. 796-797).

Weiser, M. (1998). The future of ubiquitous computing on campus. *Communications of the ACM, 41*(1), 41-42.

KEY TERMS

Compatibility: The ability to transmit data from one source to another without losses or modifications to the data or additional programming requirements.

Interoperability: The ability of two or more systems or components to exchange information and to use the information that has been exchanged (Institute of Electrical and Electronics Engineers [IEEE], 1990).

Peripheral Devices: Hardware devices, separate from the computer's central processing unit (CPU), which add communication or other capabilities to the computer.

Personal Digital Assistant (PDA): A personal digital assistant is a handheld device that integrates computing, telephone, Internet, and networking technologies.

Telecare: The use of information and communications technology to provide medical services and resources directly to a patient in his or her home.

Telehealth: The use of information and communications technologies to provide a broader set of healthcare services including medical, clinical, administrative, and educational ones.

Telemedicine: The use of information and communications technologies to provide medical services and resources.

Wireless Application Protocol (WAP): The wireless application protocol promotes the interoperability of wireless networks, supporting devices, and applications by using a common set of applications and protocols (http://www.wapforum.org).

ENDNOTES

* This article is based on work supported by the National Science Foundation under Grant No. 0443599. Any opinions, findings, and conclusions or recommendations expressed in this content are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

1 PocketPCs, Palm Pilots, and other handheld devices are referred to as PDAs in this article.

This work was previously published in Encyclopedia of Human Computer Interaction, edited by C. Ghaoui, pp. 457-462, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.8
Nurses' Perceptions of Using a Pocket PC for Shift Reports and Patient Care

Karen Chang
Purdue University, USA

Kyle D. Lutes
Purdue University, USA

Melanie L. Braswell
Purdue University, USA

Jacqueline K. Nielsen
Purdue University, USA

ABSTRACT

Nurses working in hospitals with paper-based systems often face the challenge of inefficiency in providing quality nursing care. Two areas of inefficiency are shift-to-shift communication among nurses and access to information related to patient care. An integrated IT system, consisting of Pocket PCs and a desktop PC interfaced to a hospital's mainframe system, was developed. The goal was to use mobile IT to give nurses easier access to patient information. This paper describes the development of this system and reports the results of a pilot study: a comparison of time spent in taking and giving shift reports before and after the study and nurses' perceptions of the mobile IT system. Results showed significant difference in taking shift reports and no significant difference in giving shift reports. Nurses stated that quick and easy access to updated patient information in the Pocket PC was very helpful, especially during mainframe downtime.

INTRODUCTION AND BACKGROUND

The quality of the American healthcare delivery system has been problematic. The Institute of Medicine (IOM) (2001) identified six dimensions of quality: safe, effective, patient-centered, timely, efficient, and equitable. However, in hospitals with paper-based delivery systems, nurses often face challenges in meeting these expectations. Two areas of inefficiency are communication among nurses and accessing information relevant to patient care.

Paper-based shift reports and patient management can be time-consuming. At the beginning of each shift, nurses take reports of their patients. They use self-designed paper worksheets to write key patient information (e.g., diagnosis, physical or psychosocial status, treatments, tests, etc.) and to organize patient care activities. Audiotaped reports, face-to-face reports, or walking rounds are commonly used for nurses to pass patient information from shift to shift. Face-to-face reports and walking rounds require more time than audiotaped reports. For face-to-face and walking rounds, outgoing nurses and incoming nurses need to be present at the same time and additional communication occurs when both groups are present. Audiotape reports do not require the presence of outgoing nurses and incoming nurses simultaneously. Outgoing nurses give reports by speaking into an audiotape recorder about one hour before the end of the shift. They give reports based on their memory and notes written on self-designed worksheets. Incoming nurses listen to the audiotape reports at the beginning of the shift and take notes on their self-designed paper worksheets. Since the 1990s, audiotaped reports have become prevalent and save about 15 minutes in each shift because outgoing nurses do not need to wait for incoming nurses to give reports (Mason, 2004). However, incoming nurses spend about 30 to 60 minutes taking shift reports from the audiotape reports before starting to take care of their assigned patients. The majority of the time is spent on transcribing key information from audiotape reports to their self-designed paper worksheets. Audiotaped reports continue to limit nurses in providing care in a timely and efficient manner.

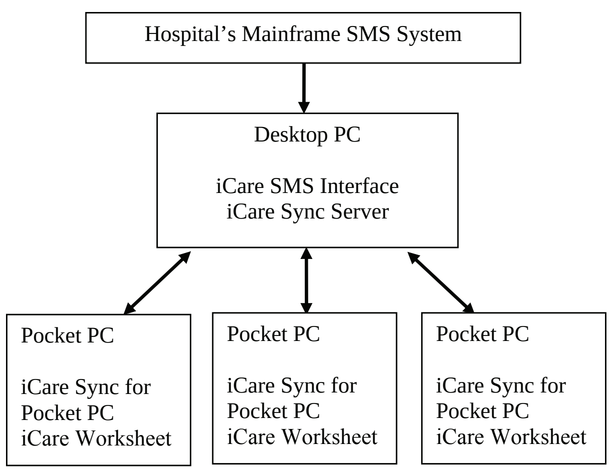
Nurses need to efficiently and effectively communicate information to provide safe, effective, and patient-centered care. When comprehensive patient information is transferred efficiently, nurses can identify patient needs, monitor patient conditions, prevent or detect complications, and implement physician orders safely and accurately. However, accessing all relevant information for patient care in the paper-based delivery system can be difficult and time-consuming. For example, when nurses need to know patient health history and health status, they need to find the patient's paper medical records, which could be misplaced or in use by other healthcare providers. When nurses need to know the results of a patient's most recent laboratory or diagnostic tests that are stored in the hospital's mainframe system, they need to go to a nursing station to log in to the mainframe to obtain the information. When nurses need to give unfamiliar medications, they need to look up the information in a drug reference, which could be misplaced, in use by other nurses, or outdated. If nurses give medications without accurate knowledge of drug information, such as safe dosage, rate of administration, drug interactions, and side effects, medication errors may occur. Thus, paper-based information systems are fraught with inefficiencies that can compromise patient care. Specifically, if nurses do not have pertinent information, they may not detect or prevent complications due to medications, procedures, or treatments or may not provide adequate patient education to help patients care for themselves.

The Institute of Medicine (2001) asserted that information technology (IT) must be used in the 21st century to improve the quality of healthcare. The American Academy of Nursing Technology and Workforce urged the use of IT to support nurses' work and to eliminate waste and redundancy (Sensimeier, Raiford, Taylor, & Weaver, 2002). President Bush announced that electronic medical records (EMR) would be available for residents in the United States in 10 years (iHealth Beat, 2004b). However, there are many barriers to integrating IT in the hospital, such as high cost, lack of support from physicians and administration, and lack of applications that are easy to use (Gillespie, 2002; iHealth Beat, 2004a; Sensimeier et al., 2002). Thus, many hospitals are still using paper-based medical records and mainframe computer systems to store patient information.

To identify studies using Personal Digital Assistants (PDAs) integrated with the hospital mainframe system for nurses to use in shift reports, an extensive search of MEDLINE and CINAHL databases was performed using "Computers, Handheld" as the key word in combination with other key words, such as "Database Management Systems, Hospital Information Systems, Hospital Nursing Staff, or Nursing Information Systems". There is an increasing trend in using PDAs in the healthcare field. For examples, orthopedic surgeons used programs loaded in the PDAs to access patient data in the office and the operating room (Laskin & Davis, 2004); a wireless PDA was developed for patient transport to monitor patient's vital signs (heart rate, three-lead electrocardiography, and SpO2) remotely (Lin, Jan, Ko, Chen, Wong, & Jan, 2004); nurses used PDAs to collect utilization review data that were sent to hospital's mainframe system for quality evaluation (Lanway, 2003); and physicians at Cedars-Sinai Hospital in California used wireless PDAs to access patient EMRs wirelessly and used them during ward rounds or at shift changes (Shabot, 2002). The healthcare providers of Shock Trauma Intensive Care Unit at Intermountain Health Care, Salt Lake City, used EMR and reported that they spent 10 minutes for shift reports (Nelson et al., 2003). However, no studies were found regarding nurses' use of PDAs interfaced with the hospital mainframe system for shift reports.

A collaborative effort among Purdue's School of Nursing, Purdue's Department of Computer Technology, and a local hospital's nursing staff resulted in the development of the iCare system to improve the efficiency and effectiveness of shift reports in the paper-based medical records environment. This paper reports the development of the iCare system and nurses' perceptions of the iCare system for shift reports and patient care. The iCare system consists of handheld computers (Pocket PCs), a desktop personal computer (PC), and a hospital's mainframe system (Figure 1). The goal of the iCare system is to use mobile IT, in the form of Pocket PCs, to give nurses easier access to patient information so that nurses have more time to provide quality patient care. Additionally,

Figure 1. The iCare system architecture



an electronic version of a worksheet, similar to nurses' paper worksheets and audiotape reports, was developed to improve the efficiency of communicating patient information between shifts.

METHOD

Design and Security Considerations

By design, Pocket PCs are small, mobile computing devices and can easily be lost or stolen. Privacy of patient information on the Pocket PCs was a serious concern. To protect this information, several layers of security were used. First, password protection was built into the Pocket PCs, which requires a password each time a Pocket PC was powered on. Second, the custom applications required nurses to log in using a unique user ID and password. And finally, all patient information used by the system was stored in encrypted and obfuscated files on the Pocket PCs and desktop PC (Rosenthal, 2004).

Other considerations were technology-related. Because the hospital did not have a wireless network infrastructure in place, nor would they allow wireless network radios to be installed, the iCare system was not developed to use wireless network access of any kind. Instead, the system required the nurses to send and receive patient information to and from a Pocket PC by periodically synchronizing it with the desktop PC through a sync cable.

In designing the study, a decision was made to limit the impact this research would have on the workload of the hospital IT staff. The project was designed so that no new custom mainframe programs or changes to the existing systems were required. Retrieving patient information from the hospital's mainframe computer system was done using custom software designed by the second author.

System Architecture

After an analysis period and meeting with several of the nursing staff to understand the process of shift reports and the process of accessing the hospital's mainframe system for patient information, four custom software applications were developed. These were named "iCare SMS Interface," "iCare Worksheet for Pocket PCs," "iCare Sync for Pocket PCs," and "iCare Sync Server". An iterative process was used to simulate paper-based worksheets and audiotape shift reports during the development of the iCare worksheet.

A desktop PC was centrally located in a unit of the hospital in which the study was carried out. A program on the central PC periodically accessed patient information from the hospital's mainframe computer systems for all patients in that unit. At the beginning of a shift, each nurse synchronized the Pocket PC by placing it in a cradle connected to the PC. Custom software automatically copied patient information, including data from the mainframe systems and notes from the nurses on previous shifts, to the Pocket PC. During the shift, the nurse recorded new notes about each patient using typed text, handwriting, and/or audio voice recordings. Whenever the nurse wanted to retrieve updated patient information, to send the notes to other nurses, or to give reports at the end of the shift, the nurse connected his or her Pocket PC to the central desktop PC where the software automatically copied any new patient notes.

iCare SMS Interface

The hospital tracked patient information using an information system from Siemens Health Services (SMS) that runs on its mainframe computers. Nurses access this system using dumb terminal hardware, or terminal emulation software running on PCs. To avoid changes to programs running on the mainframe system, the iCare Interface program was written to use a technique called "screen scraping" to gather patient information.

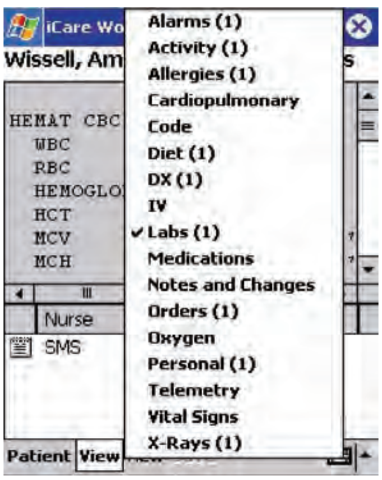
The iCare SMS Interface program functions by communicating with a terminal emulation component running on the same PC. The iCare SMS interface program sends keystrokes to the terminal emulation component that mimics the keystrokes a nurse might use when using the SMS system. When patient information is returned as text to the terminal emulator, the iCare SMS Interface program copies the text from the emulator screen buffer, reformats it, and saves it to encrypted files stored on the PC. Although not efficient by modern IT standards, screen scraping does provide an interface to legacy systems that cannot otherwise be modified to interface with a new system.

The iCare SMS Interface program was developed using Microsoft Visual Basic 6 and a terminal emulation component from Attachmate. These technologies were chosen for this application because it was developed by the authors primarily onsite at the hospital's IT department, and these technologies were already being used by the hospital's IT staff for other projects.

iCare Worksheet for Pocket PCs

The iCare worksheet application is the primary program for nurses to use for shift reports and patient care management. It simulates nurses' self-designed paper worksheets to organize patient information and patient care activities. Nurses can view patient information either gathered from the mainframe system via iCare SMS Interface or recorded notes. They can also record new notes. The user interface allows nurses to see a list of all patients in the unit. After selecting a patient, nurses can see information about the patient. Because of the limited screen space on a Pocket PC, information is grouped into various categories to make finding the information quicker (see Figure 2). The patient information obtained from the mainframe system by the SMS Interface program is automatically sent to the categories of Allergies, Cardiopulmonary (test results), Diet, Lab, Orders (current medications and laboratory orders), Personal (date of admission, room number, physicians), and X-Rays (radiology test results). Nurses can take notes in any of the categories

Figure 2. Categories of the iCare worksheet



Note: Fictitious patient information shown. A number inside of parenthesis indicates the number of notes in that category.

Figure 3. An example of text, audio, and handwriting notes in the iCare worksheet

Note: Fictitious patient information shown. Red, blue, green, and black pen colors can be used when taking handwritten notes.

using either typed text, handwriting, or audio voice recordings (see Figure 3). Alarm features are also built into the iCare worksheet. Nurses can set alarms to remind them to do certain tasks for each patient at a specified time, such as time to give medications or treatments.

This program was written using Microsoft Visual C# .NET and the .NET Compact Framework. These technologies were chosen because the study also included plans to test the effectiveness of Tablet PCs in addition to Pocket PCs for use by nurses. By using Visual C#, the authors hoped that much of the Pocket PC program source code could be ported to the Tablet PC platform easier than other software development tools would allow.

iCare Sync for Pocket PCs and iCare Sync Server

The iCare Sync for Pocket PCs and iCare Sync Server work in tandem to automatically copy patient information to and from the Pocket PCs. When a Pocket PC is connected to the central desktop PC using the sync cable, the iCare Sync for Pocket PCs program running on the Pocket PC establishes a connection to the iCare Sync Server running on the central PC. The list of patient information files on the Pocket PC is compared to the list of patient information files on the central PC. Any note files on the Pocket PC not present on the central PC are copied to the PC. Likewise, any note files on the PC are copied to the Pocket PC.

As with the iCare worksheet application, the iCare Sync programs were developed using Microsoft Visual C# .NET to allow easier porting to the Tablet PC platform.

Study Setting and Participants

This study was conducted in a medical unit of an acute care hospital in the Midwestern United States, where paper-based medical records were used. In addition, key patient information was available using dumb terminals in the nurses' stations. There was no computer access at the bedside. A prospective pre- and post-test design was used to examine the efficiency and effectiveness of the iCare system for shift reports and patient care. Nurses of all shifts who worked during the study period participated in the study.

Procedure

Ten Dell Axim X5 Pocket PCs were used in the study because the maximum number of nurses working at one time on the unit was eight. All Pocket PCs were kept on the unit and maintained by unit secretaries. After the approval from Institutional Review Board, this study was implemented over four weeks (two weeks of learning and two weeks of actual use). In the first week, six one-hour orientation sessions were provided. Nurses were encouraged to practice using the iCare worksheet for shift reports. The purpose and the procedure of the study, confidentiality of collected data, and voluntary nature of participation were explained to participants. After consent was obtained, nurses learned to use Pocket PCs and the iCare worksheet to find patient information, and take notes (text, handwriting, or voice format) in the iCare worksheet. During the second week, nurses continued to practice using the iCare system for shift reports and patient care. In the third week, nurses used both the iCare worksheet and audiotape for all shift reports. In the fourth week, the iCare worksheet was the only means for shift reports and audiotape recorders were removed. Three nursing faculty provided orientation and support for the iCare system. They visited the unit periodically during each shift and talked to every nurse to inquire about concerns or difficulties. They also helped each nurse to use the iCare worksheet.

Questionnaires

To examine efficiency and effectiveness of using the iCare worksheet, Baseline Information and Perceptions of the iCare worksheet questionnaires were used. Nurses were asked to indicate their comfort level using Pocket PC (very uncomfortable, somewhat uncomfortable, somewhat comfortable, and very comfortable) and the number of years working in the hospital and in the unit in the Baseline Information. Nurses were also asked to provide the average time spent in giving and taking shift reports before starting the study.

Perceptions of the iCare worksheet were given to nurses at the end of the study in a self-addressed stamped envelope. Information obtained included: average time spent in giving reports and taking reports using the iCare worksheet, positive and negative experiences in using the iCare worksheet, and suggestions for improvement. In addition, nursing faculty who provided support during the study period took anecdotal notes about nurses' comments and experiences in using the iCare system.

RESULTS

Baseline Information

The response rate was 74.3% (35 received iCare training, 26 responded to questionnaires; 4 of the 26 contained incomplete data). The average years of working in the hospital was 12.62 (range = 1.5 to 36; S.D. = 10.61). The average years of working in this unit was 7.28 (range = 1 to 32, S.D. = 7.57). Most nurses reported that they were comfortable in using Pocket PCs to find information (44% somewhat comfortable; 44% very comfortable).

Efficiency

The results of descriptive statistics showed that on average, nurses using the iCare worksheet method spent 8.54 minutes less in taking reports (24.95 – 16.41 = 8.54) and 2.84 minutes less in giving reports (23.05 – 20.21 = 2.84) than the audiotape method (see Table 1). The average total time spent in giving and taking shift reports using the iCare worksheet method was 11.38 minutes less than the audiotape method.

Paired sample T-Test procedure showed significant differences in taking reports ($t = 3.29$, $p = 0.004$), but no significant differences in giving reports between these two methods (see Table 2).

Table 1. Descriptive statistics of time spent (in minutes) for giving and taking shift reports using audiotape and iCare worksheet methods

		Audiotape		iCare Worksheet	
		Giving Reports	**Taking Reports**	**Giving Reports**	**Taking Reports**
N	Valid	22	22	23	24
	Missing	4	4	3	2
Mean (S.D.)		23.05 (14.25)	24.95 (11.04)	20.21 (14.90)	16.41 (11.52)
Median		20	20	16.25	13
Minimum	7		10	0	3
Maximum	6	0	55	60	55

Table 2. Paired samples statistics: Time spent (in minutes) in giving and taking shift reports — Audiotape vs. iCare worksheet

		Mean	N	SD
Pair 1	Time spent in giving reports per audiotape	23.05	22	14.25
	Time spent in giving reports per iCare worksheet	20.09	22	15.36
Pair 2	Time spent in taking reports per audiotape	25.20	20	11.32
	Time spent in taking reports per iCare worksheet	16.38	20	11.77

		Paired Differences				
		Mean	SD	t	df	Sig. (2-tailed)
Pair 1	Time spent in giving reports per audiotape – Time spent in giving reports per iCare worksheet	2.96	13.45	1.03	21	0.314
Pair 2	Time spent in giving reports per audiotape – Time spent in giving reports per iCare worksheet	8.83	12.01	3.29	19	0.004

** Paired samples statistics (exclude cases analysis by analysis) was performed*

Perceived Benefits

Nurses wrote many positive comments about the accessibility of the iCare worksheet. All respondents (n = 26) stated that quick access to patient information was very helpful. They were glad to see laboratory results, radiology reports, diagnoses, and medications at the beginning of the shift and while providing patient care. Several nurses gave examples to describe how the iCare worksheet helped them provide patient care. For example, one patient had an emergency situation. A nurse responded and did not know this patient. She accessed information in the iCare worksheet (code status and treatment) quickly so that she could provide proper care until support staff arrived. In another example, a nurse answered the call lights of other nurses' patients. She used information in the iCare worksheet to answer patients' requests. A nurse answered physicians' and patients' questions about laboratory and diagnostic tests results promptly by accessing information in the iCare worksheet. Three nurses reported that they were glad to have comprehensive patient information in the Pocket PCs when the hospital's mainframe systems were unavailable.

Eight nurses wrote positive experiences about entering nurses' notes in the iCare worksheet. They also reported that using the iCare worksheet was faster than using audiotape to give reports to the next shift. In addition, being able to see nurses' notes of the previous shift helped them know patients' condition quickly without duplicating work. One nurse liked to use the alarms feature. She set alarms to remind her when to give medications and treatments, to check laboratory results, and to call physicians.

Key Findings from Anecdotal Notes

Anecdotal notes provided further perceived benefits of the iCare system. Initially, nurses stated that they spent more time entering data into the iCare worksheet than they did on their paper-based worksheets. Gradually, they began to voice the benefits of using the iCare worksheet. They could readily access specific patient information in the iCare worksheet without having to leave the patient's room to look up information in dumb terminals during physician rounds.

Taking notes in the iCare worksheet involved a learning curve. Nurses who took notes in the iCare worksheet stated that they became more proficient with repeated use. Some nurses voiced that the time required taking notes and synchronization was far less than the time required to manually abstract updated patient information from patient's medical records.

One nurse reported that an unexpected incident convinced more nurses to use the iCare system. One day, the hospital mainframe systems went down. No new orders or new diagnostic results could be retrieved on the dumb terminals. The only way to find this information was to call various departments. Physician rounding became quite challenging. On this particular day, nurses discovered that they could still access updated patient information in the iCare worksheet if their Pocket PCs were synchronized. Subsequently, whenever there was an indication that hospital's mainframe systems would go down, there was a flurry of activity at the sync station.

Being able to access all patients' data was another significant nurse satisfier. Nurses covered for other nurses during meal breaks or time off the unit. With the iCare worksheet, any nurse could access the most updated information on all patients. Nurses reported that this benefit did not exist with the paper-based worksheet. Some verbalized that they spent less time abstracting information from medical records and the mainframe system and had more time available for patient care.

A train-the-trainer strategy was used to have nurses helping nurses for orientation, training, and support. Many times computer-based training was performed by computer personnel. Nurses with limited computer background tended to become apprehensive because they could not comprehend the information. When the nursing faculty provided the training and were available to help them, they reported feeling less threatened in learning the new technology. Some nurses who began with apprehension due to limited computer skills became enthusiastic participants later.

Areas for Improvement and Suggestions

There was no consistent system to take notes in the iCare worksheet. Some nurses entered useful notes and some nurses who were slow to use the technology did not enter notes. Some nurses worked only once or twice during the study period and were uncomfortable with the technology. To them, talking to the audiotape recorder was faster than entering notes on the Pocket PC. Scribbled handwriting was hard to read sometimes. Voice notes could be hard to hear. Nurses would like to have had all information in one screen instead of going to each category to view. They would like to have had more information from the mainframe systems, such as physician's dictated notes. They would like to print out patient reports and to use the iCare worksheet along with computerized charting and physician-order entry system in the future.

The size of the Pocket PC was identified both as a positive and a negative. The portable size was beneficial for carrying and immediate access. The disadvantage was in reading reports with large amounts of data on a small screen, which necessitated a significant amount of side-to-side and up-and-down scrolling. They would like to have had a larger screen to view reports.

Several suggestions were related to hardware improvement. They would like to hear louder sounds for recorded notes, a lighter weight Pocket PC, a keyboard, and a longer battery life. They would like to have more sync stations, voice translation, faster handwriting recognition program, and a faster sync process. They also voiced a desire to see if the patient call lights system could be integrated with the Pocket PC in the future.

DISCUSSION

The goal of the iCare system is to use an integrated IT system using Pocket PCs to give nurses working in hospitals with paper medical records and mainframe systems quick access to patient information and to improve quality patient care. Most of the nurses in this study were using this mobile technology for the first time. Initially, they were skeptical about using mobile technology. Results indicated that nurses enjoyed the quick access to pertinent patient information at any time and any place without patients' paper medical records or dumb terminals. Their comments and anecdotal notes indicated that the iCare system enabled them to have more time to provide quality patient care.

The iCare system shortens incoming nurses' time taking reports at the beginning of the shift. When the Pocket PC is synchronized with the central PC, pertinent and updated patient information can be viewed in the iCare worksheet. There is less information that nurses need to write down. However, the iCare system does not save nurses' time when giving reports. The discrepancy reflects variations of how each nurse uses the iCare worksheet. One nurse reported spending zero minutes in giving reports while another nurse reported spending 60 minutes in giving reports. The first nurse wrote notes throughout the shift while caring for patients. This nurse used an iCare worksheet the same way as using the paper-based worksheet, in an ongoing fashion throughout the shift. The second nurse tried to enter notes in the iCare worksheet at the end of the shift to give the report. The second nurse did not use this technology the way it was intended and thus took 60 minutes to give her report. IT was often a tough sell to nursing staffs (Gillespie, 2002). Strategies are needed to help nurses embrace the technology to use it in a consistent manner.

The iCare worksheet provides a standard format to record and retrieve patient information. This format decreases variability in transferring and communicating patient information among nurses. In the paper-based worksheet, there is no common ground, each nurse uses his or her own system to transfer and communicate patient information for patient care and shift reports. Variability increases risks for errors. The iCare worksheet has the potential to improve the quality of nursing care by standardizing nurses' information transfer and communication in the hospital with paper-based medical records and the mainframe system.

This project was limited to four weeks, but the impact of this study continues to be experienced. Prior to this study, the hospital's nursing administrators had not considered using Pocket PCs for nurses. After the study, the hospital's nursing administrators saw the benefits of using Pocket PCs to access patient information in the mainframe, but chose not to expand the iCare system hospital-wide because the hospital was going to change the mainframe system to another vendor. An effort to refine the iCare worksheet to interface with the new mainframe system has been sought to continue this technology for nurses as long as the hospital is using the paper-based medical records and the mainframe system.

At the end of this study, it was also identified that nursing students could benefit from using the iCare worksheet to access key patient information instead of looking for paper medical records, which are often used by other healthcare providers. Thus, the hospital allowed the iCare system to be placed in two additional units for students to use during the school year. The iCare worksheet was loaded in both Pocket PCs and Tablet PCs for students to use. Preliminary students' feedback showed the preference of using Tablet PCs rather than Pocket PCs.

CONCLUSION

This study demonstrated that the use of mobile technology, integrated with a hospital's mainframe system, improved the efficiency of communication in shift reports and in accessing information relevant to patient care. Nurses in the study perceived that with quick access to pertinent information in the Pocket PC, they could provide better patient care. This technology provided a standardized means to communicate patient information from shift to shift. However, whether or not the patients perceived an improvement in their care as a result of nurses using this technology was not examined in this study. The future goal of this project is to revise the iCare worksheet to better meet nurses' needs to communicate patient information from shift to shift. A larger and longer study will be conducted in the future.

At the beginning of the study, it was unclear if nurses would be receptive because they had become accustomed to the paper-based system and audiotape reports. However, as they identified the benefits of having quick access to patient information with the mobile technology, they gradually responded positively to the iCare system. At the conclusion of the study, they eagerly offered suggestions for improvement. They even requested to continue using the Pocket PCs after the study ended.

The iCare system has great potential for use in hospitals that use paper-based medical records and mainframe systems. Hospitals face many constraints to fully implement IT for nurses, such as costs, human factors, regulation and standards, technology, and information systems (Androwich, Bickford, Button, Hunter, Murphy, & Sensimeier, 2003; Blair, 2003; iHealth Beat, 2005; Schneider, 1997). As hospitals decide to move away from mainframe systems to EMR systems, the iCare system could be a valuable tool to help nurses during the interim.

In the future, the iCare worksheet could be further improved by engaging nurses in the process of refinement. The iCare system could incorporate advanced technology, such as wireless network infrastructure, to expedite communication, information access, and synchronization. Tablet PCs could be the solution for small-screen issues. Further studies using Tablet PCs in the integrated IT system are warranted and underway by the authors.

ACKNOWLEDGMENTS

This is a collaborative effort between Purdue's School of Nursing, Purdue's Department of Computer Technology, and the Greater Lafayette Health Services, Inc. (GLHSI). This project received supports from Purdue University e-Enterprise center and Microsoft Research.

We would like to acknowledge the following persons from GLHSI for their continuous support and collaborative effort to make this innovative project possible: Sally Lowrey, Vice President, Nursing Services; Cathy Newman, Clinical Director, Nursing Services; Jim Siemers, Director, Information Technology; Lynn Wimberly, Director of 2 PCW at Home Hospital; Elizabeth Bratcher and Heather Askren, education coordinators.

REFERENCES

Androwich, I. M., Bickford, C. J., Button, P. S., Hunter, K. M., Murphy, J., & Sensimeier, J. (2003). *Clinical information systems: A framework for reaching the vision*. Washington, DC: American Medical Informatics Association and American Nurses Association.

Blair, J. (2003). The electronic health record today: The Medical Records Institute's annual survey shows the trends and challenges ahead. *Healthcare Informatics, 20*(1), 45-46.

Gillespie, G. (2002). IT: Often a tough sell to nursing staffs. *Health Data Management, 10*(4), 56.

iHealth Beat. (2004a, March 10). *Health affairs: Despite potential, EMR adoption faces barriers*. Retrieved on May 9, 2005, from http://www.ihealthbeat.org/index.cfm?Action=dspItem&itemID=100727

iHealth Beat. (2004b, April 26). *Bush to call for EMRs in technology speech*. Retrieved on May 9, 2005, from http://www.ihealthbeat.org/index.cfm?Action=dspItem&itemID=102114

iHealth Beat. (2005, April 19). *Study: Barriers to EMR adoption remain*. Retrieved on May 10, 2005, from http://www.ihealthbeat.org/index.cfm?action=dspItem&itemid=110216

Institute of Medicine. (2001). *Crossing the quality chasm — A new health system for the 21st century*. Washington, DC: National Academy Press, Institute of Medicine.

Lanway, C., & Graham, P. (2003). Mobile documentation. Wireless PDAs boost job satisfaction for utilization review nurses. *Healthcare Informatics, 20*(10), 80.

Laskin, R. S., & Davis, J. P. (2004). The use of a personal digital assistant in orthopaedic surgical practice. *Clinical Orthopaedics & Related Research,* 421, 91-98.

Lin, Y. H., Jan, I. C., Ko, P. C., Chen, Y. Y., Wong, J. M., & Jan, G. J. (2004). A wireless PDA-based physiological monitoring system for patient transport. *IEEE Transactions on Information Technology in Biomedicine, 8*(4), 439-447.

Mason, D. J. (2004). Shift to shift: Reclaiming patient 'hand-offs'. *American Journal of Nursing, 104*(9), 11.

Nelson, E. C., Batalden, P. B., Homa, K., Godfrey, M. M., Campbell, C., Headrick, L. A., Huber, T. P., Mohr, J. J., & Wasson, J. H. (2003). Microsystems in health care: Part 2. Creating a rich information environment. *Joint Commission Journal on Quality & Safety, 29*(1), 5-15.

Rosenthal, K. (2004). Get "smart" with a PDA — Handy devices keep information retrieval and protection within your control [Electronic Version]. *Nursing Management, 35 Supplement,* 5, 17-18. Retrieved on April 30, 2005, from http://www.findarticles.com/p/articles/mi_qa3619/is_200410/ai_n9458501

Schneider, P. (1997). Married to the mainframe. *Healthcare Informatics.* Retrieved on May 15, 2005, from http://www.healthcare-informatics.com/issues/1997/09_97/cover.htm

Sensimeier, J., Raiford, R., Taylor, S., & Weaver, C. (2002, July 12-14). *Using innovative technology to enhance patient care delivery.* American Academy of Nursing Technology and Workforce Conference, Washington, DC. Retrieved on May 16, 2005, from http://www.himss.org/content/files/AANNsgSummitHIMSSFINAL_18770.pdf

Shabot, M. M. (2002). Wireless technologies in medicine. *NET Journal.* Retrieved on June 14, 2005, from http://www.csmc.edu/pdf/Wireless.pdf

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 1, Issue 1, edited by J. Tan, pp. 1-28, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 3.9
The Evaluation of Wireless Devices Used by Staff at Westmead Hospital

Sandra Synthia Lazarus
University of Sydney, Australia

ABSTRACT

This chapter reports on a study to research and evaluate the use of latest generation wireless devices—typically personal digital assistant devices (PDAs)—by clinical staff at the large Westmead Hospital located in the west of Sydney, Australia. Currently, medical reports in this and other hospitals are primarily recorded on paper supported by personal computers at nursing stations. However, there is very little or no access to medical reports and decision-making tools for medical diagnosis at the patient's bedside—the precise location at which most medical decision-making occurs. Delays in access to essential medical information can result in an increased time taken for accurate diagnosis and commencement of appropriate medical management of patients. This chapter discusses the application of hand held devices into more powerful processing tools connected to a centralised hospital data repository that can support medical applications.

INTRODUCTION

Currently, personal digital assistant (PDA) devices are used by most junior hospital staff to view formulary information. A short survey carried out showed that almost 90% already own a PDA device and use it for drug and medication referencing. Access to drug information databases on PDAs provides a fast and easy way to reference medication details including interactions. Hospital staff members also use PDA devices for day-to-day calendar and appointment notification (Carroll & Cristakis, 2004). But it is just recently, due to the evolution of PDA devices, that it has become possible to take the next step and utilise the full capacity of the PDA devices in a hospital environment.

The study was conducted with a variety of staff members in the hospital; these included senior clinical staff, specialist staff, and junior staff. This was from across the hospital—from general wards, special units, and the emergency department. The duration of this trial was for nine months, with a rotation of staff members. Staff members were assessed before becoming part of the trial as to how long it takes the individual to carry out a particular task without the assistance of a PDA device; these tasks were timed. Then a complete instruction booklet was given to the hospital staff with the PDA device to make use of as they found it necessary in their daily duties.

After a certain period when the staff members were comfortable in using the device, time was again measured for different tasks with and without PDA devices. This set the background for the trial. To evaluate the social accepts of the trial, a survey was conducted to gather the thoughts of the users on the functionality and the useability of the device in a hospital environment.

THE EVOLUTION OF PDA DEVICES

Over the last few years, the PDA has evolved in ways originally unimaginable. One of the first PDAs to enter the market was the Palm Pilot (now more commonly known as the Palm). When first released, the Palm came with a monochrome screen and had a limited amount of onboard memory. Since then the Palm evolved into a PDA that featured expandable slots for additional memory and expandability options (i.e., modems, printers, and other peripherals). In addition, Palm PDAs now feature high-resolution colour screens which are continuously improving. Furthermore, the Palm also comes with built-in wireless access that permits it to connect to various types of devices and networks, allowing it to be truly portable (as originally intended).

Pocket PCs entered the market shortly after Palm PDAs were introduced; however, unlike Palms the Pocket PC is more of a PDA standardisation rather than a brand name. Pocket PCs are made by several companies, for example Dell, Toshiba, and HP/Compaq; however, the core operating system is written by Microsoft. On the other hand, the Palm PDA operating system (Palm OS) is maintained by Palm itself, which ensures that all features of the operating systems are fully functional. Small distinctions such as this have allowed both devices/brands to compete for the same market space and are still considered independent niche markets.

Since Palm has been on the market for a considerably longer period of time, there is a remarkable amount of software support found throughout the Internet. Palm PDAs now have the ability to operate seamlessly with Windows, Linux, UNIX, Novell, Macintosh, and practically any type of network. There is a considerable amount of open source development done within the communities that make featured software available for Palms. Similarly, Pocket PCs are supported through a large number of Internet sites. Since standards are maintained, which include desktop and server operating systems, these have the ability to ensure that their operating systems designed for Pocket PCs will meet the same quality control and compatibility assurance. Nevertheless, with the rivalry between these two PDA vendors, it is possible to see further improvements such as superior screen resolution, improved standard of expandability options, inclusions of newer wireless technology such as 802.11g, and most importantly reduced costs, allowing more people to access such technology. Since their evolution, PDA devices can now interact with the latest available wireless technology and can be used in a hospital environment. The two PDA devices used in this case study are the Palm Tungsten T3 and the Toshiba e800 Pocket PC.

Palm

The Tungsten T3 is selected for this study because of its flexible features. At the time of the study, the T3 was the most highly developed Palm device, with built-in Bluetooth wireless technology, 64 MB memory, and moderate screen resolution. To reduce screen scrolling, the T3 has a stretch screen display as well as portrait and landscape modes.

Figure 1. The Palm Tungsten T3 has the following features: Bluetooth technology synchronisation, 64 MB memory, stretch display, portrait and landscape, moderate resolution 320´480 pixels (Palm One, 2005)

Figure 2. The Toshiba e800 Pocket PC has the following features: Wi-Fi wireless technology—synchronisation, easy configurations for 802.11b wireless environment, 32 MB memory, portrait and landscape, VGA high-resolution screen/480´640 pixels (Toshiba, 2005)

Toshiba e800 Pocket PC

The Toshiba e800 Pocket PC has both Bluetooth and Wi-Fi capabilities, with a memory capacity of 32MB. This has proven to be sufficient memory for the PDA-based medical programs. The e800 has a high-resolution VGA screen, 640 by 480; this improves the images on the Pocket PC and allows medical staff to read data with ease.

AVAILABLE WIRELESS TECHNOLOGIES

Wireless technology has been used for several years in the information technologies area; however, more recently there has been more fascinating development in wireless technology, allowing its useability in a wide range of areas. One such area is medicine: more and more new areas are being developed in which wireless technology is becoming a necessity (Bird, Zarum, & Renzi, 2001). Wireless technology in medicine allows mobility which provides greater advantage when it comes to patient care and patient diagnosis.

Bluetooth technology is an effective radio technology that allows for point-to-point or point-to-multipoint support which connects multiple Bluetooth devices. It has proven to be the most effective cable-free connection for wireless technology.

Some of the features of Bluetooth technology are: the transmission distance is real-world range between 10 and 42 meters, and transmission speed which provides throughput comparable to modems and DSL and security with Bluetooth integrates security mechanisms within several layers of its protocol. Connections are generally configured to use authentication and 128-bit encryption. Applications can build their own security on top of Bluetooth connections to make communications even more secure; radio power output is as little as 1% of the power output of Wi-Fi (Palm One, 2004).

The 802.11x network is more commonly known as Wi-Fi, of which there are three general types: 802.11a, 802.11b, and 802.11g.

The most popular Wi-Fi is 802.11b, which has a throughput capacity of 11 Mbps and an approximate range of 50 meters. Due to bandwidth overheads and congestion, 802.11b will deliver data at about 7 Mbps, which is more than 10 times the performance of Bluetooth (Colin, 2002). The advantage of 802.11a and 802.11g is a five-fold increase in data rate over 802.11b (Colin, 2002).

These are only a few examples of the variety of wireless devices, some of which are being developed for specific use.

Medical Programs on PDA Devices

Apart from the Cerner Pocket Power Chart, other PDA medical software programs were installed on the PDA devices used in the study to assist in decision making and to increase efficiency. This allows medical staff members to access medical information at patient bedsides, reducing the physical load of books and decreasing search time. For example, on a ward round, a staff member may need to enquire about interaction between two drug types. There are two options in this situation: call the pharmacist, or look up the information at the nurse's station on the Internet or in a reference book. With the PDA at the patient bedside, the staff member now has access to the relevant information immediately. These medical programs used on the PDA were selected based on suggestions from medical staff members at the hospital.

iSilo is a program that converts file formats utilising compression for reducing document size. iSilo was used to reduce the size of the Australian Medical Handbook and the Westmead Medical Handbook so that they could be installed on the PDA device.

Medical Calculator is used in performing common calculations useful for clinical decisions or for analysing data in clinical decisions.

Medical Handbook is used by most junior medical staff as a resource to manage common medical and surgical problems.

MIMS (Monthly Index of Medical Specialities) provides medical staff and healthcare professionals with easy access to drug information required for making medical decisions.

Therapeutic Guidelines provides information to support decision making in the prescription of drugs.

HotSync allows programs to be uploaded to PDA devices from desktop computers.

Cerner Pocket Power Chart is a program that provides a mobile solution for hospital results reporting. It enables doctors and other healthcare providers to view patient results such as pathology and radiology reports.

The Cerner Pocket Power Chart has a username and password login as part of its program functions. Each doctor is allocated a username and password that are required for login to view patient lists. All hospital databases are secured behind firewalls, adding to the security of the device. The Cerner client software has built-in security—it is designed to erase all information on a PDA device if that information is not reviewed within 72 hours. This ensures that if a Palm device is lost or stolen, all Cerner Pocket Power Chart patient reports will be erased after 72 hours. This feature protects all patient information and upholds patient confidentiality.

Evaluation Research Method

To evaluate the useability and functionality of the PDA devices in a clinical environment, a survey was conducted. PDA devices were uploaded with medical support programs currently in use by some medical staff, such as MIMS, Therapeutic Guidelines, iSilo, MedCalc, the Westmead and Australian RMOs Handbooks, along with the current available version of Pocket Power Chart Palm OS client from Cerner Corporation, the vendor of the enterprise's results reporting system. The Citrix Client is used to view Cerner patient information and data on the Pocket PC handheld devices.

Doctors from various backgrounds were given PDA devices, complete with necessary programs, a user manual, and surveys to complete. They were allowed to work their way through the different programs in an unstructured way. After a week of using the PDA devices, observations were carried out to determine how often the programs and PDA devices were being used to assess patients and assist with decision making.

Survey

The survey was divided into three sections: the first section contained statements which were answered with agree, strongly agree, disagree, and strongly disagree. The second section consisted of yes and no answers, and the last section was a short, open answer section that allowed the medical staff to express their opinions about the PDA trial. The survey was completed after the trial period.

Statistical Analysis

A statistical analysis was carried out to calculate time utilisation in the clinical trial. Each medical staff member was timed before the clinical trial and during the clinical trial. They were timed on their hospital duties on different hospital shifts, such as ward rounds, overnight shift, patient discharge, and new patient evaluation. This was carried out to determine how long it takes a medical staff member to carry out his or her duties with or without a PDA device. Times were noted and then put into a graph showing the range of time it took to complete, and the average time it will take to complete different duties in different hospital shifts (as outlined in the results).

Method Outline

- A questionnaire was completed by the medical staff before the clinical trial. This determined general knowledge of PDA-type devices.
- Medical staff members were given a PDA device with a complete set of medical software and a user manual for a period of two to three weeks or more if required.
- After the trial period a survey was completed to determine how effective the medical staff found the PDA devices and the PDA-based medical programs.
- Observations were carried out to calculate the average time to access five patients with or without the PDA device.
- How frequently different medical programs were being used was also noted.
- The average time it took to prescribe medication with or without electronic medical databases or PDA-based programs was determined.

Synchronisation of Patient Information

HotSync

HotSync is a Palm-based computerised platform that allows data to be transferred between the Palm devices and the computer. This transfer of data is known as synchronisation and can be carried out through a cable connection or a wireless connection. Palm-based applications and programs are uploaded to the devices using HotSync synchronisation, as shown in Figure 3.

Citrix

Citrix, a revolutionary software, delivers terminal services to a number of operating systems such as Windows, Linux, UNIX, and Macintosh. In addition to these operating systems, Citrix also supports Palm and Pocket PC devices. A Citrix client allows a PDA to connect to a Citrix server, making it compatible to a standard desktop PC or laptop. As a result there are now practically no limitations to portability and compatibility of software restrictions on PDA devices. In addition, there is no longer a need to make a specialised application for both formats—Pocket PC and Palm. Rather, with the power of a terminal session through Citrix, the only requirements would be for the application to run on a desktop such as a Windows operating system.

However, Citrix is not only a medium for reporting applications in terminal sessions, but also a client/server-based application that provides built-in layered security protocols allowing security and confidentiality. A Citrix client can be configured with a wide range of security provisions, including standard authentication using a username and password to extremely high levels of security using EAP (Extensible Authentication Protocol) (Citrix, 2004). Security controls such as certificates, smartcards, or even fingerprint and retinal scans can be used for identification and authentication. A Citrix client can also connect to a Citrix server via the Internet using the World Wide Web, allowing it to be published externally to the Internet, thus allowing ease of use and accessibility worldwide (a handy feature when you are out of the office and require a document stored on a network drive) (Citrix, 2004).

Figure 3. Using HotSync and Bluetooth, wireless technology patient information on Cerner Power Pocket Chart is uploaded to the Palm Tungsten T3

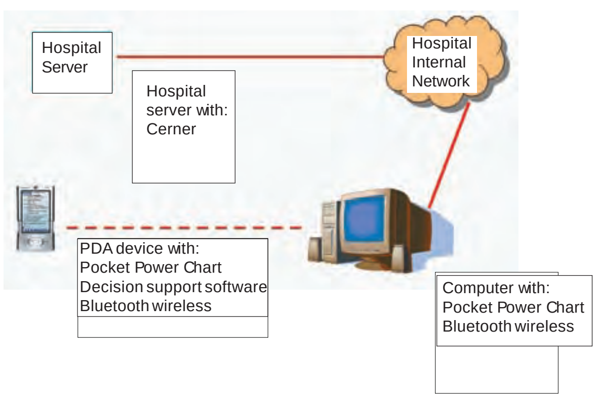
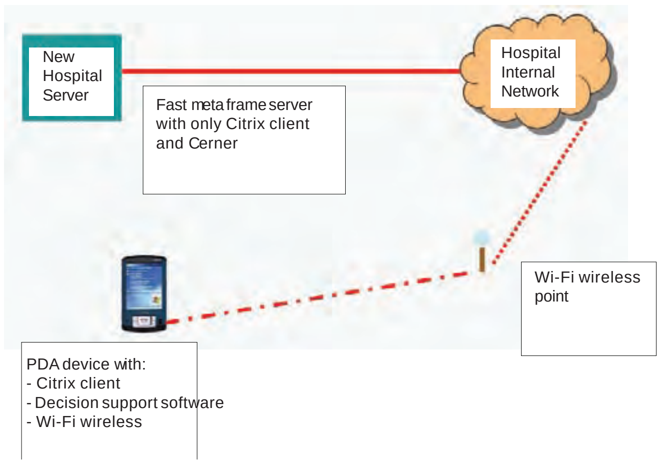


Figure 4. Through Citrix terminal sectioning and Wi-Fi wireless technology patient information on the Cerner program is captured on the Toshiba e800 Pocket PC



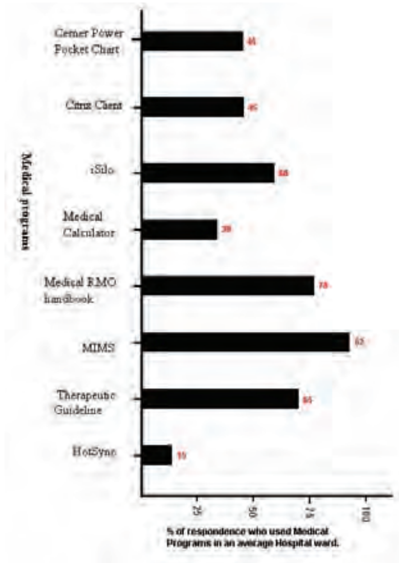
The Use of PDA-Based Medical Programs

Graph 1 outlines the percentage of PDA programs utilised during various hospital ward shifts. Cerner Pocket Power Chart (CPPC) was used by 45% of the doctors; the lower percent was due to a few problems with the CPPC program which includes the time (up to 24 hours) taken to upload and access new patient information and results on a PDA device. The usage may increase if the CPPC program is more efficient in uploading patient information and results. The Citrix client and CPPC had identical usage rates, as CPPC uses Citrix to access patient information. Furthermore, iSilo was used by 60% of the doctors. This was due to medical staff uploading other medical reference programs and using iSilo to convert them into PDA-based programs. The Medical Calculator was mainly used by specialist medical staff to convert and calculate test results when they were required, with approximately 30% reporting use of the application. The Medical RMO handbook was the second most popular program used by 78% of the doctors in a hospital shift. MIMS was the most used application, with 82% of the doctors each shift reporting that they used it, mainly for drug information and drug interactions. Therapeutic Guidelines was found to be the third most popular medical program used by 65% of the trial participants. HotSync had the lowest value of 15%, as its only role is to assist in uploading CPPC from the hospital Cerner desktop program. During an average shift, patient information is only refreshed if a new patient is admitted or new test results are required, meaning that HotSync was not required as often as the other programs.

Time Analysis of PDA Devices While Performing Medical Tasks

Graph 2 illustrates the varying timeframes required for medical activities with or without a PDA device. The graph outlines the average time range taken to carry out a task during different shifts at the hospital. It is evident from the graph

Graph 1. The percentage of respondents who used each medical Palm-based program on an average hospital ward (not to scale)



Graph 2. The range and average time to carry out medical tasks on ward rounds, overnight shift, patient discharge, and new patient assessment (not to scale)

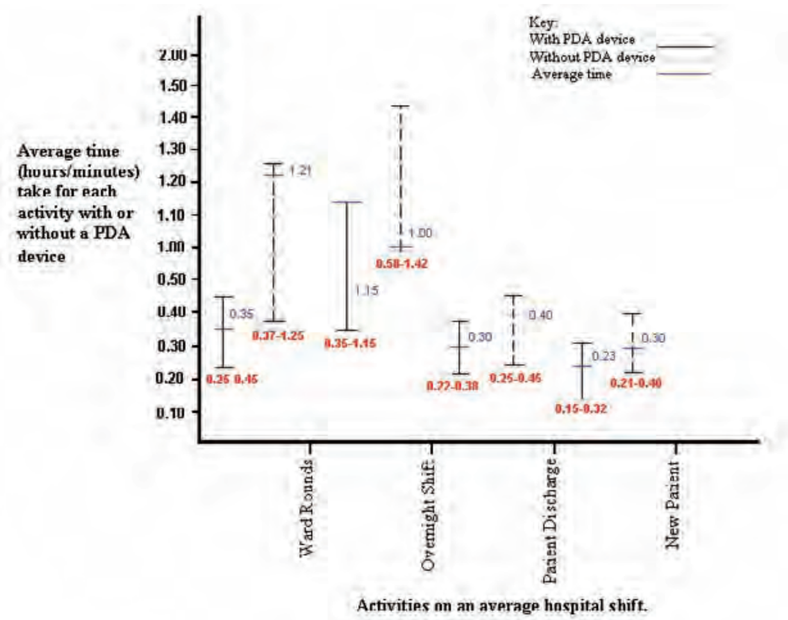


Table 1. Outlines the user percentage on a survey statement

Survey Questions	**Percentage Outcome (%)**			
	Strongly Agree	**Agree**	**Strongly Disagree**	**Disagree**
Did the PDA device improve access to patient information and necessary medical information at the patient's bedside?	23	58 5	1	4
The PDA device with its medical programs assisted in decision making and diagnosis.	18	78 0	4	
Security level in this trial and on PDA application is sufficient.	15	44 1	2	29
Mobile devices such as PDA devices and mobile tailored medical programs can be seen as a daily part of a hospital environment.	21	59 9	1	1

that regardless of the hospital shift or task, the PDA devices and PDA-based programs assisted in reducing time taken to evaluate patients. For example, managing common medical problems and searching for medication for patients requires medical staff members to go back to the nurse station from the patient bedside to look up the necessary information. This can take up to 30 minutes as they go through several book references. But with the PDA devices, the medical staff can stay at the patient's bedside and search for the necessary information in less than a minute (this time is in accordance to observations made during the trial period).

Survey Outcome and Options

The survey conducted outlined the options of the users as well as the productivity of the PDA devices and medical applications and programs. Mobile medical programs and applications were seen as a convenient and essential part of daily duties around the hospital (Tschopp, Lovus, & Geissbuhler, 2002). A larger majority of the users considered the programs a fast and easy tool while making discussions and diagnosis, and approved its use and availability when required at the patient's bedside. The main concern was security of patient information in an event of a theft or lost property; this was secured with the 72-hour limitation on all patient information (Cerner program function) and a complete shutdown of the PDA device if power is not maintained which will erase all stored information in the memory. The security is handled with a user login and server firewalls which maintain the overall security of the hospital wireless system. It was noted that the Palm T3 added functionality of an expandable screen, reduced screen scrolling, and increased visibility. Also the VGA screen on the Pocket PC Toshiba e800 with high resolution provided x-ray images for reference. These increased the potential and further the use of such devices, making them suitable for medical operations.

CONCLUSION AND FUTURE DIRECTIONS

From the above results and feedback from clinical staff at Westmead Hospital, it can be concluded that such trial can be seen as a success and a route to possible implementation of wireless technology. Similar clinical studies have looked at the possibility of PDA devices being introduced into hospitals for clinical use. Previous clinical trials of handheld devices have demonstrated both limitations and benefits to these devices. The PDA clinical trial had a few limitations; these were found in PDA based medical programs. Cerner Power Pocket Chart was found to provide only a limited amount of patient information and patient results. It limited patient results' viewing time to 72 hours and was unable to show previous results for comparison, a major disadvantage in clinical management. This reduced the use of CPPC, as medical staff had to go back to traditional methods to view complete patient details and results. For the CPPC program to be a successful tool in patient evaluation, improvement is required in areas such as patient administration details. However, other PDA medical programs combined together proved to be a very useful collection of reference tools for the medical staff. This also encouraged users to take the initiative and upload other PDA medical programs. To gain a better understanding of the benefits of PDA device usage, it may be important to extend the trial period to allow intermixing periods with and without the staff member having access to a PDA device. This will allow evaluators to measure the effects of PDA devices on hospital staff when the device is used for a longer period.

The main aim of this trial was to evaluate time taken to carry out different medical duties in the hospital setting (Ruland & Cornelia, 2002). Previous trials have concentrated on the functionality of the PDA device or opinion of the medical staff (Carroll, Tarczy-Hornoch, O'Reilly, & Christakis, 2004). This trial considered all three factors: staff option through the use of different medical programs and software, use ability of the PDA device in a hospital environment, and also accessed time during different medical duties. As noted through the results, it is evident that the trial has proven to be successful in determining how effective this method of patient assessment can be, with the correct software environments.

ACKNOWLEDGMENT

The author gratefully acknowledges Chris Liddle of the University of Sydney, the Westmead Hospital Clinical Pharmacology Department, the Westmead Hospital Information Technology Department, the Westmead Hospital Medical Staff, Citrix, Cisco, Palm One, Toshiba, and Therapeutic Guidelines.

REFERENCES

Bird, S., Zarum, R., & Renzi, F. (2001). Emergency medical resident patient care documentation using a handheld computer devices. *Academic Emergency Medicine, 8*(12), 1200-1203.

Carroll, A., & Cristakis, D. (2004). Pediatricians' use of and attitudes about personal digital assistants. *Pediatrics, 113*(2), 238-242.

Carroll, A., Tarczy-Hornoch, P., O'Reilly, E., & Christakis, D. (2004). The effects of point-of-care personal digital assistance use on resident documentation discrepancies. *Pediatrics, 113*(3), 450-454.

Citrix. (2004). Retrieved June 24, 2004, from http://www.citrix.com/lang/English/home.asp?source=google&keyword=citrix&sitrackingid=399741

Colin, D. (2002). *Going wireless.* Retrieved August 1, 2004, from http://www.projector central.com/wi-fi.htm

Palm One. (2004). *Bluetooth wireless technology.* Retrieved March 11, 2004, from http://www.palmos.com/dev/tech/bluetooth/

Palm One. (2005). *Palm One Tungsten T3 image*. Retrieved February 10, 2005, from http://www.palmone.com/us/support/tungstent3/

Ruland, C., & Cornelia. (2002). Evaluating a support system for preference based care planning at the bedside. *Journal of American Medical Informatics Association, 9*(2), 192-201.

Toshiba. (2005). *Toshiba e800 Pocket PC image*. Retrieved February 10, 2005, from http://www.isd.toshiba.com.au/cgi-bin/

Tschopp, M., Lovus, C., & Geissbuhler, A. (2002). Understanding usage patterns of handheld computers in clinical practice. *American Medical Informatics Association, 1531*(605), 806-809.

This work was previously published in Handbook of Research in Mobile Business, edited by B. Unhelkar, pp. 96-105, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.10
A Preliminary Study Toward Wireless Integration of Patient Information System

Abdul-Rahman Al-Ali
American University of Sharjah, UAE

Tarik Ozkul
American University of Sharjah, UAE

Taha Landolsi
American University of Sharjah, UAE

ABSTRACT

This paper presents the results of a study toward generating a wireless environment to provide real-time mobile accessibility to patient information system. A trial system is set up where database, Internet, and wireless personal digital assistants (PDAs) are integrated in such a way that the medical professionals like physicians, nurses, and lab assistants can create, access, and update medical records using wireless PDAs from any location in the hospital, which is covered by wireless LAN. The same services, which can be carried out via fixed terminals with Internet connectivity, can be carried out using wireless PDAs. The implementation has used and integrated many technologies like active server pages (ASP), Visual Basic®, structured query language (SQL) server, ActiveSync®, IEEE802.11 wireless locala area network (WLAN) technology and wireless security concepts. The paper details the architectural aspects of technology integration and the methodology used for setting up the end-to-end system. The proposed architecture, its performance data, and the common implementation barriers are reported.

INTRODUCTION

Medical professionals have already recognized the importance of keeping patient information (medical records) in an electronic format rather than paper-based format because of the sheer size of records generated daily. Due to the extensive size and costly storage requirements, keeping paper-based records became more expensive than keeping records electronically. A study conducted in a 500-bed hospital indicated that a 7-inch stack of paper-based laboratory reports must be filed daily. The informal survey was conducted with medical professionals among the American University of Sharjah medical center staff and a neighboring local hospital in which receptionists, laboratory, and X-ray technicians, nurses, and physicians participated. Most of them liked the idea of using electronic patient record (EPR) technology. Some of them expressed some concerns about the screen size and the resolution limitation of the personal digital assistants (PDAs) used in the trial. Others worried that if such ubiquitous systems are deployed, then medical staff will have to be available all the time even during vacations days. They stated that if such technology is available, they will be liable if they do not answer even during their breaks. "It is a matter of life and death" one of the nurses stated, "I should answer calls anywhere, anytime."

The cost of maintaining paper-based records and filing them in an ordered fashion to keep them accessible is over US$10 per record (Safran & Goldberg, 2000). Keeping records electronically also presented the opportunity of being able to access records over the Internet from anywhere, anytime. Together with the powerful PDAs and wireless connectivity tools, it became feasible to access EPR remotely without being tied to workstations. There are several records in the literature, which mention successful implementations of Web-based access to patient databases (Liu, Long, Li, Tsai, & Kuo, 2001; Garcia et. al., 2002). Others have reported wireless healthcare using wireless local area networks (WLAN) and discussed the electronics home healthcare concepts and challenges (Wang & Hongwei, 2005; Wickramasighe & Misra, 2004). A trial study conducted recently among medical professionals in real hospital settings indicated that medical professionals regard mobile access to the following data highly useful (Ammenwerth, Buchauer, Bludau, & Haux, 2000):

- Medical knowledge like drug data.
- Medical coding references like ICD-10 codes (International Classification of Diseases) and literature databases.
- Patient database and administrative patient data.
- General information like telephone numbers and medical databases.

After using the system for a week, the respondents indicated that mobile communication and mobile information processing power offered by PDAs are very valuable. However, the respondents also reported that they were not satisfied with the 9600-baud rate communication speed offered by the early versions of PDAs and the mobile phones based on the Global System for Mobile (GSM) standard used in the study. During the study, it also became apparent that the messaging ability offered by PDAs was much superior to personal accessibility provided by pagers and mobile phones (Ammenwerth et al., 2000). Since then, the rapid change in the technology provided better connection methods, more durable and faster handheld mobile computing devices. The wireless accessibility provided by nowadays existing WLAN standards such as IEEE802.11g can support 54 Mbits/s data rate and the soon to come IEEE802.11n standard will support 540 Mbits/s data rate. This will clearly satisfy the need to higher access bandwidth required by healthcare providers. Along with the other contemporary software and database tools, this new connectivity method promised better EPR system and motivated many researchers in the healthcare industry to develop integrated wireless applications for use on pocket PC, smart-phone PDAs and other portable device platforms (Lu, Xiao, Sears, & Jacko, 2005).

In this study, we will design a prototype electronic medical database system and evaluate its performance in near-realistic settings. Another aspect of the research conducted is the design of a Web-based database for hospital environment, which could be equally accessed by wireline and wireless networks.

TECHNOLOGY IMPLEMENTATION BARRIERS

As is the case for many newly introduced technologies, even with uncontested technological and economical benefits of the proposed architecture the implementation may be impeded by the perceived steep learning curve for potential users. This fact is especially true for medical professionals who are in general uncomfortable introducing new technology due to the risks involved and required "protected time" to integrate it in their work environment (Van Ginneken, 2002). The flexibility and adaptability to change are regarded as the key factors for medical applications. Standardized and open-vendor systems are also important factors for getting new technology accepted by the user community. Acceptance of EPR by the medical community has improved after an initial hesitation. The technology to access patients' information varies. Literature search indicated that there are several techniques that are currently being used:

- Hybrid architecture using PDAs and wireless GSM modems
- IEEE802.11 WLAN standard
- IEEE802.15 known as Bluetooth standard

Software applications were based on Wireless Application Protocol (WAP). The hybrid system using PDA and GSM modems combination was used in a trial study conducted in real hospital settings (Van Ginneken, 2002). This valuable study was aimed at determining expectations of medical professionals from potential wireless enabled hospital settings. The study indicated many points, which are regarded important by medical professionals. But the study also indicated that the data rate offered by the hybrid solution is far lower than expected to be considered useful. Another connectivity method used in some studies was WAP connectivity, which is used by mobile phones to transfer Web contents to mobile devices. In a successful WAP-based system, patient data was sent successfully from patients monitored to WAP-enabled phone. Information like Electro-Cardiogram (ECG) signal was displayed graphically on the medical professionals WAP enabled phone (Hung & Zhang, 2003). Although the project was successful, the authors reported that the data transfer rate was low. Another major drawback of the system is the limited screen size of the WAP phone. Another similar study used WAP services for connectivity to biological databases (Riikonen, Boberg, Slakoski, & Vihinen, 2002). The Mobile and PDA phones successfully queried the databases and displayed the contents in a browser page on the mobile device. PDA-based client-server architecture is preferred for flexible telemedicine systems (Nazeran, Setty, Haltiwanger, & Gonzalez, 2004). The client uses a Pocket PC platform because of its processing capabilities, low cost, and compact size. The system could be used to transmit audio, still images and vital signs from a remote site to a clinic or a hospital Web-server that implements standard Internet protocols.

Adoption of wireless solutions in the healthcare sector has many advantages but also poses some challenges (Demiris, 2004; Lu et al., 2005). Time and cost saving, mobility and real-time access, reducing medical errors, enhancing productivity, and quality of care are major advantages of using PDA in healthcare (Lu et al., 2005). However, speed, screen size, data entry, maintenance, interoperability issues, patient privacy issues, interference with medical equipment, data security issue, and negative patient perception of delicate devices are notable challenges (Lin & Vassar, 2004; Lu et al., 2005).

DESIGN CONSIDERATIONS FOR THE TRIAL SYSTEM

Literature and informal surveys of medical professionals showed that the designed system should have the following requirements:

- The system should have an EPR database which is user friendly, robust and Web-enabled.
- The system should be secure. Several layers of security should be established at different levels.
- Access speed to the system, data access and update speed should be high; PDA used for the system should have reasonable size and weight.
- The system should use off-the-shelf components and the cost should be reasonable.
- The system should be accessible at any time. Hence, architecture should be robust to provide the needed high availability.

To achieve the previous requirements a client-sever model is designed and constructed. Figure 1 shows the system hardware architecture. A database is designed and managed by SQL-sever, Web site and wireless connectivity via PDA are developed. Figure 2 shows the system software tiers.

Database Design

A database system is designed to serve the requirements of the overall system. Entity relationship diagram (ERD) is used to describe the objects in the hospital database. The objects or entities and the relationship between them were translated into tables. This process helped us identify facts, known as attributes or fields, about these entities. Five entities are used namely; employee, patient, record, department and test. The entity employee consists of doctors, nurses, and lab assistants. For example, each employee is identified by his or her ID number, first name, last name, department ID number, employee designation, address, contact number, and nationality.

Figure 1. System hardware architecture

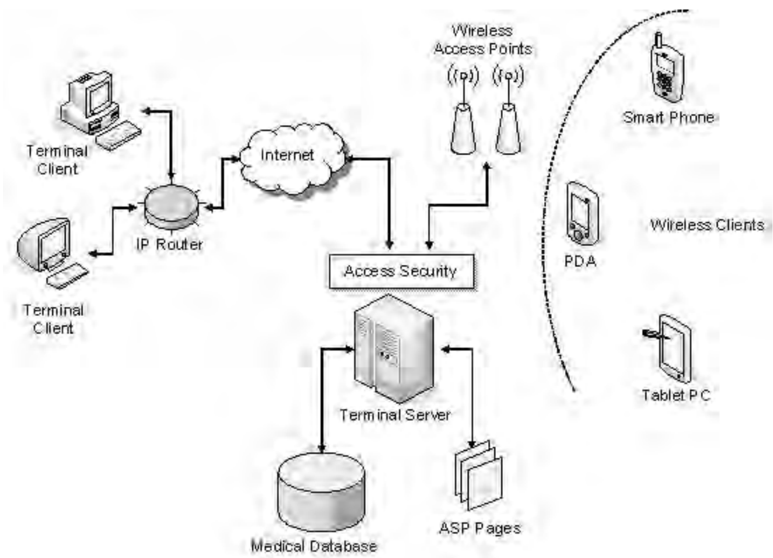


Figure 2. System software tiers

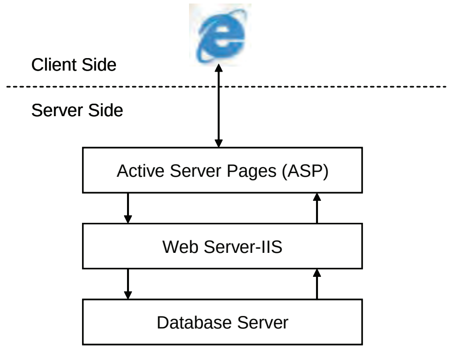
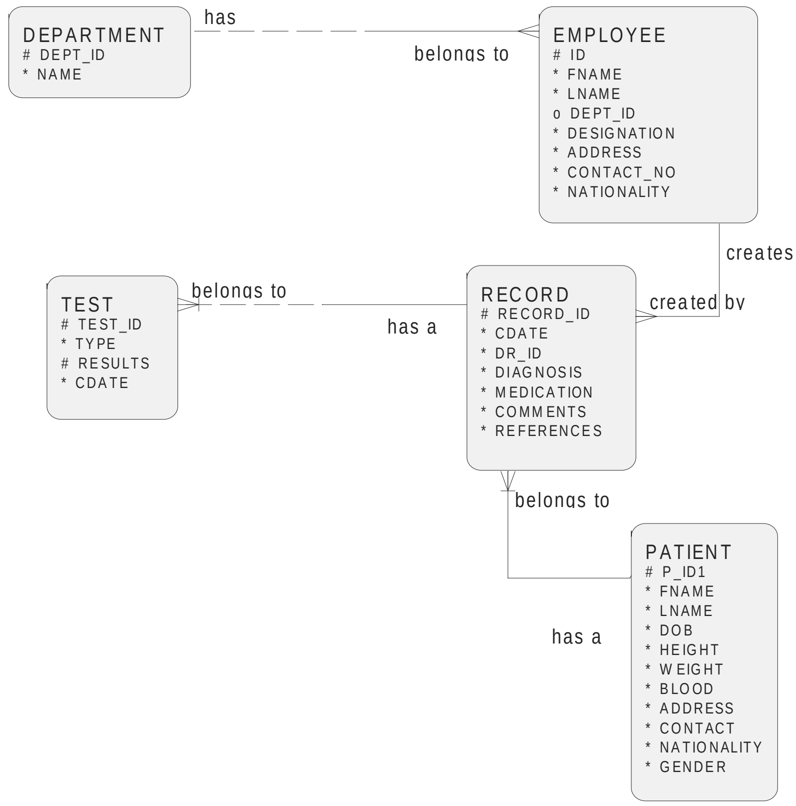


Table 1. Database fileds corresponding to the system entities

Sr. No	Entity	Attribute	Description
1	**Employee**	ID	Unique identification number of the employee
		FName	First name of the employee
		LName	Last name of the employee
		Dept_ID	Unique department to which the employee belongs
		Designation	Post of the employee in his or her respective department
		Address	Mailing address of the employee
		Contact_no	Contact information
		nationality	Nationality of the employee
2	**Patient**	P_ID1	Unique identification number(patient id) of the patient
		FName	First name of the patient
		LName	Last name of the patient
		DOB	Date of birth of the patient
		Height	Height of the patient
		Weight	Weight of the patient
		Blood	Blood group of the patient
		Address	Mailing address of the employee
		Contact_no	Contact information
		Nationality	Nationality of the employee
3	**Record**	Record_ID	Unique identification of the record along with the P_ID
		CDate	System Date when the record is stored
		Dr_ID	Doctor id
		Diagnosis	Diagnosis of the medical problem
		Medication	Medication prescribed
		Comments	Further comments regarding the record
		References	References to other doctors, medical staff, etc.
4	**Department**	Dept_ID	Unique identification of the department
		Name	Name of the department
5	**Test**	Test_ID	Unique identification of the test in combination with the record id and patient id
		Type	The type of the test
		Result	The results of the test
		CDate	System Date when the record is stored

Figure 3. Entity relationship diagram



A table for each of the entities was constructed and the unique and primary keys for the entities were issued. The final design was implemented in SQL server. The SQL Server technology was chosen because it has enterprise data features, is better in maintaining data integrity, supports triggers and rollbacks, and stores procedures and dynamic data processing. Table 1 describes the entities and the attributes along with a brief description of each. Figure 3 shows the Entity Relationship Diagram.

Web Site Design

A complete functional and interactive Web site for the hospital "Care Well Hospital" medical centre, which is accessible through local intranet and from the Internet, is designed. This site is developed using Active Server Pages (ASP®) and Visual Basic® scripting that was embedded in HTML files. The advantages of using ASP are fast execution, no client-side constraints, ODBC links to any data source and orientation towards Microsoft products. A snapshot of the home page is shown in Figure 4. A dropdown menu helps the users to activate one of the hotlinks to access any of the system functions such as medical services, visiting hours, login, logout, etc. An authorized person can access personal records, patient records, patients table, and add new patients' records. For example, the patients' table link returns a list of all patients who are already registered in the system. Figure 5 shows the existing patients list. Search by a specific key such as "Patient ID" or "Last Name." Figure 6 shows that a patient search by last name "Kumar" returns a list of patients records with matching last name.

Figure 4. Homepage of "Care Well Hospital" Web site designed for the study

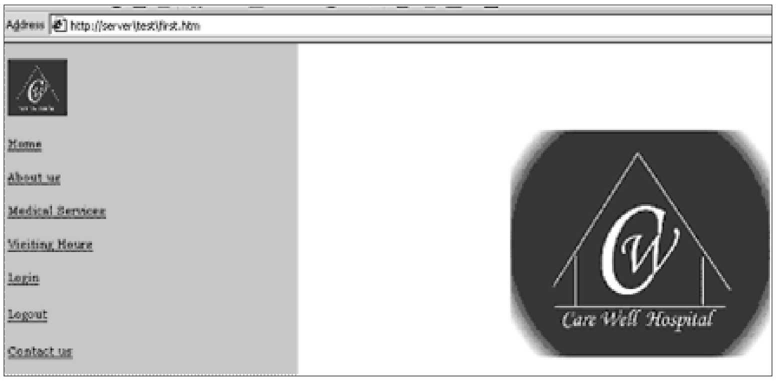
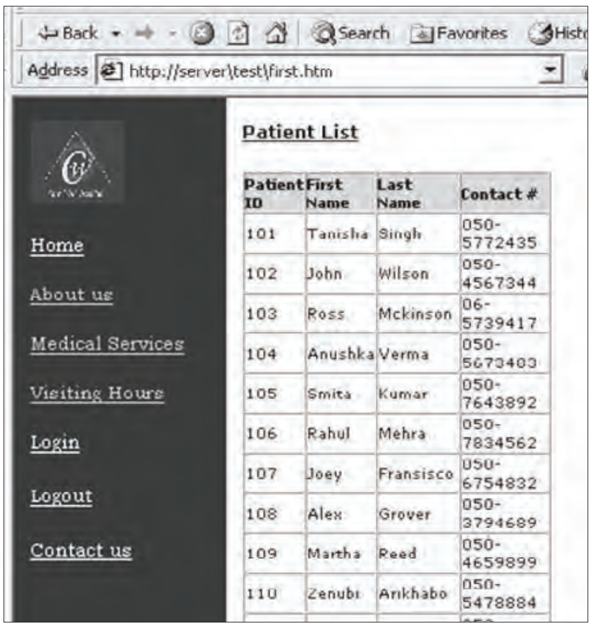


Figure 5. Patients record table



Patient ID	First Name	Last Name	Contact #
101	Tanisha	Singh	050-5772435
102	John	Wilson	050-4567344
103	Ross	Mckinson	06-5739417
104	Anushka	Verma	050-5673403
105	Smita	Kumar	050-7643892
106	Rahul	Mehra	050-7834562
107	Joey	Fransisco	050-6754832
108	Alex	Grover	050-3794689
109	Martha	Reed	050-4659899
110	Zenubi	Arikhabo	050-5478884

PDA INTERFACE AND CONNECTIVITY

System Hardware and Software Description

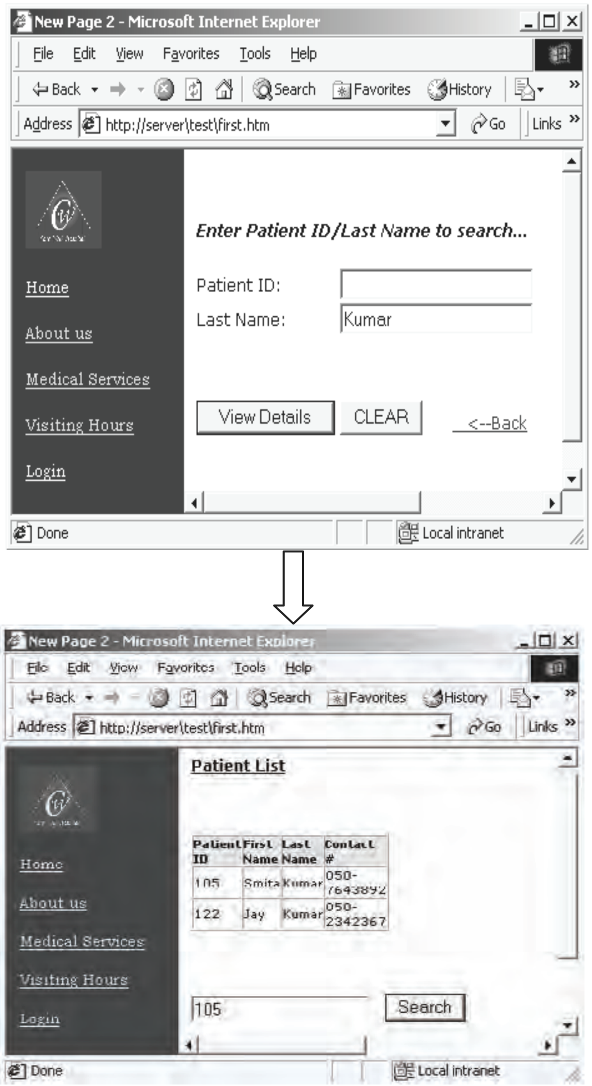
To access the Web site that contains the patient information system via PDA, combination of hardware devices and software drivers are needed.

As shown in Figure 1, the hardware requirements are:

- A Web server along with Internet connection to the service provider.
- PDA with WLAN access interface card.
- Wireless access points.

The software drivers used a set of protocols that provide client utilities such as client manager,

Figure 6. Search by last name output



link status, wireless client login, and encryption manager. The USB connection between the server and the PDA is used to install the above-mentioned software utilities on the PDAs. The USB cable is later removed and the wireless connection became available between the PDA and the LAN via the wireless access point.

The system provides two modes of connectivity, namely server-client model and wireless model. The server-client model uses standard access method to the LAN of the hospitable where doctors, nurses, and lab assistant can create, modify, and access the patient's information system via the Internet Service Provider (ISP) wired conductivity.

In the wireless model, the study used an off-the-shelf Compaq-200 PDA with the following specifications:

- 400MHz processor, Windows CE for Pocket PC operating system, 64-MB SDRAM & 32-MB Flash ROM memory.
- Interfaces: WLAN (IEEE802.11b) and Bluetooth (IEEE802.15) compliant. The distance range of the PDA was 100 meters (class-2 radio).
- TFT liquid crystal display with viewable 64K color image size 2.25 inch wide by 3.02 inch tall.

For wireless access purposes, commercially available wireless access points are used in the study. The access point provides 40-bit and 128-bit wired-equivalent privacy (WEP) encryption security over the 100m coverage area. It is IEEE 802.11b compliant and works at self-adjusted data rates of 11, 5.5, 2, and 1 Megabits/s (Mbps).

Wireless Access Client-Server Synchronization

Synchronizing the wireless access between the Server and the Client is done in two ways:

- Configuration of the wireless access through the Server being connected to the Internet: This was accomplished using a hub through which the server was connected to another network, which was connected to the Internet. The PDA accessed this server through the wireless access point and was able to access sites on the Internet as well.
- Configuration of the wireless access through Terminal Services on LAN: Wireless access through terminal services was introduced as a safety procedure in case Internet access is not available. Terminal services are implemented using Windows 2000® Advanced Server.

SECURITY ISSUES

Two levels of security were implemented, one for the advanced server and the other for the SQL server.

Windows Advanced Server Security

Using the administrative tools, privileges are granted to users and groups. Privileges include restarting the server, modifying certain settings, accessing databases, etc. This is done by creating different profiles for the users and then granting the required privilege. For instance, if an employee is registered under the group "Lab Assistant," upon logging-on to the Windows 2000® system as a Lab Assistant, he or she is given those privileges assigned to the group "Lab Assistant." The privilege denies the lab assistant to reboot the system.

Microsoft SQL Server Security

The SQL Server authentication mode used is "Windows Authentication Only." In other words, SQL Server automatically authenticates users based on their Windows user account names or their group membership. If you've granted the user or the user's group access to a database, the user is automatically granted access to that database.

The SQL Server security model controls access to the database using the server login, permissions and roles.

IMPLEMENTATION AND TESTING

Trial System Implementation

One of the problems encountered during implementation was due to the mismatch of screen resolution of PDA and the Web page. The problem occurred because the contents were being directly viewed from the server's default browser. The problem was solved by using a built-in browser of the PDA. Setting of wireless configurations presented another problem. Initially, a personal LAN was set up using the 3COM wireless access points. Following this, the wireless network card was registered with the Information Technology department of the American University of Sharjah (AUS), so that the PDAs could access the AUS network. Problems such as conflicting devices and security issues had to be dealt with. The solution was to select the auto channel setting when configuring the access point.

Trial System Testing

The designed system has been implemented using AUS infrastructure and tested by doctors, nurses, and lab assistants of AUS healthcare center. AUS infrastructure provided large number of workstations and wireless points distributed throughout the campus. Two wireless PDAs with wireless jackets are used for testing wireless functions.

Each one of the users logs on to network with his or her unique ID and password. The system first validates the ID and password, and then gives access based on the assigned privileges. The following instance shows how the SQL Server database is accessed to retrieve or modify data.

Doctors have the following options and privileges:

- View Personal Records.
- Update Address and Contact Number.
- Search for a Patient using Patient's last name or ID.
- Search for the general record details of a patient using the unique patient ID.
- View the specific record details for a particular record given the record number and Patient ID.
- Create new records for the patient. View a list of all those patients who are under him.
- Can add a new patient to the Patient Database.

The options given to nurses are:

- Update, Create Personal Records.
- Update, Create Patient Records.
- View Patient Table.
- Add New Patient.

The authorized privileges for lab assistants are:

- View Personal Records.
- Can update Address and Contact Number.
- Register Test Results for a patient by entering the patient ID number and the record number view the test results for a patient by entering the patient ID number and the record number.

When the nurse chooses to view the patient table, a list of all the patients is generated. This list contains the Patient ID, First Name, Last Name, and Contact Number of the patient. Figure 7 shows the system response for such a query. It shows the number of record for the patient, data of the visit, visited physician ID and Name, and the diagnosis by each physician.

The ASP Code accesses the database from the SQL Server because the nurse has been authenticated by the login process. Following this, a selected set of columns is retrieved from the database and is displayed to the end user. All of the above activities are tested using wireless PDAs. Also, it is found that they can be carried out and viewed just as it can be done using wired servers.

The ASP Script appends the entered data into the Record Table of the Hospital Database. Furthermore, the ASP Script also stores the Doctor

Figure 7. System output using patient's ID key search

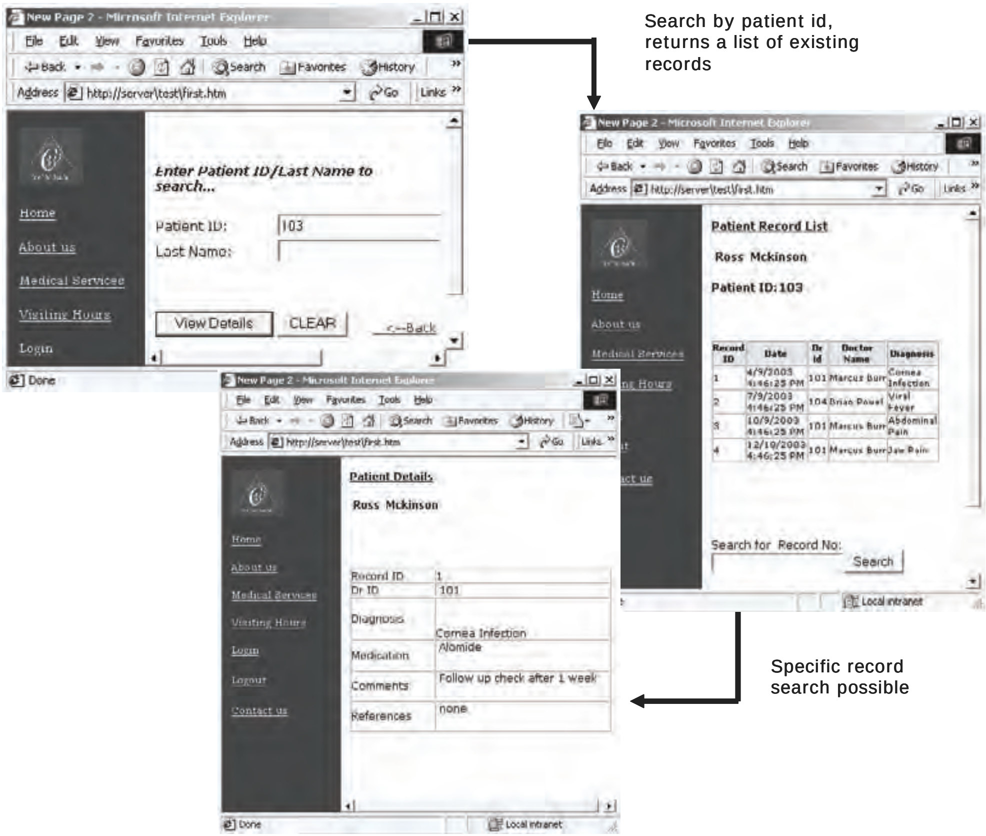
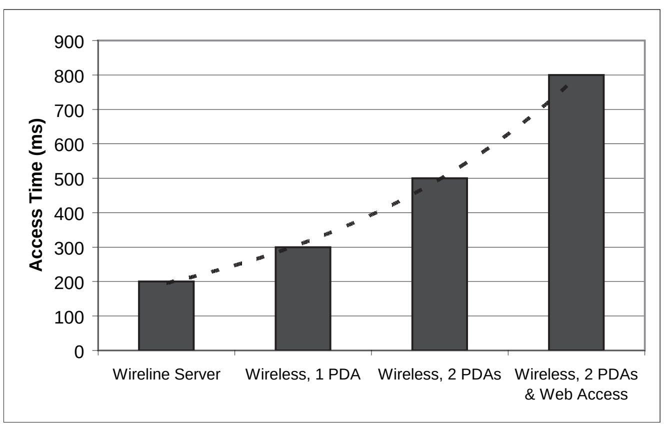


Figure 8. Simultaneous access times through Web-based database using wired and wireless PDA combinations



ID (which is a session variable) and the System date into the Record Table. It is worth mentioning that once a patient record is accessed for editing, the database locks the file and it can not be updated by the another user until the current session is finished. However, the stored version of record can be viewed while the update session in progress.

SYSTEM PERFORMANCE ANALYSIS

The system is implemented on AUS network, which covers the entire campus with wired and wireless links. Two wireless PDAs are used for measuring performance of the system. Figure 8 indicates the access time performance values measured during the trial. The access time values were measured while the PDAs were trying to access the server simultaneously. Even though the manufacturer specifications indicate that wireless devices should work with up to 100 meter distance from wireless access points, our tests indicated that the maximum reliable distance from wireless points are around 60 meters.

Our tests also indicated that network traffic over the intranet, which is not related to electronic patient record database increased the previously mentioned access times about 20% as shown in Figure 9.

System access failure rate was studied as a function of PDA-to-access point distance. The results are shown is Figure 10. As it can be seen, the success rate drops drastically as the distance approaches 60 meters.

Figure 10. System access failure rate as a function of distance for different numbers of mobile users

CONCLUSION

A wireless PDA-based patient's information system was designed, implemented, and tested. Hospital personnel can access, create, and update the patient's record using standard Internet browsing method through wireless mobile devices. The system gives the health personnel the mobility feature where they can check their patient's record from anywhere in the hospital using the wireless PDA or using Internet browser from any other location. Limited system performance tests indicated

Figure 9. Impact of concurrent Internet access on the access times through Web-based database

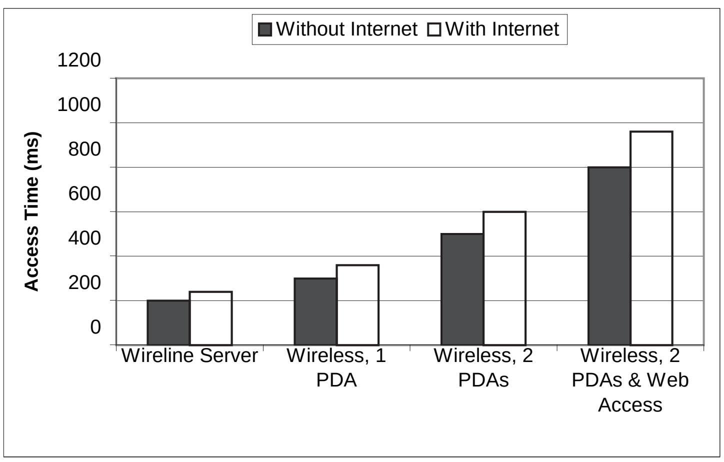
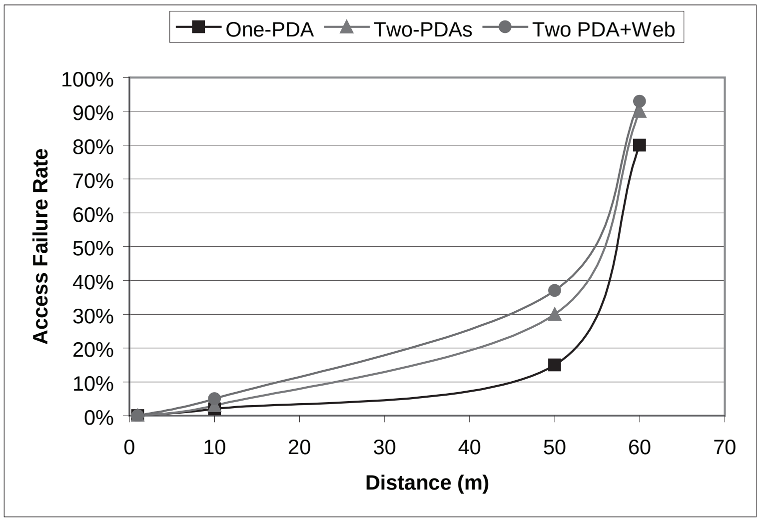


Figure 10. System access failure rate as a function of distance for different numbers of mobile users



satisfactory performances as long as the hospital's environment is well covered with wireless access points with access distance not exceeding 60 meters. Although it is not implemented in the test system, the users indicated that the inclusion of a practical messaging system similar to pagers will make the system even more useful.

ACKNOWLEDGMENT

The authors would like to express their thanks to Mr. Rahul Saharia, Ms. Sharmila Gupta, Ms. Suvina Singh and Ms. Tara Thomas for their valuable contributions to the project, AUS-IT Center, AUS Medical Center, and local representatives of HP for providing equipment support for the project.

REFERENCES

Alan, T., Lefor, M. D, & Lefor, M. K. (2003). Wireless Computing in Healthcare. *Current Surgery, 60*(4), 477-479.

Ammenwerth, E., Buchauer, A., Bludau, B., & Haux, R. (2000). Mobile information and communication tools in the hospital. *International Journal of Medical Informatics*, *57*(1), 21-40.

Demiris, G. (2004). Electronic home healthcare: Concepts and challenges. *International Journal of Electronic Healthcare*, *1*(1), 4-16.

Garcia, J., Martinez, I., Sornmo, L., Olmos, S., Mur, A., & Laguna, P. (2002). Remote processing server for ECG based clinical diagnosis support. *IEEE Transactions on Information Technology in Medicine*, *6*(4), 277-284.

Hung, K., & Zhang, Y. (2003). Implementation of a WAP-based telemedicine system for patient monitoring. *IEEE Transactions on Information Technology in Medicine*, *7*(2), 101-107.

Lin, B., & Vassar, J. A. (2004). Mobile healthcare computing devices for enterprise-wide patient data delivery. *International Journal of Mobile Communications*, *2*(4), 343-353.

Liu, C., Long, A., Li, Y., Tsai, K., & Kuo, H. (2001). Sharing patient care records over the World Wide

Web. *International Journal of Medical Informatics*, *61*(2-3), 189-205.

Lu, Y., Xiao, Y., Sears, A., & Jacko, J. (2005). A review and a framework of handheld computer adoption in healthcare. *International Journal of Medical Informatics*, *74*(5), 409-422.

Metaxiotis, K. (2005). E-health vs. KM-based health: A dilemma in researchers' minds.' *International Journal of Electronics Healthcare*, *1*(3), 303-315.

Nazeran, H., Setty, S., Haltiwanger, E., & Gonzalez, V. (2004). A PDA-based flexible telecommunication system for telemedicine applications. *Proceeding of the IEEE Engineering in Medicine and Biology Society Conference* (pp. 2200-2203), EMBC 2004.

Riikonen, P., Boberg, J., Slakoski, T., & Vihinen, M. (2002). Mobile access to biological databases on the Internet. *IEEE Transactions on Biomedical Engineering*, *49*(12), 1477-1479.

Safran, C., & Goldberg, H. (2000). Electronic patient records and the impact of the Internet. *International Journal of Medical Informatics*, *60*(2), 77-83.

Van Ginneken, A. M. (2002) 'the computerized patient record: balancing effort and benefit' International Journal of Medical Informatics, vol. 65, pp. 97-119.

Wang, J., Hongwei, D. (2005). Setting up a Wireless Local Area Network (WLAN) for a healthcare system. *International Journal of Electronic Healthcare*, *1*(3), 335-348.

Wickramasinghe, N., & Misra, S. K. (2004). *A wireless trust model for healthcare*. *International Journal of Electronic Healthcare*, *1*(1), 60-77.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 1, Issue 4, edited by J. Tan, pp. 1-14, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 3.11
Mobile Clinical Learning Tools Using Networked Personal Digital Assistants (PDAs)

Bernard Mark Garrett
University of British Columbia, Canada

INTRODUCTION

The School of Nursing at the University of British Columbia has more than 300 nursing students engaged in supervised clinical practice in hospital and community settings around Vancouver. Likewise, the Faculty of Medicine has more than 200 medical students undertaking supervised clinical experience locally and remotely in the Prince George and Vancouver Island regions. The management of these clinical experiences and the promotion of learning while in an active clinical setting is a complex process.

BACKGROUND

Supporting the students at a distance while undertaking their clinical experience is particularly resource-intensive. It requires the creation and maintenance of good communication links with the clinical and administrative staff, active management, clinical visits from faculty, and the provision and management of remotely based resources. However, there were few existing resources that helped to contextualize and embed clinical knowledge in the workplace in the practice setting (Landers, 2000). A technological solution was developed and implemented using several clinical applications designed for use on personal digital assistants (PDAs).

MOBILE CLINICAL LEARNING TOOLS

A suite of PDA-based tools were created for a pilot study with the involvement of nursing and medical students during the academic year of 2004-2005 to achieve the following objectives:

- To demonstrate the potential use of mobile networked technologies to support and improve clinical learning.
- To develop and evaluate a range of mobile PDA tools to promote reflective learning in practice and to engage students in the process of knowledge translation.

- To develop and evaluate a suite of pedagogic tools that help contextualize and embed clinical knowledge while in the workplace.
- To evaluate the value of networked PDA resources to help prevent the isolation of students while engaged in clinical practicum.

The tools developed provide a mobile clinical learning environment incorporating an e-portfolio interface for the Pocket PC/Windows Mobile (Microsoft, 2004) operating system. They were implemented on i-mate PDAs equipped with GSM/GPRS (Global System for Mobile Communications/General Packet Radio Service; GSM World, 2002). This platform offered considerable flexibility for the project. It supported the use of cellular telephone connectivity and Pocket Internet Explorer Web browser (which has a full Internet browser with support for HTML, XML/XSL, WML,cHTML, and SSL); the i-mate device had sufficient memory for the storage of text, audio, image, and video data, with a large screen and a user-friendly interface with an integrated digital camera.

The tools included a mobile e-portfolio (with a multimedia interface) designed to promote professional reflection (Chasin, 2001; Fischer et al., 2003; Hochschuler, 2001; Johns, 1995; Kolb, 1984). These mobile learning tools were designed to promote the skills of documentation of clinical learning, active reflection, and also to enable students to immediately access clinical expertise and resources remotely. Community clinical placements are being used for the testing domain, as there are currently no restrictions on using cellular network technology in these areas, whereas this is currently restricted in acute hospital settings in British Columbia and many other parts of the world.

THE PDA INTERFACE DESIGN

The main interface to the clinical tools was based on a clinical e-tools folder on the Pocket PC containing icon-based shortcuts to a number of specific applications (Figure 1).

The clinical e-portfolio tool represented the major focus for the project, allowing the student to access clinical placement information; log clinical hours; achieve clinical competencies; record portfolio entries in the form of text, pictures, or video clips; and record audio memos. This provides the user with a very adaptable interface, allowing them to choose how they input data. For example, a text-based entry describing a clinical procedure may be accompanied by a picture or audio memo.

The e-portfolio tool also incorporates a reflective practice wizard promoting the students to work through the stages of the Gibbs reflective cycle (Gibbs, 1988) when recording their experiences. This wizard also allows students to record their experiences with multimedia, including text, audio, digital images, or video input. Once the data have been recorded in the e-portfolio, they can be synchronized wirelessly (using the built-in GSM/GPRS or Bluetooth connectivity) with a Web-based portfolio. The data then can be reviewed and edited by the student or by clinical tutors.

The other icons represent the following applications:

- The synch portfolio icon initiates synchronization of the content of the student's e-portfolio on the PDA with that of a remote server.
- The University of British Columbia (UBC) library icon presents a shortcut to a Pocket Internet Explorer Web access to the UBC library bibliographic health care database search (CINAHL, Medline, etc.).

Figure 1. Screenshot of the clinical e-tools folder

- The Pocket Explorer icon presents a shortcut to Pocket Internet Explorer for mobile Web access.
- The e-mail icon presents a shortcut to the Pocket PC mobile e-mail application.

The other icons on the screen (Diagnosaurus, ePocrates, etc.) represent third-party clinical software that was purchased and loaded onto the PDAs in order to support the students learning in the clinical area (e.g. a drug reference guide).

FUTURE TRENDS

In the future, the PDA will provide a one-stop resource to support clinical learning. Students also will be able to examine their learning objectives, record their achievements, and record notes/memos attached to specific clinical records for later review. Where students have particular concerns or questions that cannot be answered immediately in the clinical area, they will be able to contact their supervisors or faculty for support using e-mail, cell phone, or multimedia messaging service (MMS) communications.

The use of multimedia in PDA interfaces is likely to become much more widespread as the cost of these devices reduces and they become more accessible to a wider spectrum of the population. This already is occurring with the merging of cell phone and PDA technologies and the uptake of MMS and use of audio and video data entry on mobile devices (deHerra, 2003).

In the long term, multimedia mobile learning tools will encourage a more structured process of professional reflection among students in supervised clinical practice (Conway, 1994; Copa et al., 1999; Palmer et al., 1994; Reid, 1993; Sobral, 2000). When unexpected learning opportunities arise, students will be able to quickly review online materials in a variety of formats and prepare for their experience, record notes, record audio memos or images during their practice, and review materials following their experience.

An expansion in the use of such mobile clinical learning tools is envisaged, and there is considerable scope for the widespread application of such tools into areas where students are engaged in work-based learning. We are likely to see the integration of these technologies into mainstream educational practice in a wide variety of learning environments outside of the classroom.

CONCLUSION

The value of these new tools to students in clinical practice remains to be demonstrated, as the evaluation stage of the project has yet to be completed. The project also has highlighted the necessity of addressing some of the weaknesses of current PDA design, such as the small display screen and the need for more built-in data security. However, initial feedback appears promising, and the interface design appears to promote reflective learning in practice and engage students in the process of knowledge translation.

REFERENCES

Chasin, M.S. (2001). Computers: How a palm-top computer can help you at the point of care. *Family Practice Management, 8*(6), 50-51.

Conway, J. (1994). Profiling and nursing practice. *British Journal of Nursing, 3*(18), 941-946.

Copa, A., Lucinski, L., Olsen, E., & Wollenberg, K. (1999). Promoting professional and organizational development: A reflective practice model. *Zero to Three, 20*(1), 3-9.

deHerra, C. (2003). *What is in the future of Windows mobile pocket PCs.* Retrieved August 12, 2004, from http://www.cewindows.net/commentary/future-wm.htm

Fischer, S., Stewart, T.E., Mehta, S., Wax, R., & Lapinsky, S.E. (2003). Handheld computing in medicine. *Journal of the American Medical Informatics Association, 10*(2), 139-149.

Gibbs, G. (1988). *Learning by doing. A guide to teaching and learning methods.* Oxford: Oxford Polytechnic.

GSM World. (2002). GSM technology: GPRS platform? Retrieved June 24, 2004, from *http://www.gsmworld.com/technology/sms/intro.shtml*

Hochschuler, S.H. (2001). Handheld computers can give practitioners an edge. *Orthopedics Today, 21*(6), 56.

Johns, C. (1995). The value of reflective practice for nursing. *J. Clinical Nurs, 4*, 23-60.

Kolb, D.A. (1984). *Experiential learning*. Englewood Cliffs, NJ: Prentice Hall.

Landers, M.G. (2000). The theory-practice gap in nursing: The role of the nurse teacher. *Journal of Advanced Nursing, 32*(6), 1550-1556.

Microsoft. (2004). *Windows mobile based pocket PCs.* Retrieved August 10, 2004, from http://www.microsoft.com/windowsmobile/pocketpc/ppc/default.mspx

Palmer, A., Buns, S., & Bulman, C. (1994). *Reflective practice in nursing: The growth of the professional practitioner.* Oxford: Blackwell Science.

Reid, B. (1993). "But we're doing it already": Exploring a response to the concept of reflective practice in order to improve its facilitation. *Nurse Ed Today, 13*, 305-309.

Sobral, D.T. (2000). An appraisal of medical student reflection-in-learning. *Medical Education, 34*, 182-187.

KEY TERMS

Bluetooth: A short-range wireless radio standard aimed at enabling communications between digital devices. The technology supports data transfer at up to 2Mbps in the 2.45GHz band over a 10m range. It is used primarily for connecting PDAs, cell phones, PCs, and peripherals over short distances.

Digital Camera: A camera that stores images in a digital format rather than recording them on light-sensitive film. Pictures then may be downloaded to a computer system as digital files, where they can be stored, displayed, printed, or further manipulated.

e-Portfolio: An electronic (often Web-based) personal collection of selected evidence from coursework or work experience and reflective commentary related to those experiences. The e-portfolio is focused on personal (and often professional) learning and development and may include artefacts from curricular and extra-curricular activities.

General Packet Radio Service (GPRS): A standard for wireless communications that operates at speeds up to 115 kilobits per second. It is designed for efficiently sending and receiving small packets of data. Therefore, it is suited for wireless Internet connectivity and such applications as e-mail and Web browsing.

Global System for Mobile Communications (GSM): A digital cellular telephone system introduced in 1991 that is the major system in Europe and Asia and is increasing in its use in North America. GSM uses Time Division Multiple Access (TDMA) technology, which allows up to eight simultaneous calls on the same radio frequency.

i-Mate: A PDA device manufactured by Carrier Devices with an integrated GSM cellular phone and digital camera**.** The device also incorporates a built-in microphone and speaker, a Secure Digital (SD) expansion card slot, and Bluetooth wireless connectivity.

Personal Digital Assistant (PDA): A small handheld computing device with data input and display facilities and a range of software applications. Small keyboards and pen-based input systems are commonly used for user input.

Pocket PC: A Microsoft Windows-based operating system (OS) for PDAs and handheld digital devices. Versions have included Windows CE, Pocket PC, Pocket PC Phone Edition, and Windows Mobile. The system itself is not a cut-down version of the Windows PC OS but is a separately coded product designed to give a similar interface.

Wizard: A program within an application that helps the user perform a particular task within the application. For example, a setup wizard helps guide the user through the steps of installing software on his or her PC.

This work was previously published in Encyclopedia of Human Computer Interaction, edited by C. Ghaoui, pp. 404-407, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.12
Choosing Technologies for Handheld and Ubiquitous Decision Support

Darren Woollatt
University of South Australia, Australia

Paul Koop
University of South Australia, Australia

Sara Jones
University of South Australia, Australia

Jim Warren
The University of Auckland, New Zealand

ABSTRACT

Wireless, handheld devices are becoming increasingly popular in health care settings, but the full potential of their role in patient-specific decision support remains to be achieved. This article presents a multicriteria framework for choosing technologies apropos to handheld and ubiquitous decision support architecture. This framework is illustrated through architectural middleware choices made in the context of a podiatry and diabetes care network. Performance issues are found to be very important in the handheld space, and minor aspects of connectivity and other constraints drive significant changes in choices of architectural approach. The resulting architecture employs layers, including serialized objects, XML payloads, event notification, Web services, and dynamic class loading, with the mix varying among the system interfaces. The overall recommendation is that organizations wishing to fully exploit mobile technology must use a flexible policy and pursue a process of technology choice that is scenario-based and iterative to take into account discoveries from prototyping and field-test experience.

INTRODUCTION

Handheld computers acting as personal digital assistants (PDAs) are growing in popularity in health care. They are increasingly trusted (particularly as sources for reference material), used, and considered to be efficient (Cimino & Bakken, 2005). A mid 2004 survey (Grasso, Yen, & Mintz, 2005) found 52% of medical students reporting handheld computer use, with drug reference and clinical calculators the major clinical applications of the technology. Characterizing the use and capability of PDAs is an exercise in measuring the position of a moving target, but the dominant trend sees students as a key user group and reference for education as a key application area. PalmCIS (Chen, Mendonca, McKnight, Stetson, Le, & Cimino, 2004) illustrates the less common use of mobile technology as a terminal for viewing patient data. In a nursing context, Chang, Lutes, Braswell, and Nielsen (2006) found that mobile technology integrated with a hospital's mainframe system improves the communications between shifts. Lu, Xiao, Sears, and Jacko (2005) and Fischer, Stewart, Mchta, Wax, and Lapinsky (2003) also have reviewed handheld applications in health care, and Wu and Straus (2006) and Lane, Heddle, Arnold, and Walker (2006) have conducted systematic reviews. Wu and Strauss found limited, but supportive, evidence that handhelds improved documentation taken by physicians. The more functionally ambitious handheld solutions in health care are generally "home-grown" (i.e., purpose-built for a specific application context).

Handheld computing and mobile communications move us closer to a ubiquitous computing environment, where the notions of "a computer" or "the information system" become less central to the attention of the end user and where many often loosely coupled subsystems collaborate to achieve an overall goal. At the University of South Australia (UniSA), we have been following a vision of ubiquitous decision support for chronic disease management with an emphasis on diabetes and foot care that integrates the roles of clinical decision support, provider education, and patient education (Warren, Lundstrom, Osborne, Kempster, Jones, Ma, & Jasiunas, 2004). Two key elements of this vision have been a handheld data collection and decision support application for students at the UniSA Podiatry Clinic (Lundström, Warren, Jones, Chung, & Jasiunas, 2003); and an individualized Web-based consumer diabetes information service (Ma, Warren, Phillips, & Stanck, 2005).

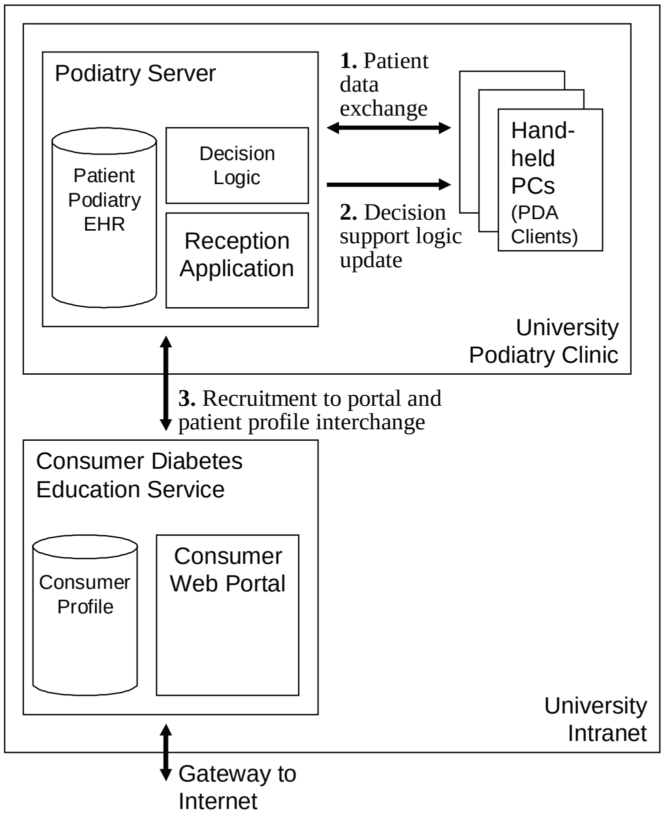
In the context of ongoing iterative development of handheld and ubiquitous decision support, one is faced with a diversity of architectural options in an environment of rapidly changing technology. The remainder of this paper presents a framework for consideration of relevant technology choices, with particular focus on middleware architecture, which is illustrated in terms of a cycle of development in the UniSA podiatry/diabetes management environment. The outcomes highlight the sensitivity of technology appropriateness to existing constraints and the ongoing relevance of performance.

METHODS

Setting

A spiral development lifecycle has been pursued with the UniSA Podiatry Clinic from mid 2002. The objectives of a cycle for calendar year 2005 entailed: (a) achieving a production client-server system in the podiatry clinic, allowing students to undertake podiatry assessments using handheld computers with results logged to a central database; (b) active decision support for the podiatry students on the handheld terminal to critique assessment and plan in terms of recorded observations; and (c) interfacing of the podiatry clinic to a diabetes consumer education service such that the clinic acts as a recruitment point for the education

Figure 1. Podiatry/diabetes care network setting (technology choice scenarios numbered)



service, and podiatry observations lead to tailoring of educational priorities. The environment, with the major communication scenarios addressed in the 2005 cycle, is illustrated in Figure 1.

Aspects of the design were fixed at the commencement of the 2005 cycle, constituting developments, investments, and decisions that were undesirable or infeasible to back away from. These fixed aspects of the design include: (a) use of Hewlett-Packard iPAQ pocket PCs with Microsoft Windows operating system as the podiatry clinic terminals; (b) the elements recorded in the podiatry review, user interface design for data capture, and podiatry clinic server database design (encompassing some 200 specific observations, including some free-hand sketches); (c) the use of Bluetooth for communication within the podiatry clinic (due in part to administrative "turf" problems entangling development on 802.11 standards); and (d) the implementation of the Web portal for the consumer diabetes education service to be impacted minimally.

Outside of these fixed constraints, the outstanding architectural decisions concerned the approach to data interchange at the application level among the major system components, including interchange of decision support logic, a realm that can be termed "middleware architecture." All of these communication interfaces were required to support modification of content in the future, but the decision support content was seen as the most volatile, with plans for active iterative refinement of the logic in a future cycle. Experience with an earlier cycle prototype indicated that the data transmission speed performance of the podiatry server to handheld data exchange link was problematic, although a naïve analysis of Bluetooth bandwidth left the source of the problem unclear.

Procedure

Our procedure was informed by Serain's (2002) "criteria based table model" (p. 211) and the practice of defining scenarios that are to occur within the architecture, as demonstrated by Goedicke and Zdun (2002) and Dobrica and Niemelä (2002).

Major architectural options were identified and each assessed in terms of a number of criteria. It should be noted that the architectural options are not, in general, mutually exclusive, and assessment on criteria is qualitative. The purpose of the resulting option-×-criteria grid is to inform the synthesis of candidate designs, not to support a mechanistic choice of a single option. Each cell of the grid should be populated with an explanation of the qualitative judgement and possibly include a number of unresolved issues. This procedure is iterative. Review of the option-×-criteria grid can result in options being dismissed, merged, or split.

Criteria

The decision criteria group was separated into what classically is called "functional" and "'non-functional" requirements, although we prefer to designate the criteria groups as *case-specific* and *generic*. One case-specific criterion is identified for each major system interfacing scenario. In the current setting, these scenarios are:

1. Distributing and synchronizing patient data (around the podiatry clinic);
2. Ability to transfer decision support logic (from the podiatry server to handheld PCs); and
3. Ability to transfer profiling information (to focus consumer diabetes education, moving data between the podiatry clinic and Web-based consumer education setting).

The generic criteria are fairly standard aspects of systems integrity, but many take on a degree of specific meaning in the health computing context:

- **Security:** Acceptable handling of patient data (notably HIPAA compliance in the United States, although not in demonstration context);
- **Performance:** Chiefly latency, and also reliability;
- **Maintainability:** Outlook for the long-term feasibility of the solution in terms of a continuing user community and availability of technical staff for maintenance;
- **Maturity/vailability:** Presence of standards (de facto or otherwise) to provide stability and practicality of acquisition;
- **Flexibility:** Supporting explicitly expected changes and reasonable estimation of long-term change (A special bonus is if one technology is flexible enough to encompass multiple requirements of the system, making the system simpler overall.); and
- **Feasibility:** To further diffuse "tick-box" thinking, a summative assessment that the technology is actually practical and can fulfil the requirements in question.

RESULTS

Technology Assessment Decisions

Initial brainstorming in the development team (authors Darren Woollatt, Jim Warren, and Paul Koop) revealed seven options for further exploration from among a diverse range of health-specific initiatives, relatively modern generic

Figure 2. Abbreviated option—criteria grid for middleware architecture technology choices

Criterion	HL7	openEHR	iROS	SOAP	CORBA	DCOM	RMI
Scenario 1	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scenario 2	Yes, Arden	N/A	1, 2	1, 2	1	1	1
Scenario 3	Yes	Yes	Yes	Yes	Yes	3	Yes
Security	Add on	Add on	Add on	Add on	Built in	Built in	Built in
Performance	4	4	4	4	4, well-studied	4, well-studied	4
Maintainability	5	5	OK	Good	OK	3	Good, part of JDK
Maturity/ availability	5, 6	5, 6	6	6	6	3	6
Flexibility	Suits purpose	Suitable for scenario 1 and 3	Very flexible	General	General	General	General
Feasibility	7	7	Good	Good	8	3	Good

[1]*Can transfer guideline representation for interpretation on client*

[2]*Can serialize object with decision logic for execution on client*

[3]*Operating system and language limitations of DCOM (or .NET) are a concern; current Diabetes Consumer Education Service is on Linux*

[4]*A concern for Scenario 1—must be measured in context*

[5]*Steep learning curve and technology is rapidly changing, although adherence to a standard is a plus.*

[6]*Open with freely available implementations and user communities; somewhat more restricted to members for HL7. iROS community rather small but deemed adequate*

[7]*HL7 v3 requires process using RIM to create new messages for our highly specialized application; somewhat better supported and more approachable process using openEHR archetypes*

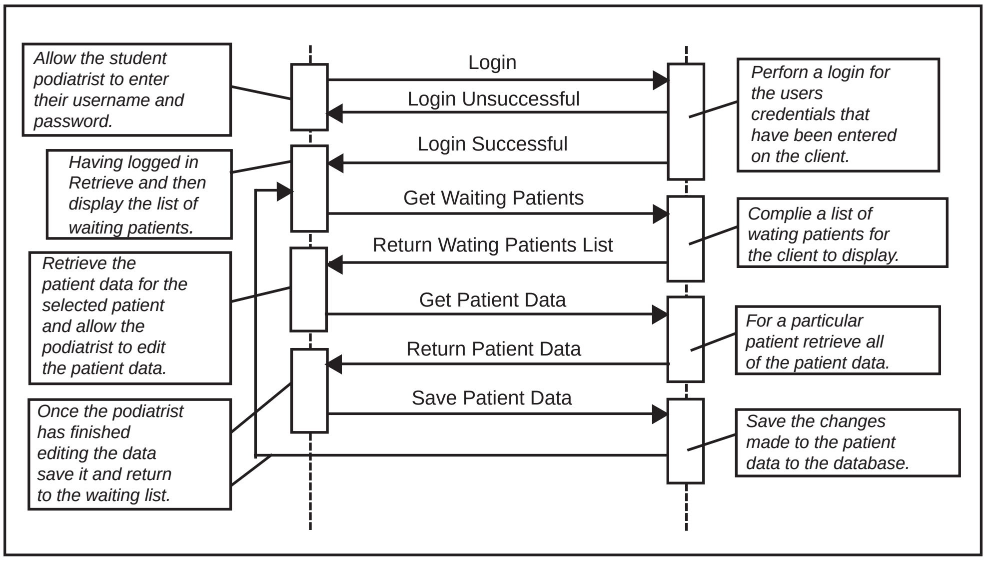
[8]*Possible, but over-spec to needs in absence of multilanguage and legacy system integration motivators of CORBA*

solutions, and solutions of historical interest for comparison. Probably the least generally well-known option under consideration was iROS (Open Source Technology Group, 2006), which had a track record of use for enabling "interactive rooms" through a messaging service and was familiar to the authors. An abbreviated first-cut option-×-criteria grid is shown in Figure 2. The full grid, as used by the development team, occupies a color A3 sheet in 10-point type with 40 footnotes on ancillary pages.

With respect to the first scenario, remote method invocation (RMI) and iROS emerged as strong options based on transfer of serialized objects. Simple object access protocol (SOAP) (W3C, 2004) also had promise. It also had an edge in extensibility, if serialized objects were not used, but required use of XML, which seemed like it might be a performance issue, especially on the client (handheld) side, where computing power for parsing a large XML payload is quite limited. All three options are capable of having XML components to the message, which could then support Health Level Seven (HL7)(2006) or OpenEHR (n.d.) as a further layer.

Performance analysis revealed that a "typical" patient review as a serialized object was 564kB in size and required about 14 seconds to transfer over our BlueTooth connection from the client to server using iROS. However, application of a

Figure 3. Sequence diagram for podiatry application client/server interaction

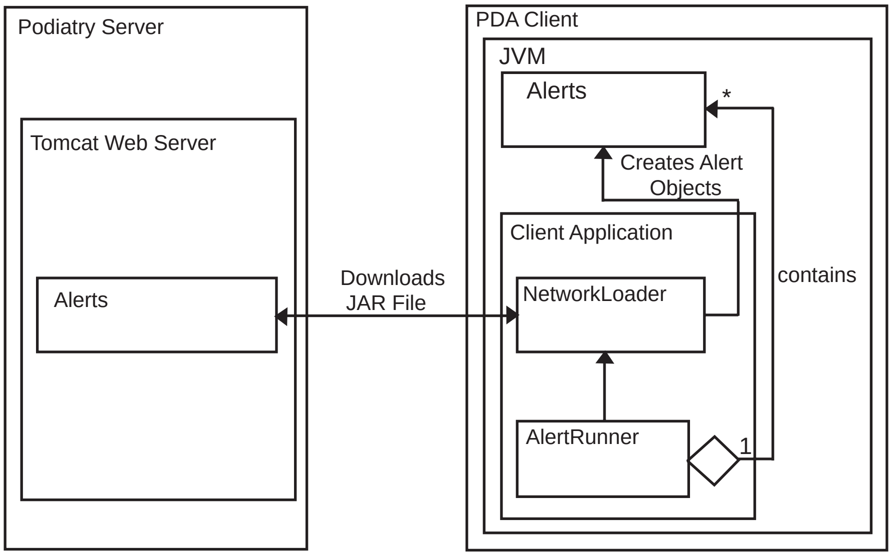


compression routine from the Sun Microsystems Java Development Kit JDK reduced this to 14.5kB and brought the transfer time (with the negligible compression time included) down to an acceptable 0.85 seconds. iROS was compared to SOAP and found to be about 2.5 times faster for the same test object. Encryption of several sensitive fields (again using a JDK routine) also was included with no observable performance impact, yielding an overall acceptable solution to scenario 1. Layering of HL7 or openEHR was not pursued in this scenario, due to performance concerns, availability issues that parallel scenario 2 and 3 as discussed below, and minimal perceived relevance for a system as tightly coupled as the podiatry server and handheld client. The iROS solution was chosen over an RMI solution largely due to developer familiarity.

With respect to scenario 2, SOAP emerges as particularly promising for invocation of decision support methods on the server from the handheld devices. However, a significant constraint in this scenario is that the users may wander outside of BlueTooth range at the time when decision support alert logic must be checked. (This is done interactively in the consulting rooms, whereas scenario 1— completion of the previous patient and start of the next—is done in close physical proximity to Clinic reception desk and the server.) As such, we are restricted to methods that actually transfer the decision support logic to the client at some time between user login and start of processing on the first patient of the user session.

One promising option for transfer of decision support logic was to use a representation scheme, such as the Arden syntax (Jenders, 2006) of HL7. However, this requires the client to interpret logic statements in a high-level specification language, raising issues around performance, as well as the question of availability of such an interpreter to run on a handheld PC. Given the difficulties of an interpreted logic solution, the technology choice taken was to write the decision logic in Java such that all decision alerts shared a common Java interface. Central update of decision logic is achieved by redirecting the handheld client's class loader to the server (see Figure 4), exploiting the compactness of a Java archive (JAR) file to move the logic as Java byte code.

Figure 4. Architecture for updating client decision support logic



In scenario 3, SOAP wins out. While the podiatry server and Diabetes Consumer Education Service currently reside on a common university intranet, the probability of the clinic and Web-education services becoming more widely distributed is high, and hence the ability of the SOAP protocol to cross organizational firewalls provides maximum maintainability of the solution.

The case is strong for layering either HL7 or openEHR onto the SOAP message that transfers the patient profile (observations collected in the podiatry clinic of relevance to diabetes education; some 40 data items in the current data sets). The application fits well with the model of either an HL7 message or an openEHR EHR extract. At the time of implementation, however, neither HL7 v3 nor openEHR were offering a convenient Java toolkit; this situation is changing rapidly (see the Future Research section). The medium-term solution implemented was to provide the profile as an XML payload, but to leave the reconciliation of that payload to the HL7 RIM or an openEHR archetype for future extension of the system. W3C standards of XML signatures (Eastlake & Reagle, 2006) and XML encryption (Reagle, 2006) are applied to meet the security requirements of scenario 3.

Resulting System

With the technology assessment decisions taken as per above, the result is a close coupling of the podiatry server to the PDA clients (handheld PCs) and a looser coupling of the podiatry clinic (through the podiatry server) to the Diabetes Consumer Education Service. The PDA clients communicate to the server over BlueTooth. A sequence of iROS messages, as illustrated in Figure 3, keeps the client up-to-date with the available patients in the waiting room (tracked via a Microsoft .NET application local to the podiatry server) and transfers patient data prereview and postreview. The PDA client implements a common user interface for decision support alerts, the decision logic, which is loaded from the podiatry server when a user logs onto the podiatry application on the PDA (as per Figure 4).

At the end of the podiatry review for patients with diabetes, the PDA client prompts the user to ask whether the patient would like access to the

Consumer Diabetes Education Service. When the podiatry server receives the message to update the patient database and sees the request for an education service account, a SOAP request is issued for an account to be created on the consumer portal. Upon receiving the SOAP result, the server prints a hardcopy invitation letter for the patient with login information, including their specific username and password as provided by the Consumer Diabetes Education Service. Upon login to the portal, the consumer is prompted for consent for the transfer of information from the podiatry clinic, at which time a getProfile method is invoked from the education service to the podiatry server, causing clinical data to be extracted from the patient's podiatry reviews. This data is composed into an XML snippet by the podiatry server and returned to the education service for parsing into its own consumer/patient profile database.

DISCUSSION

We have reached an age that is ripe for innovation in handheld and ubiquitous information technology. It has never been easier to integrate applications across a range of platforms to achieve decentralized behavioral changes, such as improved chronic disease management. Specifically, growing acceptance of handheld computing in the health workplace, ubiquity of short-range radio communications, and the advent of Web services standards are making it feasible and inexpensive to field health data collection and decision support tools. Moreover, it is increasingly easy to "tack on" new avenues for the exploitation of existing functionality (taking an existing application and putting it on the Web, adding a mobile portal, mounting it as a Web Service for other applications, etc.). And many of these integration solutions are well packaged for programmers using common development environments, such as JDK and Microsoft .NET.

The range of available technology choices presents in itself a challenge. An open-minded, explicit and semistructured (but not overstructured) process of technology evaluation is warranted. Mandating a technology choice will cull potential opportunities by making them infeasible for the present, where a more flexible policy could allow them. The process of technology choice must be iterative to take into account discoveries from prototyping and field test experience, even if these emerge uncomfortably late in the project schedule (better late, or even cancelled, than an on-time, on-budget failure). Choices must be scenario based, not just generic criteria based, as subtly different constraints can make major differences in the feasibility or acceptability or particular technology choices.

Performance is a particularly sensitive issue in that technologies can yield surprisingly high latencies and surprising variances in performance among technologies and their implementations, latencies that can render a solution unusable under field conditions. This problem is acute with low-powered mobile devices, especially when using advanced solutions recently ported from mainstream computer systems. While performance of mobile devices will continue to improve, one can expect it to continue to lag the performance of nonmobile counterparts.

The issues of choice are well illustrated in our experiences with middleware architecture. A show stopper emerged on the issue of disconnected operation (for access to decision support alert methods), removing all choices based on remote method invocation from one of our scenarios. Latency was a strong issue in another scenario, driving a technology choice (and also requiring a well positioned use of compression). In addition, we placed a high value on programmer experience and toolkit availability to go ahead with what was known and involved a short learning curve. We found that very different middleware architecture technology choices came through as appropriate for each of our three major system interfacing scenarios.

Many (if not most) experienced developers will disagree with our specific choices. We fully admit to our own biases, idiosyncratic experience levels, and valuation of the features and limitations of certain options. The important message is not in our choices, but the process of choice and acknowledgment of the diversity of outcomes, even within the same development group, in differing scenarios with differing constraints.

An interesting feature of our solutions for the reported development cycle is that we did not use some of the obvious health-specific technology standards. However, we left room for their adoption in the near future.

FUTURE DIRECTIONS

We may be poised for a time where the easy uptake of health informatics standards begins to parallel their mainstream counterparts, such as SOAP and other XML-based methods. The HL7 Java SIG API (Hendler & Schadow, n.d.) indicates the emergence of appropriately usable mechanisms to integrate HL7 with ubiquitous computing applications, such as the one described in this article. In a case, such as integrating the podiatry clinic with a Consumer Diabetes Education Service, this API should prove an excellent fit. Similarly, recent release of the openEHR specification project release 1.0 (OpenEHR, n.d.) should promote further development of Java APIs on this paradigm. Proposal for a Java Compiler for GELLO,an object-oriented query expression language,(Index of /hl7/arden/2005-05-AMS, n.d.) is also timely. As per our decision support scenario, allowing use of the ubiquitous JVM as the decision support interpreter, rather than needing to port (and get adequate performance from) a high-level guideline expression interpreter, is an excellent solution where system capacity and performance are issues.

A question remains as to whether performance of new health informatics packages will be acceptable for use on small devices, including PDAs. We can expect that these devices will remain dependent on application-specific coupling to servers for many scenarios, where more than a few fields of patient data must be exchanged or where the devices must sometimes operate in disconnected mode.

ACKNOWLEDGMENT

The work reported in this article was partly funded by the Co-operative Research Centre (CRC) for Enterprise Distributed Systems Technology (DSTC) through the Australian federal government's CRC Programme (Department of Education, Science, and Training). All authors were employees or students of the University of South Australia at the time the work was conducted. Special thanks to Helen Banwell and all the UniSA Podiatry Clinic staff and students.

REFERENCES

Chang, K., Lutes, K., Braswell, M., & Nielsen, J. (2006). Nurses' perceptions of using a pocket PC for shift reports and patient care. *International Journal of Healthcare Information Systems and Informatics, 1*(1), 64-77.

Chen, E., Mendonça, E., McKnight, L., Stetson, P., Le, J., & Cimino, J. (2004). PalmCIS: A wireless handheld application for satisfying clinician information needs. *Journal of the American Medical Informatics Association*, *11*(1), 19-28.

Cimino, J., & Bakken, S. (2005). Personal digital educators. *New England Journal of Medicine, 352*(9), 860-862.

Dobrica, L., & Niemelä, E. (2002). A survey on software architecture analysis methods. *IEEE*

Transactions on Software Engineering, 28(7), 638-653.

Eastlake, D., & Reagle, J. (2006). IETF/W3C XML-DSig working group. Retrieved March 13, 2007, from http://www.w3.org/Signature/.

Fischer, S., Stewart, T., Mehta, S., Wax, R., & Lapinsky, S. (2003). Handheld computing in medicine. *Journal of the American Medical Informatics Association 10,* 139-149.

Goedicke, M., & Zdun, U. (2002). A key technology evaluation case study: Applying a new middleware architecture on the enterprise scale (LNCS). In *Proceedings of the Second International Workshop on Engineering Distributed Objects* (pp. 8-26).

Grasso, M., Yen, J., & Mintz, M. (2005). Trends in handheld computing among medical students. In *Proceedings of the AMIA Annual Symposium* (p. 968).

Health Level Seven. (n.d.). Health Level 7. Retrieved March 13, 2007, from http://www.hl7.org/.

Hendler, P., & Schadow, G. (n.d.). Java SIG. Retrieved March, 13, 2007, from http://hl7.org/Special/committees/java/index.cfm#WorkProduct.

Index of /hl7/arden/2005-05-AMS. (n.d.). Retrieved March, 13, 2007, from http://cslxinfmtcs.csmc.edu/hl7/arden/2005-05-AMS/.

Jenders, R.A. (2006). *The Arden syntax for medical logic systems.* Retrieved March 13, 2007, from http://cslxinfmtcs.csmc.edu/hl7/arden/.

Lane, S., Heddle, N., Arnold, E., & Walker, I. (2006). A review of randomized controlled trials comparing the effectiveness of hand held computers with paper methods for data collection. *BMC Medical Informatics and Decision Making, 6*(23).

Lu, Y., Xiao, Y., Sears, A., & Jacko, J. (2005). A review and a framework of handheld computer adoption in health care. *International Journal of Medical Informatics, 74,* 409-422.

Lundström, M., Warren, J., Jones, S., Chung, F., & Jasiunas, M. (2003). A wireless handheld terminal for podiatry that combines routine care, decision support and professional learning. In *Proceedings of the HIC 2003/RACGP 12CC Combined Conferences.*

Ma, C., Warren, J., Phillips, P., & Stanek, J. (2005). Adaptive filtering and prioritising of diabetes consumer information. *International Journal of Medical Informatics, 75,* 577-596.

OpenEHR. (n.d.). OpenEHR home. Retrieved March 13, 2007, from http://www.openehr.org/.

Open Source Technology Group. (n.d.). Interactive room software infrastructure. Retrieved March 13, 2007, from http://sourceforge.net/projects/iros/.

Reagle, J. (2006). W3C XML encryption working group. Retrieved March 13, 2007, from http://www.w3.org/Encryption/2001/.

Serain, D. (2002). *Middleware and enterprise application integration*. London: Springer-Verlag.

W3C (2004). SOAP specification. Retrieved March 13, 2007, from http://www.w3.org/TR/soap/.

Warren, J., Lundström, M., Osborne, D., Kempster, M., Jones, S., Ma, C., & Jasiunas, M. (2004). A multi-interface, multi-profiling system for chronic disease management learning. In *Proceedings of the 37th Hawaii International Conference on System Sciences.*

Wu, R., & Straus, S. (2006). Evidence for handheld electronic medication records in improving patient care: A systematic review. *BMC Medical Informatics and Decision Making, 6*(26).

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 2, Issue 3, edited by J. Tan, pp. 1-11, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 3.13
ECG Diagnosis Using Decision Support Systems

Themis P. Exarchos
University of Ioannina, Greece

Costas Papaloukas
University of Ioannina, Greece

Markos G. Tsipouras
University of Ioannina, Greece

Yorgos Goletsis
University of Ioannina, Greece

Dimitrios I. Fotiadis
University of Ioannina, Greece
Biomedical Research Institute—FORTH, Greece
Michaelideion Cardiology Center, Greece

Lampros K. Michalis
Michaelideion Cardiology Center, Greece
University of Ioannina, Greece

ABSTRACT

ECG is one of the most common signals used in medical practice due to its noninvasive nature and the information it contains. Several systems and various automated approaches have been developed that use computer technology to provide ECG diagnosis. These systems detect abnormalities and other features in the ECG signal and produce a decision which helps the physician when performing diagnosis. ECG decision support systems can serve as a diagnostic tool for specific cardiac anomalies such as myocardial ischaemia and arrhythmia.

ELECTROCARDIOGRAM

The electrocardiogram (ECG) is a clinical test that records the electrical activity of the heart.

ECG is used to measure the rate and regularity of heartbeats as well as the size and position of the chambers, the effects of drugs or devices used to regulate the heart, and the presence of any damage to the heart. An ECG is useful in determining whether a person suffers from a heart disease. If a person has chest pain or palpitations, an ECG will determine if the heart is beating normally. If a person is under medications that affect the heart or if the patient is on a pacemaker, an ECG can provide information on the immediate effects of changes in activity or medication levels. An ECG may be included as part of a routine examination in patients over 40 years old.

ECG ANALYSIS

Automated ECG analysis consists of a series of procedures that can be utilized in order to produce useful clinical information to help the physician to reach a diagnosis concerning the pathophysiological condition of the patient's heart faster and safer. ECG analysis consists of four stages: (a) signal acquisition, (b) processing, (c) feature extraction, and (d) diagnosis. Signal acquisition should fulfill certain specifications concerning the sampling frequency (100Hz to 1 KHz), the resolution (number of bits for each sample, 6 to 16), and the sensitivity, which expresses the signal's amplitude range (usually 5 mV or 6 mV). The digital ECG signal is then processed and filtered to suppress noise and enhance the relevant ECG characteristics.

In the feature-extraction stage, all the relevant ECG characteristics are recognized and some of their features are computed. The extracted features vary from simple ones like the duration and amplitude to the more complex like slopes, intervals, frequencies, or other discriminating indices. These are used in the diagnosis stage since the values of certain features are indicators of the existence of an underlying disease. Apparently, the measurement accuracy (Acc) is vital at this stage, and computerized methods are used to address it efficiently.

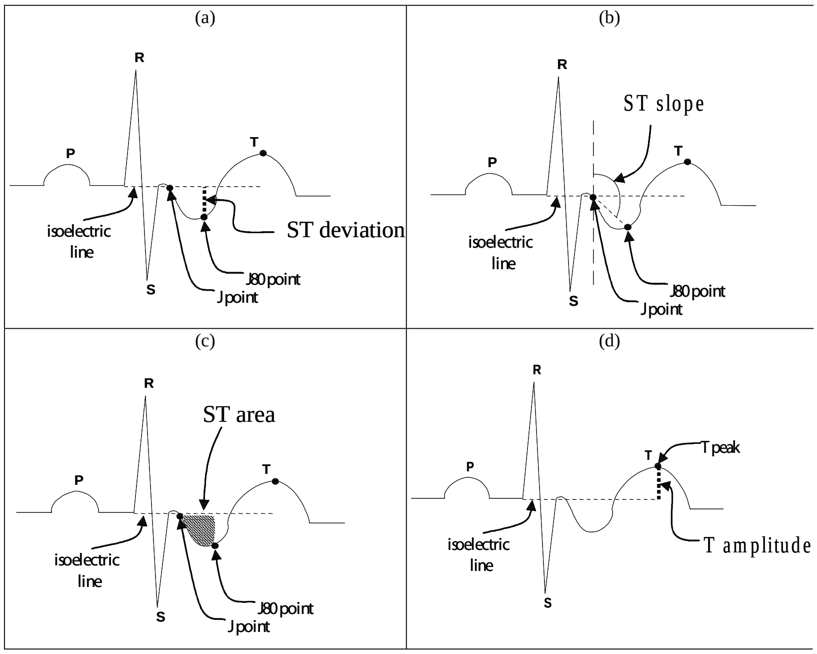
The last stage in the ECG analysis is the diagnosis, where explicit medical knowledge is utilized. Collaboration with medical experts is necessary, and the individual characteristics of each patient complicate the decision-making task. Various automated approaches have been proposed. These systems detect the abnormalities in the ECG and some of them can also produce interpretations for the decisions made. ECG analysis can help diagnose specific cardiac anomalies such as myocardial ischaemia and arrhythmia.

MYOCARDIAL ISCHAEMIA DIAGNOSIS

Myocardial ischaemia is the condition in which oxygen deprivation to the heart muscle is accompanied by the inadequate removal of metabolites due to reduced blood flow or perfusion. This reduced blood supply to the myocardium causes alterations in the ECG signal, such as deviations in the ST segment and changes in the T wave (Goldman, 1982). The detection and assessment of those alterations in long-duration ECGs is a simple and noninvasive method to diagnose ischaemia. In Figure 1, some of the typical ECG features employed for the diagnosis of myocardial ischaemia are shown.

Myocardial-ischaemia diagnosis using the ECG signal can be described as a sequence of two tasks: ischaemic beat detection and ischaemic episode definition. The first is related to the classification of beats as normal or ischaemic. Several techniques have been proposed for ischaemic beat classification, which evaluate the ST-segment changes and the T-wave alterations using different methodological approaches. More specifically, they use parametric modeling (Papaloukas, Fotiadis, Likas, & Michalis, 2002a; Pitas, Strintzis, Grippas, & Xerostylides, 1983), wavelet theory (Senhadji, Carrault, Bellanger, & Passariello,

Figure 1. Typical ECG features extracted for myocardial-ischaemia diagnosis: (a) ST-segment deviation, (b) ST-segment slope, (c) ST-segment area, and (d) T-wave amplitude



1995), a set of rules (Papaloukas, Fotiadis, Liavas, Likas, & Michalis, 2001; Papaloukas, Fotiadis, Likas, Stroumbis, & Michalis, 2002), artificial neural networks (Maglaveras, Stamkopoulos, Pappas, & Strintzis, 1998; Papadimitriou, Mavroudi, Vladutu, & Bezerianos, 2001; Papaloukas, Fotiadis, Likas, & Michalis, 2002b; Stamkopoulos, Diamantaras, Maglaveras, & Stintzis, 1998), multicriteria decision analysis (Goletsis, Papaloukas, Fotiadis, Likas, & Michalis, 2003), genetic algorithms (Goletsis, Papaloukas, Fotiadis, Likas, & Michalis, 2004), and classification using association rules (Exarchos, Papaloukas, Fotiadis, & Michalis, in press).

Cardiac beat detection and classification is a key process in the definition of the ischaemic episodes in the ECG signal. The accuracy of the beat classification influences ischaemic episode definition, in which sequences of ischaemic beats need to be identified. Various methods have been proposed for ischaemic episode detection based on a set of rules (Papaloukas, Fotiadis, Liavas, et al., 2001; Silipo, Taddei, & Marchesi, 1994), artificial neural networks (Papadimitriou et al., 2001; Papaloukas, Fotiadis, Likas, & Michalis, 2002b; Silipo & Marchesi, 1998), fuzzy logic (Vila, Presedo, Delgado, Barro, Ruiz, & Palacios, 1997), and other signal-analysis techniques (Jager, Mark, Moody, & Divjak, 1992; Lemire, Pharand, Rajaonah, Dube, & LeBlanc, 2000; Taddei, Costantino, Silipo, Edmin, & Marchesi, 1995).

The most common techniques for the beat-classification problem are neural- and rule-based approaches. Neural-based approaches have resulted in high performance but exhibit an important drawback due to their inability to provide explanations for the classification decisions. Rule-based approaches exhibit the highly desirable feature of interpreting the decisions, but their performance is not equally satisfactory in terms of accuracy.

A recent knowledge-based system analyzes the ECG signal using a four-stage algorithm (Papaloukas, Fotiadis, Liavas, et al., 2001). The four stages correspond to noise handling and ECG feature extraction, beat classification, window classification, and the identification of ischaemic episode duration. In the first stage, the preprocessing of the ECG recording is performed to achieve noise removal and the extraction of the signal features to be used for beat characterization. In the second stage, each beat is classified as normal, abnormal (ischaemic), or artefact. This information is used in the third stage (the window-characterization stage), where each 30-second ECG window is classified as ischaemic or not. In the fourth stage, the identification of start and end points of each ischaemic episode is performed. The above system is used to detect the overall episodes of ischaemia, but to distinguish also the ST episodes from the T episodes (Papaloukas, Fotiadis, Likas, Stroumbis, et al., 2002). The latter is of great clinical importance since the prognosis of ischaemic episodes with ST-segment changes is worse than those with T-wave alterations only.

Another methodology, which is based on ANNs, has been developed for the detection of ischaemic episodes in long-duration ECG recordings (Papaloukas, Fotiadis, Likas, & Michalis, 2002b). The raw ECG signal containing the ST segment and the T wave of each beat is the input to the beat-classification system, and the output is the classification of the beat. The input to the network is produced using principal component analysis to reduce dimensionality. The network performance in beat classification was tested on a subset of the European Society of Cardiology ST-T Database (*European ST-T Database Directory*, 1991), providing 90% sensitivity (Se) and 90% specificity (Sp). The neural beat classifier is integrated in a four-stage procedure for ischaemic episode detection.

Another technique employs a similar approach as fuzzy logic, multicriteria decision analysis, for ischaemic beat recognition (Goletsis et al., 2003). It deals with assigning objects, namely the cardiac beats, into predefined categories. In order to characterize each beat as ischaemic or not, the beat is compared to already assigned category prototypes. Similarity between each beat and the prototype is computed, and each beat is assigned to the category to which the most similar prototype belongs. The fuzzy pairwise comparison is made for a number of criteria that employ the ST-segment deviation and slope, the T-wave amplitude and polarity, and the patient's age. For each criterion, two parameters are estimated, the similarity and the dissimilarity, while the comparison outcome is aggregated into an indifference index with the use of criterion weights. All the parameters of the method, thresholds and weights, were adjusted using medical experience. A more recent approach uses genetic algorithms for the automatic calculation of the thresholds and the weights (Goletsis et al., 2004).

The latest developed approach for ECG beat classification employs data-mining techniques and especially algorithms that use association rules for the classification (Exarchos et al., in press). A methodology based on a three-stage schema was developed. The three stages correspond to noise handling and ECG feature extraction, feature discretization, rule mining, and beat classification. In the first stage, the preprocessing of the ECG recording is performed to achieve noise removal and the extraction of the signal features that are used for beat characterization. In the second stage, every continuous valued feature is discretized (it is transformed to categorical) in order to be utilized in the next stage. In the third stage, association-rule mining algorithms are applied to generate association rules, which are used to establish the beat-classification model. The methodology was evaluated using data from the European Society of Cardiology ST-T Database, and the obtained sensitivity and specificity were 87% and 93%, respectively.

Other approaches for beat classification are based on the combination of an auto-associative, nonlinear ANN and a radial basis function ANN (Stamkopoulos et al., 1998), ANNs and parametric

Table 1. Comparison of the performance of several methods for myocardial-Ischaemia diagnosis

Method	Se[1] (%)	Sp[2] (%)	Acc[3] (%)
ANN & parametric modelling (Papaloukas et. al., 2002)	81	8 4	
Rule-based (Papaloukas et. al., 2001)	70	63	
Feed forward ANN and nonlinear principal components analysis (Stamkopoulos et. al., 1998).	79	7 5	
Bidirectional associative memories ANN (Maglaveras et. al.,1998)			56
ANN (Classification partitioning-Self organising map) (Papadimitriou et. al., 2001)			74
ANN (Classification partitioning-Self organising map & radial basis function) (Papadimitriou et. al., 2001)			77
ANN (Classification partitioning-Self organising map & support vector machine) (Papadimitriou et. al., 2001)			80
ANN & principal components analysis (Papaloukas et. al., 2002a)	90	90	
Multicriteria decision analysis (Goletsis et. al., 2003)	90	89	
Genetic algorithms & multicriteria decision analysis (Goletsis et. al., 2004)	91	91	
Association rule mining (Exarchos et. al., 2005)	87	93	90

- *1Se: Sensitivity*
- *2Sp: Specificity*
- *3Acc: Accuracy*

modeling (Papaloukas, Fotiadis, Likas, & Michalis, 2002a), bidirectional associative memories (Maglaveras et al., 1998), and the Kohonen self-organizing map algorithm combined with radial basis functions or support vector machines (Papadimitriou et al., 2001).

In order to evaluate the performance of automated systems for myocardial-ischaemia diagnosis, a standard reference database has been developed: the European Society of Cardiology ST-T Database (*European ST-T Database Directory*, 1991).

In Table 1, we can see the performance of several systems designed for beat classification.

ARRHYTHMIA DIAGNOSIS

Arrhythmia can be defined as either an irregular single heartbeat (arrhythmic beat), or as an irregular group of heartbeats (arrhythmic episode). Arrhythmias can affect the heart rate, causing irregular rhythms, such as slow or fast heartbeat. Arrhythmias can take place in a healthy heart and be of minimal consequence (e.g., respiratory sinus arrhythmia, which is a natural periodic variation in heart rate, corresponding to respiratory activity), but they may also indicate a serious problem that may lead to stroke or sudden cardiac death (Sandoe & Sigurd, 1991).

Several researchers have addressed the problem of the automatic detection and classification of cardiac rhythms. Some techniques are based on the detection of a single arrhythmia type and its discrimination from normal sinus rhythm, or the discrimination between two different types of arrhythmia utilizing time-domain analysis (Throne, Jenkins, & DiCarlo, 1991), the sequential hypothesis-testing algorithm (Thakor, Zhu, & Pan, 1990), threshold-crossing intervals (Clayton, Murray, & Campbell, 1993), artificial neural networks (Clayton, Murray, & Campbell, 1994; Yang, Device, & Macarlane, 1994), time-frequency analysis (Afonso & Tompkins, 1995; Tsipouras & Fotiadis, 2004), fuzzy adaptive resonance theory mapping (Ham & Han, 1996), and the sequential detection algorithm (Chen, Clarkson, & Fan, 1996). Another category of methods for arrhythmia detection and classification is based on the detection of different heart rhythms and their classification in two or three arrhythmia types, and the normal sinus rhythm. Techniques belonging to this category include multiway sequential-hypothesis testing (Thakor, Natarajan, & Tomaselli, 1994), wavelet analysis (Khadra, Al-Fahoum, & Al-Nashash, 1997), artificial neural networks (Minami, Nakajima, & Toyoshima, 1999), complexity measure (Zhang, Zhu, Thakor, & Wang, 1999), multifractal analysis (Wang, Zhu, Thakor, & Xu, 2001), wavelet analysis combined with radial basis function neural networks (Al-Fahoum & Howitt, 1999), and nonlinear dynamical modeling (Owis, Abou-Zied, Youssef, & Kadah, 2002). It is noticeable that all methods address the detection of only a few types of arrhythmia (atrial tachycardia, ventricular tachycardia, atrial fibrillation, and ventricular fibrillation). ECG beat-by-beat classification is another field of interest, where each beat is classified into several different rhythm types utilizing artificial neural networks (Dokur & Olmez, 2001), fuzzy neural networks (Osowski & Linh, 2001), the "mixture of experts approach" (Hu, Palreddy, & Tompkins, 1997), hermite functions combined with self-organizing maps (Lagerholm, Peterson, Braccini, Ebendrandt, & Sornmo, 2000), and time-frequency analysis combined with knowledge-based systems (Tsipouras, Fotiadis, & Sideris, 2002). These methods classify more arrhythmic beat types, but they focus on single-beat classification and not arrhythmic episode detection.

Most of the studies are based on the analysis of the ECG signal. In these methods, ECG features are extracted and used for the detection and/or classification of arrhythmias. However, the presence of noise makes feature extraction difficult and in some cases impossible. Also, most of the methods are time consuming and ineffective for real-time analysis. An alternative would be to use only the RR-interval signal. In this case, it is expected that certain types of arrhythmias can be detected and classified.

A recent work for an arrhythmia-detection method based on time and time-frequency analysis (Tsipouras & Fotiadis, 2004) utilizes only the RR-interval signal and heart-rate features. Initially, the RR-interval duration signal is extracted from ECG recordings and segmented into small intervals. The analysis is based on both time and time-frequency features. Time-domain measurements are extracted and several combinations between the obtained features are used to train a set of neural networks. Short-time Fourier transform, and several time-frequency distributions are used in the time-frequency analysis. The final decision is made using a set of rules. The proposed approach is tested using the *MIT-BIH Arrhythmia Database* (Harvard-MIT Division of Health Sciences and Technology, 1997), and satisfactory results are obtained for both sensitivity and specificity for arrhythmic segment detection (87.5% and 89.5% respectively for time-domain analysis, and 90% and 93% respectively for time-frequency-domain analysis).

Another approach for arrhythmic beat classification and arrhythmic episode detection and classification (Tsipouras, Fotiadis, & Sideris, in press) is also based only on the RR-interval signal. A three RR-interval sliding window is used in the arrhythmic beat classification algorithm. Classification is performed for four categories of beats: normal, premature ventricular contractions, ventricular flutter and fibrillation, and 2^{o} heart block. The beat classification is used as the input of a knowledge-based, deterministic automaton to achieve arrhythmic episode detection and classification. Six rhythm types are classified: ventricular bigeminy, ventricular trigeminy, ventricular couplet, ventricular tachycardia, ventricular flutter and fibrillation, and 2^{o} heart block. The achieved scores for the *MIT-BIH Arrhythmia Database* indicate high performance: 98% accuracy for arrhythmic beat classification and 94% accuracy for arrhythmic episode detection and classification.

In Table 2, several systems and methodologies that address the problem of arrhythmic beat classification are presented.

Table 2. Comparison of several research attempts for arrhythmic beat classification

Method	Signal	Dataset	Acc[1] (%)
Feature extraction: cumulants of the second, third and fourth order Classification: fuzzy hybrid neural network (Osowski, & Linh, 2001)	ECG	7,185 beats from MIT-BIH 4,035 training – 3,150 testing N : 2,250 A : 658 L : 1,200 V : 1,500 R : 1,000 I : 472 E : 105	96.06%
Feature extraction: discrete wavelet transform Classification: intersecting Spheres network (Dokur, & Olmez, 2001)	ECG	3,000 beats from MIT-BIH N, L, R, P, p, a, E, V, F, f: 300 from each category 1,500 training – 1,500 testing	95.7%
Feature extraction: PCA in 29 points from QRS, instantaneous and average RR-interval, QRS complex width Classification: mixture of experts (SOM, LVQ) (Hu et. al., 1997).	ECG	25 min from each record in MIT-BIH 200 series excluding records 212, 217, 220, 222 and 232 N : 43897 V : 5363	95.52%
Feature extraction: hermite functions, RR-interval Clustering: self organizing maps (Lagerholm et. al., 2000)	ECG	108,963 beats from MIT-BIH N : 74053 F : 803 L : 8074 b : 472 R : 7259 e : 16 A : 2544 j : 229 a : 150 E : 106 J : 83 P : 7028 S : 2 f : 982 V : 7129 Q : 33	98.49%
Feature extraction: RR-interval Classification: knowledge-based system (Tsipouras et. al., 2002)	RR-interval signal	30,000 beats from MIT-BIH N, P, f, p, Q, L, R : 25,188 V, F : 2,950 e,j,n,E : 265 A, a, J, S : 1,213 [, !,] : 384	95.85%
Feature extraction: RR-interval Classification: knowledge-based system (Tsipouras et. al., 2005)	RR-interval signal	93,349 beats from MIT-BIH N, P, f, p, L, R, Q : 86,262 V : 6,183 [, !,] : 484 BII : 420	98.20%
		109,880 beats from MIT-BIH N, P, f, p, L, R, Q : 102,793 V : 6,183 [, !,] : 484 BII : 420	94.26%

1Acc: Accuracy

CONCLUSION

We presented a general review of ECG diagnosis using decision-support systems. Computerized ECG processing can be divided into several stages. The most important ones are the analysis and the diagnosis. The variety of the systems developed in order to address the automation of these two tasks shows the great interest for this scientific area and its clinical importance. The developed methods differ in terms of accuracy, but each one of them has unique advantages. A proper combination of the above techniques can improve the results of ECG analysis and diagnosis. The physician, however, remains responsible to evaluate the systems' output and make the final diagnosis.

REFERENCES

Afonso, V. X., & Tompkins, W. J. (1995). Detecting ventricular fibrillation. *IEEE Engineering in Medicine and Biology, 14,* 152-159.

Al-Fahoum, A. S., & Howitt, I. (1999). Combined wavelet transformation and radial basis neural networks for classifying life-threatening cardiac arrhythmias. *Medical & Biological Engineering and Computing, 37,* 566-573.

Chen, S. W., Clarkson, P. M., & Fan, Q. (1996). A robust sequential detection algorithm for cardiac arrhythmia classification. *IEEE Transactions on Biomedical Engineering, 43,* 1120-1125.

Clayton, R. H., Murray, A., & Campbell, R. W. F. (1993). Comparison of four techniques for recognition of ventricular fibrillation of the surface ECG. *Medical & Biological Engineering and Computing, 31*, 111-117.

Clayton, R. H., Murray, A., & Campbell, R. W. F. (1994). Recognition of ventricular fibrillation using neural networks. *Medical & Biological Engineering and Computing, 32*, 217-220.

Dokur, Z., & Olmez, T. (2001). ECG beat classification by a hybrid neural network. *Computer Methods and Programs in Biomedicine, 66*, 167-181.

European Society of Cardiology. (1991).*European ST-T database directory*. Pisa: S.T.A.R.

Exarchos, T. P., Papaloukas, C., Fotiadis, D. I., & Michalis, L. K. (in press). An association rule mining based methodology for automatic diagnosis of myocardial ischemia using ECGs. *IEEE Transactions on Biomedical Engineering.*

Goldman, M. J. (1982). *Principles of clinical electrocardiography* (11th ed.). Los Altos, CA: LANGE Medical Publications.

Goletsis, Y., Papaloukas, C., Fotiadis, D. I., Likas, A., & Michalis, L. K. (2003). A multicriteria decision based approach for ischemia detection in long duration ECGs. *Proceedings of IEEE EMBS Fourth International Conference of Information Technology and Applications in Biomedicine (ITAB 2003)* (pp. 230-233).

Goletsis, Y., Papaloukas, C., Fotiadis, D. I., Likas, A., & Michalis, L. K. (2004). Automatic ischemic beat classification using genetic algorithms and multicriteria decision analysis. *IEEE Transactions on Biomedical Engineering, 51*, 1717- 1725.

Ham, F. M., & Han, S. (1996). Classification of cardiac arrhythmias using fuzzy ARTMAP. *IEEE Transactions on Biomedical Engineering, 43*, 425-430.

Hu, Y. Z., Palreddy, S., & Tompkins, W. J. (1997). A patient-adaptable ECG beat classifier using a mixture of experts approach. *IEEE Transactions on Biomedical Engineering, 44*, 891-900.

Jager, F., Mark, R. G., Moody, G. B., & Divjak, S. (1992). Analysis of transient ST segment changes during ambulatory monitoring using the Karhunen-Loève transform. *Proceedings of the IEEE Computers in Cardiology* (pp. 691-694).

Khadra, L., Al-Fahoum, A. S., & Al-Nashash, H. (1997). Detection of life-threatening cardiac arrhythmias using wavelet transformation. *Medical & Biological Engineering and Computing, 35*, 626-632.

Lagerholm, M., Peterson, C., Braccini, G., Ebendrandt, L., & Sornmo, L. (2000). Clustering ECG complexes using hermite functions and self-organizing maps. *IEEE Transactions on Biomedical Engineering, 47,* 838-848.

Lemire, D., Pharand, C., Rajaonah, J.-C., Dube, B., & LeBlanc, A. R. (2000). Wavelet time entropy, T wave morphology and myocardial ischemia. *IEEE Transactions on Biomedical Engineering, 47*, 967-970.

Maglaveras, N., Stamkopoulos, T., Pappas, C., & Strintzis, M. (1998). ECG processing techniques based on neural networks and bidirectional associative memories. *Journal of Medical Engineering in Technology, 22*, 106-111.

Minami, K., Nakajima, H., & Toyoshima, T. (1999). Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network. *IEEE Transactions on Biomedical Engineering, 46*, 179-185.

Harvard-MIT Division of Health Sciences and Technology. (1997). *MIT-BIH arrhythmia database* (3rd ed.) [CD-ROM]. Cambridge, MA.

Osowski, S., & Linh, T. H. (2001). ECG beat recognition using fuzzy hybrid neural network. *IEEE Transactions on Biomedical Engineering, 48*, 1265-1271.

Owis, M. I., Abou-Zied, A. H., Youssef, A. M., & Kadah, Y. M. (2002). Study of features based on nonlinear dynamical modelling in ECG arrhythmia detection and classification. *IEEE Transactions on Biomedical Engineering, 49*, 733-736.

Papadimitriou, S., Mavroudi, S., Vladutu, L., & Bezerianos, A. (2001). Ischemia detection with a self-organizing map supplemented by supervised learning. *IEEE Transactions on Neural Networks, 12*, 503-515.

Papaloukas, C., Fotiadis, D. I., Liavas, A. P., Likas, A., & Michalis, L. K. (2001). A knowledge-based technique for automated detection of ischemic episodes in long duration electrocardiograms. *Medical and Biological Engineering and Computing, 39*, 105-112.

Papaloukas, C., Fotiadis, D. I., Likas, A., & Michalis, L. K. (2002a). An expert system for ischemia detection based on parametric modeling and artificial neural networks. *Proceedings of the European Medical Biological Engineering Conference* (pp. 742-743).

Papaloukas, C., Fotiadis, D. I., Likas, A., & Michalis, L. K. (2002b). An ischemia detection method based on artificial neural networks. *Artificial Intelligence in Medicine, 24*, 167-178.

Papaloukas, C., Fotiadis, D. I., Likas, A., Stroumbis, C. S., & Michalis, L. K. (2002). Use of a novel rule-based expert system in the detection of changes in the ST segment and the T wave in long duration ECGs. *Journal of Electrocardiology, 35*, 27-34.

Pitas, I., Strintzis, M. G., Grippas, S., & Xerostylides, C. (1983). *Machine classification of ischemic electrocardiograms.* Proceedings *of the IEEE Mediterranean Electrotechnical Conference (MELECON)*, Athens, Greece.

Sandoe, E., & Sigurd, B. (1991). *Arrhythmia: A guide to clinical electrocardiology.* Bingen, Germany: Publishing Partners Verlags GmbH.

Senhadji, L., Carrault, G., Bellanger, J. J., & Passariello, G. (1995). Comparing wavelet transforms for recognizing cardiac patterns. *IEEE Engineering in Medicine and Biology Magazine, 14*, 167-173.

Silipo, R., & Marchesi, C. (1998). Artificial neural networks for automatic ECG analysis. *IEEE Transactions on Signal Processing, 46*, 1417-1425.

Silipo, R., Taddei, A., & Marchesi, C. (1994). Continuous monitoring and detection of ST-T changes in ischemic patients. *Proceedings of the IEEE Computers in Cardiology* (pp. 225-228).

Stamkopoulos, T., Diamantaras, K., Maglaveras, N., & Strintzis, M. (1998). ECG analysis using nonlinear PCA neural networks for ischemia detection. *IEEE Transactions on Signal Processing, 46*, 3058-3067.

Taddei, A., Costantino, G., Silipo, R., Edmin, M., & Marchesi, C. (1995). A system for the detection of ischemic episodes in ambulatory ECG. *Proceedings of the IEEE Computers in Cardiology* (pp. 705-708).

Thakor, N. V., Natarajan, A., & Tomaselli, G. (1994). Multiway sequential hypothesis testing for tachyarrhythmia discrimination. *IEEE Transactions on Biomedical Engineering, 41*, 480-487.

Thakor, N. V., Zhu, Y. S., & Pan, K. Y. (1990). Ventricular tachycardia and fibrillation detection by a sequential hypothesis testing algorithm. *IEEE Transactions on Biomedical Engineering, 37*, 837-843.

Throne, R. D., Jenkins, J. M., & DiCarlo, L. A. (1991). A comparison of four new time-domain techniques for discriminating monomorphic ventricular tachycardia from sinus rhythm using ventricular waveform morphology. *IEEE Transactions on Biomedical Engineering, 38*, 561-570.

Tsipouras, M. G., & Fotiadis, D. I. (2004). Automatic arrhythmia detection based on time and time-frequency analysis of heart rate variability. *Computer Methods and Programs in Biomedicine, 74*, 95-108.

Tsipouras, M. G., Fotiadis, D. I., & Sideris, D. (2002). Arrhythmia classification using the RR-interval duration signal. In A. Murray (Ed.), *Computers in cardiology* (pp. 485-488). Piscataway: IEEE.

Tsipouras, M. G., Fotiadis, D. I., & Sideris, D. (in press). An arrhythmia classification system based on the RR-interval signal. *Artificial Intelligence in Medicine*.

Vila, J., Presedo, J., Delgado, M., Barro, S., Ruiz, R., & Palacios, F. (1997). SUTIL: Intelligent ischemia monitoring system. *International Journal of Medical Informatics, 47*, 193-214.

Wang, Y., Zhu, Y. S., Thakor, N. V., & Xu, Y. H. (2001). A short-time multifractal approach for arrhythmia detection based on fuzzy neural network. *IEEE Transactions on Biomedical Engineering, 48*, 989-995.

Yang, T. F. Device, B., & Macarlane, P. W. (1994). Artificial neural networks for the diagnosis of atrial fibrillation. *Medical & Biological Engineering and Computing, 32*, 615-619.

Zhang, X. S., Zhu, Y. S., Thakor, N. V., & Wang, Z. Z. (1999). Detecting ventricular tachycardia and fibrillation by complexity measure. *IEEE Transactions on Biomedical Engineering, 45*, 548-555.

KEY TERMS

Artificial Neural Network: An artificial neural network is a massive parallel, distributed processor made up of simple processing units. It has the ability to learn from knowledge, which is expressed through interunit connection strengths, and can make this knowledge available for use.

Data-Mining Association Rules: The process of discovering valuable information from large amounts of data stored in databases, data warehouses, or other information repositories is called data mining. Association-rule mining is the discovery of association relationships or correlations among a set of items.

Genetic Algorithm: Genetic algorithms are derivative-free, stochastic optimization methods based loosely on the concepts of natural selection and evolutionary processes.

Heart-Rate Variability: The alterations of the heart rate between consecutive heartbeats.

Ischaemic Episode: An ischaemic episode is defined as a time period of no less than 30 seconds containing ischaemic beats.

RR-Interval Signal: The signal produced from the durations between consecutive R waves.

Sensitivity: The sensitivity of a test defines the probability that the test will be positive (pathologic) when the outcome is positive (pathologic).

Specificity: The specificity of a test defines the probability that the test will be negative (normal) when the outcome is negative (normal).

This work was previously published in Handbook of Research on Informatics in Healthcare and Biomedicine, edited by A. A. Lazakidou, pp. 135-146, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.14
Three Dimensional Medical Images

Efstratios Poravas
National and Kapodistrian University of Athens, Greece

Nikolaos Giannakakis
National and Kapodistrian University of Athens, Greece

Dimitra Petroudi
National and Kapodistrian University of Athens, Greece

ABSTRACT

The revolution of technology has lead to a change; from the analogic to the digital function of medical devices. Some of them were produced in the last years to improve the quality of images. Although the procedure of acquiring and using the devices has been very complicated, the analysis of the images is so dependable that a big amount of the annual budget is spent for their acquisition.

INTRODUCTION

The rapid development of science and the continuous manufacture of pioneer medical technological products, together with the modern requirements for high-quality medical services, led to the growth and introduction of modern technologies in the health sector. This development has been really impressive and rapid.

At the beginning of the 20th century, the progress of applied sciences, such as chemistry, physics, microbiology, physiology, pharmacology, and so forth enforced medical research, which resulted in the continuous discoveries in medical technology. In 1895, W. K. Roentgen discovered X-rays, which was a turning point for medical imaging and diagnostics in general. In the 1950s, there was the development of computerized systems, while in the 1960s, there were applications such as the transport of biological signals from equipped space missions and teletransfers of information.

To be more specific, the dynamic entry and the enforced intervention of sciences such as informatics gave an enormous impulse to the medical field and created new data for treatment and diagnosis. Thus, we now have the interaction and participation of many different scientific sectors, which aim at the best medical care. One of these sectors is the imaging and treatment of medical pictures, whose development and use are a crucial point in modern therapeutics.

Medical imaging is related to issues such as the descriptive principles of medical images and their elaboration, together with whatever they include. Therefore, it is easy for someone to understand the importance and the role of imaging in diagnosis, treatment, and recovery in general.

This chapter will discuss the basic concepts, such as the analysis of medical images, their elaboration, various uses and applications, as well as an analysis of the functional requirements of such applications in order for them to be fulfilled.

DIGITAL ELABORATION OF IMAGES

The digital elaboration of images is actually a new sector of informatics and has been applied for only 15 years. Its development is revolutionary in the health field and constitutes a major contribution for the promotion of health.

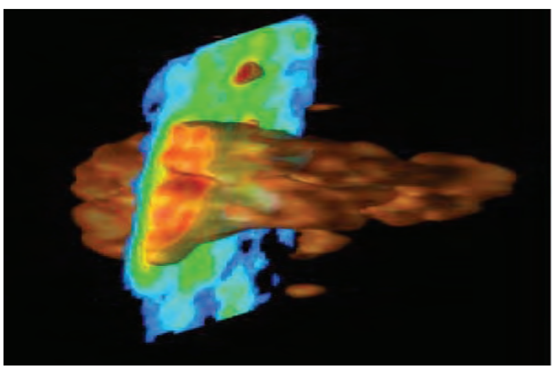
The vast amount of optical information and the need for elaborating them led scientists and technicians to the discovery of storage means for the images and their elaboration by using computers combined with the development and improvement of biomedical equipment.

Digital elaboration, as the title itself declares, deals with the digital registration of images and their elaboration by computers. The objective of elaboration can be the quality improvement of images, the straining of recording or transmission noises, the compression of the amount of information, the storage of images, and their digital transmission and depiction.

In the picture below, we schematically represent all the necessary equipment that is needed to fulfill a digital registration of images to be elaborated by the appropriate personnel in order to give all the necessary information and conclusions.

Thus, everyone can understand that for the elaboration of three-dimensional medical images, a device for the admission of images is required, and this can be an axial tomographer, a magnetic tomographer, an ultrasound device, or even more developed depiction systems. A computer is also required, with all the necessary equipment in order to elaborate the images, analyse them, and store and transmit them, and finally all the appropriate exit devices that will project the images will be needed, which can be terminal stations, special films, and printers. Finally, it is very important that the whole system will support network

Figure 1. Three dimensional imaging



communications so that all information can be available to the scientific community fast and safely with both quality and reliability. Before we introduce some of the depiction devices of three-dimensional medical images that are widely used by the scientific community nowadays, it is important to study some necessary notions for the comprehension of the way this elaboration of three-dimensional images occurs.

An image is a two-dimensional signal. Thus, for the analysis and elaboration of this signal, all the techniques and mathematic relations of digital signal elaboration can be used. Specifically, in the health field, we introduce the notion of modality, which is a biomedical signal that represents one view or one function of the organization. Therefore, the depiction devices receive modalities and with the appropriate processes turn them into images. The dimension of a signal depends on the number of independent variables it has. Consequently, the two-dimensional signals have as independent variables the two dimensions of the surface.

Therefore, the elements that should be included in a digital system for the elaboration of images are the following:

- **Image Elaborator:** This is the hardcore of the system that receives the image, temporarily stores and elaborates it on a low level, and finally demonstrates it on a primary level.
- **Digitizer:** It arithmetically represents the image in order for it to be received by a computer.
- **Digital Computer:** It performs an extensive elaboration.
- **Storage Devices**
- **Screens and Recording Devices**

We will now proceed to analyse the above elements in order to realize their participation in the elaboration of images, and at the same time we will present all the processes that occur at each level.

Image Elaborator

The evolution of technology, as mentioned above, was the vaulting bar for the development and evolution of medical depiction devices and machines, something that proved to be beneficial for health services. Today, there is a large number of medical apparatuses that fills any kind of diagnostic demands. More specifically, in the three-dimensional depiction area, there are now many options

Figure 2. Image elaborator

that suit every demand. The categories of medical images are as follows.

- Digital abstraction angiography
- Ultrasound system
- Computed tomography (CT)
- Magnetic resonance imaging (MRI)
- Gamma camera
- Computed tomographies of nuclear medicine (single-photon-emission computed tomography [SPECT] and positron-emission tomography [PET])

Digital Abstraction Angiography

It is a medical diagnostic examination that depicts the condition of the vessels. It is extremely useful in the sector of cardiology and helps to with the prevention of many diseases of the cardiac system. Its characteristic is that it produces a vast number of data, especially as we move from one-dimensional systems to two-dimensional ones and from a still image to a moving one (video).

Ultrasound System

It depicts the resonance of high-frequency sound waves, which depend on the auditory properties of tissues that are produced by different organs and are examined as brightness in the image. One application of the ultrasound system is the ultrasound of the heart and vessels (Doppler), which is developed in real time and depicts the flux speed of the blood.

Computed Tomography

It uses X-rays for the creation of images, which means that the patient is exposed to radiation during the examination. The part of the body that needs to be examined receives X-rays through different transmission angles. With the use of mathematic transformations and calculations on the counted prices from these different angles, we receive images that are cross and plane sections. Each ray penetrating the body is recording densities of tissues.

Figure 3. Digital angiography

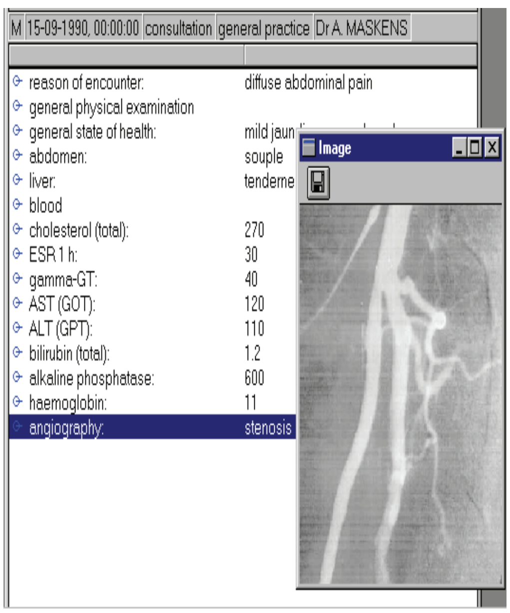


Figure 4. MRI depiction

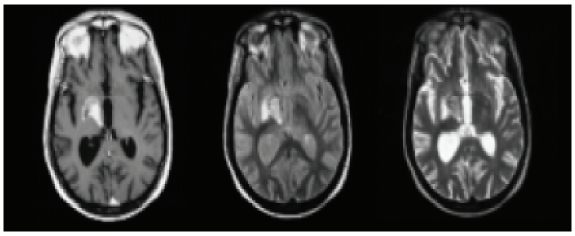
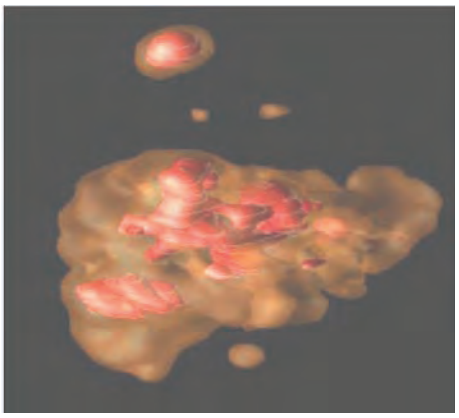


Figure 5. Depiction in SPECT



Magnetic Tomography

It is a nonpenetration diagnostic method that produces a series of images that represent biological differences among tissues. Furthermore, the patient does not have to be exposed to harmful radiation. It is based on the application system of a magnetic field in order to obtain magnetic tomographies.

By applying a magnetic field, the cells orient to the rotation frequency and thus we have the production of resonance through the pulse of radiofrequency. Afterward, we have a rest so that the cells can reorient to their initial positions. The produced signal depends on the density and type of cells and, thus, magnetic tomographies show a high contrast on the soft tissues.

Gamma Camera

It is a diagnostic method that belongs to the field of nuclear medicine. It is based on the calculation of the position and concentration of a radioactive isotope that is provided to the patient before the examination. It gathers clinical information about the normal function while its discernibility is very low.

Computed Tomographies of Nuclear Medicine (PET and SPECT)

This section is about diagnostic examinations that are related to the physiology of the organization and are the most modern in technological terms in the field of diagnostics. SPECT involves the computed tomography of photon transmission, something that shows that the signal-production process is based on photon transmission. PET involves tomography with positron transmission.

The diagnostic devices mentioned are the means of acquiring medical images in analogous or digital forms. The device used may have the ability to store digitally; otherwise, the digitalization occurs with the use of a digitizer before storage. Next we have the process of compressing images.

Digital Image Compression

The term compression is mentioned in a number of technicalities and algorithms aiming to reduce the required memory for the representation and storage of digital images. The storage of a digitalized image leads to a squandering of the memory of the computing system. Thus, in the medical field, where the amount of image information is enormous, a major problem rises because of the amount of memory required for image storage. For this reason, faster compression and decompression algorithms for images have been searched for. Besides sparing memory, compression provides a reduction in the time and the width needed for transmission.

The techniques for the compression of digital images can be divided into two categories.

In medicine, where even the smallest detail in depiction can be of vital importance, we choose lossless techniques.

Compression is based on redundant information that is included in the images. The more redundancy there is, the more is the achieved compression. The whole process is called source coding, and it is based on a system that includes the following:

- **Transformer:** It transforms the initial image into a more appropriate one for compression.
- **Quanter:** It quants the transformed image either with graduation or with vector.
- **Coder:** A sequence of bits corresponds with each quant level. The code can be of a specific or variable length.

Compression with no losses does not use a quanter system because a quant always shows a loss of information.

The used patterns of image compression are as follows:

- G3, G4, GBIG—binary images
- JPEG—firm images
- H.261 MPEG1, MPEG2, MPEG4, MPEG7—movable images

After the compression process, we have the process of storage to the PACS system and recuperation from RIS.

Storage of Medical Images and Their Recuperation

Medical images, according to the diagnostic device used, have large sizes due to the need for high-quality analysis and clearness, thus demanding a large storage space. The recuperation time differs according to the kind of examination and the demanded number of images in each examination.

When the image reaches the final stage of elaboration, the goals are then the improvement of the quality of the given information with the use of the appropriate graphic-elaboration software (e.g., clearness, brightness, contrast), and also the exploitation of specific information with the use of special operations and transformations like division and recombining.

Quality Improvement of Images

The quality improvement of images is achieved with the use of software that helps to reach the desired levels of analysis aiming to exploit and evaluate the information given from an image. During the process of recording images, certain deformities appear. These are:

- Dimness,
- Noise of recording, and
- Geometrical deformities.

All these deformities should be corrected in order to eliminate all kinds of alterations. The elimination of dimness is done with the process of restoration. The process of restoration is extremely important in the elaboration of moving images because movement creates dimness. In most cases, a filtering of the image is also required in order to remove sound. This can be achieved with various linear or nonlinear filters. Usually, nonlinear filters are preferred because they preserve the contrast of outlines, which are an important factor for human vision. Moreover, the general contrast of the image can be improved with special nonlinear techniques.

One other important procedure for image analysis is the recognition of outlines. Usually, the areas of an image are coloured with false colours.

Finally, we achieve improvement of quality by adjusting the proper clearness, brightness, and contrast with the use of appropriate technical equipment that allows us to exploit all these tools provided by the graphic software for quality improvement.

Image Division

By dividing a medical image, we are able to divide the image in different sections that show some cognitive interest according to the speciality under examination. Thus, according to the reason why the diagnostic examination occurs, the scientist can focus on a specific part of the image to process suitably and find the needed information without the redundancy of information to trouble him or her.

Furthermore, with the appropriate software, we can store to the computed system only those parts of the image that interest us, thus sparing storage space and making it more easy to recuperate the images from the used medical databases.

Medical Image Recombination

With medical image recombination, we can have a combination of different views from various depiction examinations (e.g., ultrasound+CT, MRI+PET, MRI+SPECT), giving us a three-dimensional representation of composition.

Indeed, medical image recombination is extremely important in medical depiction. Each diagnostic examination allows the scientist to extract specific information through it. With this recombination method, the scientist can have the elements and information provided by different diagnostic examinations in only one image, can elaborate them and analyse them in any way he

Figure 6. Division of image received by an MRI

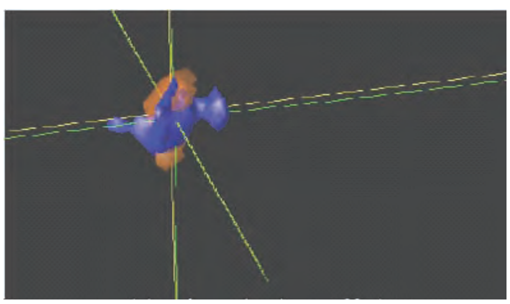


Figure 7. Image recombination MRI + PET + CT

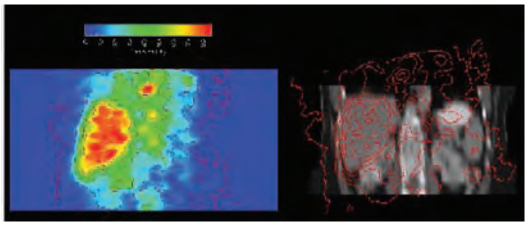
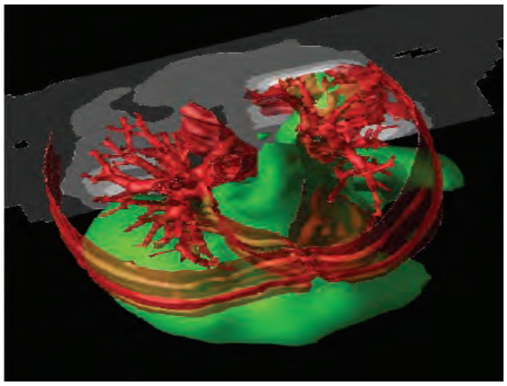


Figure 8. Image recombination SPECT + CT



or she desires, and, finally, can reproduce them in a three-dimensional environment. The whole undertaking occurs through complex graphic software, which should have certain characteristics in order to be an actual diagnostic tool in the hands of the scientist vs. an obstacle in his or her effort to do the job.

The characteristics of the elaboration software include the following:

- Friendly user interface
- Large number of choices that are easy to overcome
- Support of the established models of compression and image management
- Support of future applications
- Safe use
- Is easy to learn

CONCLUSION AND PERSPECTIVES

Three-dimensional image elaboration is a very strong tool in medical diagnostics. Many of the diseases of tissue and microscopic interest are able to be diagnosed early and therefore can be cured. Especially benefited are the fields of research in medicine, pharmacology, and genetics, where the elaboration and depiction of three-dimensional models is extremely crucial for their development. The appearance of more and more evolved depiction devices and the growth of more powerful and more complete software for digital graphic elaboration will predispose even bigger steps of evolvement in the near future.

This work was previously published in Handbook of Research on Informatics in Healthcare and Biomedicine, edited by A. A. Lazakidou, pp. 287-293, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.15
Mobile Medical Image Viewing Using 3G Wireless Network

Carrison K. S. Tong
Pamela Youde Nethersole Eastern Hospital, Hong Kong

Eric T. T. Wong
Hong Kong Polytechnic University, Hong Kong

INTRODUCTION

Teleradiology is a routine practice for radiologists to make urgent diagnosis by remote viewing radiological images such as computed tomographic (CT), magnetic resonance (MR), computed radiographic (CR), and digital radiographic (DR) images outside their hospitals. Traditionally, due to limited network bandwidth and huge image file sizes, this technique was limited to fixed-point communication using an integrated services digital network (ISDN) and broadband network. Without any prior information, most radiologists would invariably require high-quality display units and lossless compressed images for their clinical diagnosis. Besides the technical issues involved in the uninterrupted provision of a 24-hour teleradiology service, most hospital administrators have to consider a series of management issues on the quality of this service such as data confidentiality, integrity, and accessibility.

This article presents the implementation process of a high-quality teleradiology service using the third-generation (3G) wireless network. In the provision of this service, several high-quality notebook computers with a 15-inch liquid crystal display (LCD) screen of resolution 1,024 x 768 pixels and 32-bit color have been configured to view medical images in the digital imaging and communications in medicine (DICOM) format using a Web browser. These notebook computers are connected with 3G mobile phones so that users could access the Internet using Web browsers through the 3G network at a speed of at least 384 kbps. The users could also use the Web browser for logging into the hospital network through an application tunneling technique in a virtual private network (VPN). When logging into the VPN, for security purposes the network authentication is enhanced by a one-time and two-factor authentication (OTTFA) mechanism. In OTTFA, the user password contains two parts: a personal password and a randomly generated password. After successfully logging into the hospital network, the user has to log into the image server using another account name and password. The above are all important to ensure the high standard of confidentiality of the system.

The data volume of the image server is about 1 TB, stored in a level-5 configured redundant array of inexpensive disks (RAID). For management of this huge amount of data, the location of each image in the storage unit is stored in a Structural Query Language (SQL)-based database. Each image also has DICOM tags for storage of the patient name, identity (ID) number, study date, and time. After the success of each login, the user can query the image server for related images using the patient's demographic data such as the study date. These are used to enhance the integrity of the system.

There are three image servers configured in a high available (HA) cluster using a load-balancing switch. The user could access any one of the servers for diagnostic purpose using the teleradiology technique. This setting is used to ensure the availability of the service 24 hours a day/7 days a week. The above system has operated for six months, and zero downtime was recorded. This leads to the belief that it is feasible to operate a quality teleradiology system using 3G networking technology with the important concerns of data confidentiality, integrity, and accessibility being dealt with in an effective manner.

BACKGROUND

Teleradiology is the process of sending radiologic images from one point to another through digital, computer-assisted transmission, typically over standard telephone lines, a wide-area network (WAN), or over a local area network (LAN). Through teleradiology, images can be sent to another part of the hospital or around the world.

In a hospital environment, it is not unusual that sometimes certain senior or experienced clinical staff would not be available onsite. These senior clinical staff may standby at home, on business trip, or just on their way to work. For urgent medical cases, remote consultation is required. It is important to have multimedia communication, including voice, text, and picture, between the senior clinicians and the hospital. A reliable, secure, easy-access, manageable, high-speed, standardized, multimedia medical consultation system is required.

PROBLEM

Today, teleradiology is still facing many limitations such as low network bandwidth, limited locations, and implementation issues associated with security, standards, and data management.

Limited Locations and Low Network Bandwidth

Depending on data-transfer rate requirements and economic considerations, images can be transmitted by means of common telephone lines using twisted pairs of copper wire, digital phone lines such as ISDN, coaxial cable, fiber-optic cable, microwave, satellite, and frame relay or T1 telecommunication links.

Today most teleradiology systems run over standard telephone lines. Over the next couple of years, we should see a substantial migration to switched-56 and ISDN lines, which offer higher speed and better line quality than standard dial-up phone lines. Other high-speed lines, including T1 and SMDS (shared multimegabit data services), will also become more popular as prices continue to drop.

However, remote consultation on fixed lines can only be performed in pre-installed locations such as a radiologist's home. A wide-area wireless network can provide a more flexible teleradiology service for the users (Oguchi, Murase, Kaneko, Takizawa, & Kadoya, 2001; Reponen et al., 2000; Tong, Chan, & Wong, 2003).

Security in a 3G Network

The fragile security of 2.5G and 3G wireless applications was abundantly evident in Japan recently when malicious e-mails to wireless handsets unleashed a malevolent piece of code which took control of the communications device and, in some cases, repeatedly called Japan's national emergency number. Other cell phones merely placed several long-distance calls without the user's knowledge, while others froze up, making it impossible for subscribers to use any of the carrier's services. Incidents like this and others involving spamming, denial of service (DoS), virus attacks, content piracy, and malevolent hacking are becoming rampant. The security breaches that have posed a constant threat to desktop computers over the last decade are migrating to the world of wireless communications where they will pose a similar threat to mobile phones, smart phones, personal digital assistants (PDAs), laptop computers, and other yet-to-be-invented devices that capitalize on the convenience of wireless communications

SOLUTIONS

Standard

In 2003, the American College of Radiology (ACR) published a technical standard of teleradiology in which the DICOM standard (Bidgood & Horii, 1992) was used as a framework for medical-imaging communication. The DICOM standard was developed by the ACR and the National Electrical Manufacturers Association (NEMA) with input from various vendors, academia, and industry groups. Based upon the open system interconnect (OSI) reference model, which defines a seven-layer protocol, DICOM is an application-level standard, which means it exists inside layer 7. DICOM provides standardized formats for images, a common information model, application service definitions, and protocols for communication.

3G Network

3G stands for third generation (Collins & Smith, 2001) and is a wireless industry term for a collection of international standards and technologies aimed at increasing efficiency and improving the performance of mobile wireless networks (data speed, increased capacity for voice and data, and the advent of packet data networks vs. today's switched networks). As second-generation (2G) wireless networks evolve into third-generation systems around the globe, operators are working hard to enable 2G and 3G compatibility and worldwide roaming, including WCDMA, CDMA2000, UMTS, and EDGE technologies. In this project 3G technology was applied in teleradiology service for improving the speed of communication.

Types of 3G

Wideband Code Division Multiple Access (WCDMA)

This is a technology for wideband digital radio communications of Internet, multimedia, video, and other capacity-demanding applications. WCDMA has been selected for the third generation of mobile telephone systems in Europe, Japan, and the United States. Voice, images, data, and video are first converted to a narrowband digital radio signal. The signal is assigned a marker (spreading code) to distinguish it from the signal of other users. WCDMA uses variable rate techniques in digital processing and can achieve multi-rate transmissions. WCDMA has been adopted as a standard by the ITU under the name IMT-2000 direct spread.

Code Division Multiple Access 2000 (CDMA 2000)

Commercially introduced in 1995, CDMA quickly became one of the world's fastest-growing wireless technologies. In 1999, the International Telecommunications Union selected CDMA as the industry standard for new "third-generation" wireless systems. Many leading wireless carriers are now building or upgrading to 3G CDMA networks in order to provide more capacity for voice traffic, along with high-speed data capabilities. Today, over 100 million consumers worldwide rely on CDMA for clear, reliable voice communications and leading-edge data services.

Universal Mobile Telecommunication (UMTS)

This is the name for the third-generation mobile telephone standard in Europe, standardized by the European Telecommunications Standards Institute (ETSI). It uses WCDMA as the underlying standard. To differentiate UMTS from competing network technologies, UMTS is sometimes marketed as 3GSM, emphasizing the combination of the 3G nature of the technology and the GSM standard which it was designed to succeed. At the air interface level, UMTS itself is incompatible with GSM. UMTS phones sold in Europe (as of 2004) are UMTS/GSM dual-mode phones, hence they can also make and receive calls on regular GSM networks. If a UMTS customer travels to an area without UMTS coverage, a UMTS phone will automatically switch to GSM (roaming charges may apply). If the customer travels outside of UMTS coverage during a call, the call will be transparently handed off to available GSM coverage. However, regular GSM phones cannot be used on the UMTS networks.

Enhanced Data for Global Evolution (EDGE)

EDGE is a technology that gives GSM the capacity to handle services for the third generation of mobile telephony. EDGE was developed to enable the transmission of large amounts of data at a high speed, 384 kilobits per second. EDGE uses the same TDMA (time division multiple access) frame structure, logic channel, and 200 kHz carrier bandwidth as today's GSM networks which allows existing cell plans to remain intact.

Image Resolution

Digital images, whether viewed on a computer monitor, transmitted over a phone line, or stored on a hard disk or archival medium, are pictures that have a certain spatial resolution. The spatial resolution, or size, of a digital image is defined as a matrix with a certain number of pixels (information dots) across the width of the image and down the length of the image. The more pixels, the better the resolution. This matrix also has depth. This depth is usually measured in bits and is commonly known as shades of gray: a 6-bit image contains 64 shades of gray; 7-bit, 128 shades; 8-bit, 256 shades; and 12-bit, 4096 shades.

The size of a particular image is referenced by the number of horizontal pixels "by" (or "times") the number of vertical pixels, and then by indicating the number of bits in the shades of gray as the depth. For example, an image might have a resolution of 640×480 and 256 shades of gray, or 8 bits deep. The number of bits in the data set can be calculated by multiplying $640 \times 480 \times 8$ equals 2,457,600 bits. Since there are 8 bits in a byte, the 640×480 image with 256 shades of gray is 307,200 bytes or .3072 megabytes of information.

Data Compression

Although images should be permanently archived as raw data or with only lossless data compression (no data is destroyed), hardware and software technologies exist that allow teleradiology systems to compress digital images into smaller file sizes so that the images can be transmitted faster. Compression is usually expressed as a ratio: 3:1, 10:1, or 15:1. The compression ratio refers to the ratio of the size of a compressed file to the original uncompressed file.

Certain images can withstand a substantial amount of compression without a visual difference: computed tomography and magnetic resonance images have large areas of black background surrounding the actual patient image information in virtually every slice. The loss of some of those pixels has no impact on the perceived quality of the image nor does it significantly change reader-interpretive performance.

Image Transmission

Image-transmission time is directly proportional to the file size of the digital image. The greater the amount of digital information in an image which involves the image matrix size and the number of bits per pixel, the longer the time required to transmit the image from one location to another. A radiological image contains a large amount of digital information. For example, an image with a relatively low resolution of 512 × 512 × 8 bits contains 2,097,152 bits of data, and a 1,024 × 1,024 × 8-bit image has 8,388,608 bits of data. Transmission time has to follow the laws of size. The only way to decrease the transmission time is either to increase the speed of the modem or reduce the number of bits (compress the image) being sent. The following formula is used to calculate the time to transmit an image:

(Matrix Size) × (Matrix Depth + 2 bits) × (Percentage of Compression) / (Modem Speed) = Seconds to Transmit

Matrix Depth is the shades of gray as shown in Table 1.

For modem or router control, most devices add 2 bits when transmitting as overhead.

Data Management

According to the ACR Technical Standard for Teleradiology, each examination data file must have an accurate corresponding patient and examination database record that includes patient name, identification number, examination date, type of examination, and facility at which the examination was performed. A Structural Query Language (SQL) database has been installed for the registration of each incoming and query of studies.

Web Technology

Web technology offers a significant advantage to physicians who need to receive images quickly,

Table 1. Shades of gray and matrix depth

Shades of Gray	Matrix Depth	Shades of Gray	Matrix Depth
256	8 bits	16	4 bits
128	7 bits	8	3 bits
64	6 bits	4	2 bits
32	5 bits	2	1 bit

and who require real-time image navigation and manipulation to perform diagnostic tasks effectively. It facilitates the use of graphical-user interfaces, making teleradiology and picture achieving and communication system (PACS) applications easier to use and more responsive. Additionally, clients and servers can be run on different platforms, allowing end users to free themselves from particular proprietary architectures. Software applications designed for client-server computing can interface seamlessly with most hospital information systems (HISs) (RCR, 1999) or radiology information systems (RISs), while providing rapid soft-copy image distribution.

Storage

RAID (Marcus & Stern, 2003) stands for redundant array of inexpensive (or identical) disks. RAID employs a group of hard disks and a system that sorts and stores data in various forms to improve data-acquisition speed and provide improved data protection. To accomplish this, a system of levels (from 1 to 5) "mirrors," "stripes," and "duplexes" data onto a group of hard disks. All images were stored in the RAID of the server for high availability of the service.

Display

Today, most of the gray shades were produced by a mixing of primary colors in the video boards. Three types of video boards are commonly available, including 16-bit, 24-bit, and 32-bit color video cards. The video board with higher bits can be configured in a lower bit mode. The 16-bit color mode is called "High Color" mode with almost "good enough" quality to show photo images, at least for most purposes; 16-bit color is 5 bits each of red, green, and blue packed into one 16-bit word (2 bytes per pixel). Five bits can show 32 shades of each primary RGB channel, and 32 × 32 × 32 is 32K colors. Green used the extra one bit for 6 bits to achieve 64K colors overall, but half of them are green. The human eye is most sensitive to green-yellow, and more shades are a bigger advantage there. Green has twice the luminance of red and six times more than blue, so this is very reasonable. Video boards do vary, but 24 bits is normally not so much better in most cases, except in wide smooth gradients.

Video boards for the last few years are 24-bit color or "true color"; 24-bit color is 8 bits each of RGB, allowing 256 shades of each primary color, and 256 × 256 × 256 = 16.7 million color combinations. Studies show that the human eye can detect about 100 intensity steps (at any one current brightness adaptation of the iris), so 256 tones of each primary is more than enough. We would not see any difference between RGB (90,200,90) and (90,201,90), but we can detect 1% steps (90,202,90) (on a cathode ray tube (CRT) tube, but 18-bit LCD panels show 1.5% steps). So our video systems and printers simply do not need more than 24 bits.

Theoretically, there is no true 32-bit color display mode. The confusion is that 24-bit color mode normally uses 32-bit video mode today, referring to the efficient 32-bit accelerator chips (word size). The 24-bit color mode and so-called 32-bit video mode show the same 24-bit colors, the same 3 bytes RGB per pixel; 32-bit mode simply discards one of the four bytes (wasting 25% of video memory), because having 3 bytes per pixel severely limits video acceleration functions. Processor chips can only copy data in byte multiples (8, 16, 32, or 64 bits). A 24-bit copy done with a hardware video accelerator would require three 8-bit transfers per pixel instead of one 32-bit transfer; 32-bit video mode is for speed, and it shows 24-bit color.

Liquid Crystal Display (LCD)

In teleradiology, it is more convenient to use LCD for image display than CRT. In LCD, there are no CRTs. Instead, thin "sandwiches" of glass contain liquid-crystal filled cells (red, green, and blue cells) that make up a pixel. Arrays of thin film transistors (TFTs) provide the voltage power, causing the crystals to untwist and realign so that varying amounts of light can shine through each, creating images. This particular sensitivity to light makes LCD technology very useful in projection such as LCD front projectors, where light is focused through LCD chips.

Specifically, there are five layers to the LCD display: a backlight, polarized glass sheet, colored pixel layering, coating of liquid crystal solution that responds to signals off a wired grid of x and y coordinates, followed by a second glass sheet. To create an image, electrical charges, precision coordinated in various degrees and volts, effect the orientation of the liquid crystals, opening and closing them and changing the amount of light that passes through specific colors of pixels. LCD technology has increased its accuracy that can produce sharp and more accurate color images than earlier passive-matrix technologies.

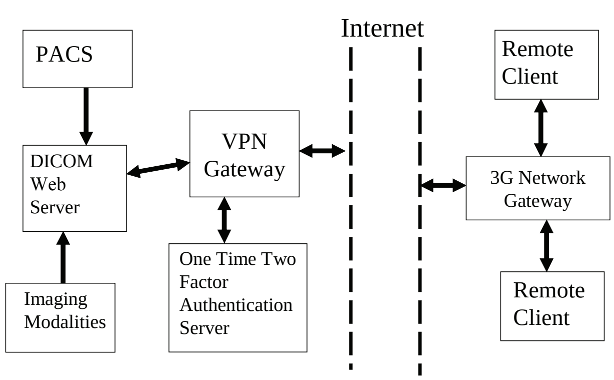
One-Time and Two-Factor Authentication

OTTFA is an authentication protocol that requires two independent ways to establish identity and privileges in which at least one is continuously and non-repetitively changing. This contrasts with traditional password authentication, which requires only one factor such as the knowledge of a password in order to gain access to a system. OTTFA technique provides a secure authentication protocol for the teleradiology service.

Application Tunneling (Port Forwarding)

Application tunneling (or port forwarding) is a combination technique of routing by port combined with packet rewriting. A convention router examines the packet header and dispatches the packet on one of its other interfaces, depending on the packet destination address. Port forwarding examines the packet header and forwards it on to another host with the header rewriting depending on the destination port. The application of application tunneling is its inability of the destination machine to see the actual originator of the forwarded packets and instead seeing them as originating from the router. One of the applications of application tunneling is in a virtual private network gateway, as shown in Figure 1. Using the application tunneling technique, the security of the teleradiology system can be strengthened considerably.

Figure 1. Schematic diagram of 3G wireless medical image viewing system



High Available Server Cluster

The purpose of high available (HA) clustering is to maintain a non-stop teleradiology service for the users. In the current design, there are three image servers configured to form an HA cluster using a load-balancing switch. The user could access any one of the servers for making of diagnosis using the teleradiology technique. Other clustering techniques being used include operating system clustering, Internet protocol (IP) failover, and fault tolerance (FT) techniques.

Implementation Result

The overall design of the 3G wireless medical image viewing system is shown in Figure 1. The medical images received from the PACS or imaging modalities were stored in the DICOM Web server.

The operation of the system is shown in Figures 2 through 10. The users can use a laptop computer as a remote client for the connection to the Internet through a 3G network gateway provided by their service provider (see Figure 2). From the Internet, the users can make a connection to the VPN gateway and be authenticated by the OTTFA server. After login, the VPN gateway will redirect the user to the DICOM Web server using the application tunneling technique. After another login of the DICOM Web server, users can query the server for related studies. Finally, once the related studies are found and selected, all related images can be retrieved and displayed as shown in Figures 9 and 10.

FUTURE TRENDS

4G Wireless Network

4G is the next generation of wireless networks that will replace 3G networks sometimes in future. In another context, 4G is simply an initiative by academic R&D laboratories to move beyond the limitations and problems of 3G to meet its promised performance and throughput. In reality, as of the first half of 2002, 4G is a conceptual framework for or a discussion point to address future needs of a universal high-speed wireless network that will interface with the wireline backbone network seamlessly. 4G also represents the hope and ideas of a group of researchers at Motorola, Qualcomm, Nokia, Ericsson, Sun, HP, NTT DoCoMo, and other infrastructure vendors who must respond to the needs of MMS, multimedia, and video applications if 3G never materializes in its full glory.

Figure 2. A laptop computer for teleradiology

Figure 3. A One-Time Two-Factor Authentication device with a 3G phone for teleradiology

Figure 4. Login screen of VPN gateway

Figure 5. Successful login of VPN gateway

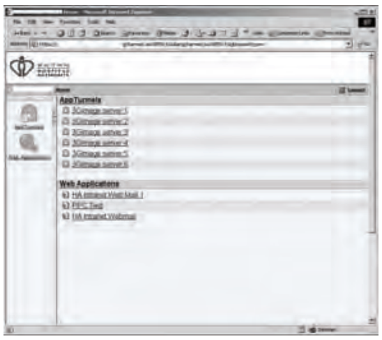


Figure 6. Application tunneling screen

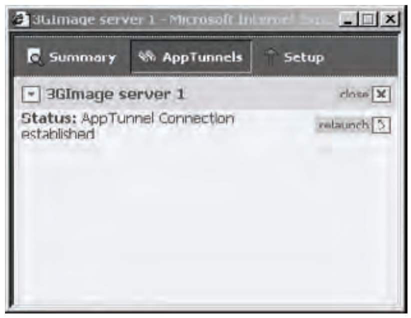
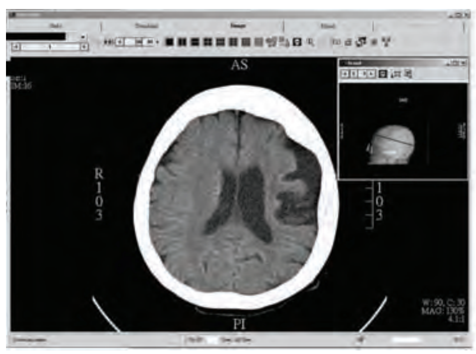


Figure 7. Connection to DICOM Web server established

Figure 8. Image data management

Figure 9. Selection of images

Figure 10. Viewing of images



A comparison of key parameters of 4G with 3G is as shown in Table 2.

CONCLUSION

The above-mentioned 3G wireless medical image viewing system is providing a transfer speed of 384 kbps, which is comparable to a T1 fixed line of a speed of 1.4 mbps, but with greater accessibility and confidentiality. This system has been used successfully to ensure the availability of the teleradiology service 24 hours a day and 7 days a week. Throughout its construction and operation, it is found that the ACR and DICOM standards have provided useful guidelines on achieving the quality assurance and integrity expectations of this kind of service.

REFERENCES

ACR (American College of Radiology). (2003). *ACR technical standard for teleradiology.*

Bidgood, W. D., & Horii, S. C. (1992). Introduction to the ACR NEMA DICOM standard. *Radiographics, 12,* 34-35.

Collins, D., & Smith, C. (2001). *3G wireless networks.* New York: McGraw-Hill.

Marcus, E., & Stern, H. (2003). *Blueprints for high availability* (2nd ed.). New York: John Wiley & Sons.

Oguchi, K., Murase, S., Kaneko, T., Takizawa, M., & Kadoya, M. (2001). Preliminary experience of wireless teleradiology system using Personal Handyphone System. *Nippon Igaku Hoshasen Gakkai Zasshi, 61*(12), 686-687.

RCR (Royal College of Radiologists). (1999). *Guide to information technology in radiology: Teleradiology and PACS*. Board of Faculty of Clinical Radiology, RCR.

Reponen, J., Ilkko, E., Jyrkinen, L., Tervonen, O., Niinimaki, J., Karhula, V., et al. (2000). Initial experience with a wireless personal digital assistant as a teleradiology terminal for reporting emergency computerized tomography scans. *Journal of Telemed Telecare, 6*(1), 45-49.

Tong, C. K. S., Chan, K. K., & Wong, C. K. (2003). Common gateway interfacing and dynamic JPEG techniques for remote handheld medical image viewing. *CARS*, 815-820.

KEY TERMS

Application Tunneling: A combination technique of routing by port combined with packet rewriting.

Digital Imaging and Communication in Medicine (DICOM): A standard developed by the American College of Radiology (ACR) and the National Electrical Manufacturers Association (NEMA) to provide standardized formats for images, a common information model, application service definitions, and protocols for communication.

One-Time Two-Factor Authentication: An authentication protocol that requires two independent ways to establish identity and privileges in which at least one is continuously and non-repetitively changing.

Port Forwarding: Another name for application tunneling.

Redundant Array of Inexpensive (or Identical) Disks (RAID): Employs a group of hard disks and a system that sorts and stores data in various forms to improve data-acquisition speed and provide improved data protection.

Teleradiology: The process of sending radiologic images from one point to another through digital, computer-assisted transmission, typically over standard telephone lines, a wide-area network (WAN), or over a local area network (LAN)

This work was previously published in Encyclopedia of Mobile Computing and Commerce, edited by D. Taniar, pp. 533-540, copyright 2007 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.16
Imaging the Human Brain with Magnetoencephalography

Dimitrios Pantazis
University of Southern California, USA

Richard M. Leahy
University of Southern California, USA

ABSTRACT

Magnetoencephalography is a relatively new medical imaging modality for the monitoring and imaging of human brain function. Extracranial magnetic fields produced by the working human brain are measured by extremely sensitive superconducting sensors, called SQUIDs, enclosed in a liquid helium-filled dewar. Mathematical modeling allows the formation of images or maps of cortical neuronal currents that reveal neural electrical activity, identify cortical communication networks, and facilitate the treatment of neuronal disorders, such as epilepsy.

INTRODUCTION

Magnetoencephalography (MEG) is a noninvasive technique for measuring neuronal activity in the human brain. Electrical currents flowing through neurons generate weak magnetic fields recorded using magnetic sensors surrounding the head. The MEG method is part of a broad area of research referred to as biomagnetism, which involves studies of magnetic fields emanating from several organs of the human body, notably the brain and heart.

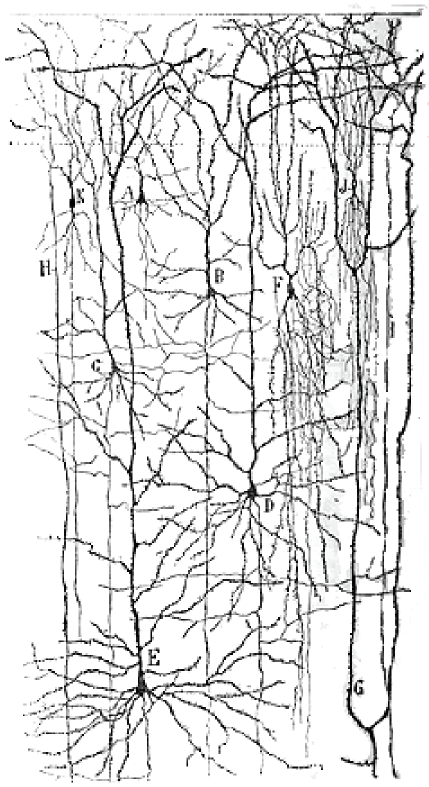
The temporal resolution of MEG is in the millisecond (ms) range, the timescale at which neurons communicate. Therefore, we can follow the rapid cortical activity reflecting ongoing signaling between different brain areas. This is a great advantage compared to other medical imaging modalities such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET), where temporal resolution is on the order of seconds. Furthermore, unlike other methodologies that measure brain metabolism or the relatively slow hemodynamic response, MEG directly measures electrical brain activity. Electroencephalography (EEG) is a complimentary method to MEG, measuring electrical scalp potentials rather than magnetic fields. It offers similar temporal resolution to MEG, but the spatial resolution is less accurate because electrical potentials measured on the scalp are heavily influenced by strongly inhomogeneous conductivity of the head, whereas magnetic fields are mainly produced by currents that flow in the relatively homogeneous intracranial space.

NEURAL BASIS OF ELECTROMAGNETIC SIGNALS

A neuron consists of the cell body (or soma), which contains the nucleus; branching dendrites, which receive signals from other neurons; and a projection called an axon, which conducts the nerve signal. When a pulse arrives at an axon of a presynaptic cell, neurotransmitter molecules are released from the synaptic vesicles into the synaptic cleft. These molecules bind to receptors located on target cells, opening ion channels (mostly Na^+, K^+, and Cl^-) through the membrane. The resulting flow of charge causes an electrical current along the interior of the postsynaptic cell, changing the postsynaptic potential (PSP). When an excitatory PSP reaches the firing threshold at the axon hillock, it initiates an action potential that travels along the axon with undiminished amplitude.

The conservation of electric charge dictates that intracellular currents, commonly called primary currents, give rise to extracellular currents flowing through the volume conductor. Both primary and volume currents contribute to magnetic fields outside the head; however, only locally structured arrangements of cells can achieve sufficient coherent superposition of currents as to produce measurable external fields. Clusters of thousands of synchronously activated pyramidal cortical neurons are believed to be the main generators of MEG signals (Figure 1). In particular, the currents associated with large dendritic trunks, which are locally oriented in parallel and perpendicular to the cortical surface, are believed to be the primary source of the neuromagnetic fields outside the head. In contrast, the temporal summation of currents for action potentials, which have duration of only 1 ms, is not as effective as for dendritic currents flowing in neighboring fibers, so action potentials are believed to contribute little to MEG measurements.

Figure 1. Cerebral frontal cortex drawn by Ramón y Cajal using a Golgi staining technique. Pyramidal (A, B, C, D, E) and nonpyramidal (F, K) cells are clearly depicted. Currents flowing in the dendritic trunks of pyramidal cells are believed to be the primary generators of magnetic signals outside the head.



INSTRUMENTATION

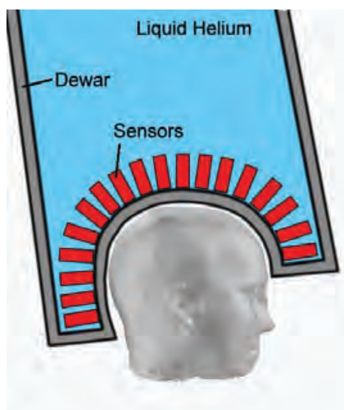
Empirical observations indicate that we observe sources on the order of 10 nA-m, and consequently, the neuromagnetic signals are typically 50 to 500 fT, that is, 10^9 or 10^8 times smaller that the geomagnetic field of the earth (Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993). The only detector that offers sufficient sensitivity to measure such fields is the superconducting quantum interference device (SQUID) introduced in the late 1960s by James Zimmerman (Zimmerman, Thiene, & Harding, 1970). The first measurement of brain magnetic fields using a SQUID magnetometer was carried out by David Cohen (1972) at the Massachusetts Institute of Technology, and it consisted of the spontaneous alpha activity of a healthy participant and abnormal brain activity in an epileptic patient.

The SQUID is a superconducting ring interrupted by thin insulating layers to form one or two Josephson junctions (Barone & Paterno, 1982). One important property associated with Josephson junctions is that magnetic flux is quantized in units of . If a constant biasing current is maintained in the SQUID device, the measured voltage oscillates as the magnetic flux increases; one period of voltage variation corresponds to an increase of one flux quantum. Counting the oscillations allows one to evaluate the flux change that has occurred, and therefore detect magnetic fields on the order of a few fT. The sensitivity of the SQUID can be increased to 1 fT by attaching a coil of superconducting wire or flux transformer. The latter is placed as close to the human head as possible, and depending on its shape, it can be configured as a first-order planar or axial gradiometer, a second-order axial gradiometer, or a simple magnetometer (Hämäläinen et al., 1993). The gradiometer configurations produce measurements proportional to the spatial gradient of the magnetic field, thus offering robustness to interference from distant magnetic field sources.

Modern MEG systems consist of a few hundred SQUID sensors placed in a liquid-helium-filled dewar, with the flux-transformer pickup coils surrounding a helmet structure (Figure 2). Worldwide, three companies build the majority of whole-head MEG systems: 4-D Neuroimaging (formerly Biomagnetic Technologies Bti), Elekta Neuromag Oy, and VSM MedTech Ltd. (manufacturers of the CTF Systems). In recent years, all three vendors have introduced dense arrays comprising over 200 SQUID channels.

Brain magnetic signals are very weak compared to ambient noise. Outside disturbances include fluctuations of the earth's geomagnetic field, power-line fields, electronic devices, eleva-

Figure 2. Whole-head CTF Omega MEG system with 275 axial gradiometers (left), and MEG sensors using low-temperature electronics cooled by liquid helium (right)



tors, and radio-frequency waves. Nearby artifacts are caused by instrumentation noise and body interference, such as heart, skeletal muscle, and spontaneous or incoherent background brain activity. Shielded rooms made of successive layers of mu-metal, copper, and aluminum effectively attenuate high-frequency disturbances. Furthermore, gradiometer flux transformers cancel distant noise sources that produce magnetic fields with small spatial gradients.

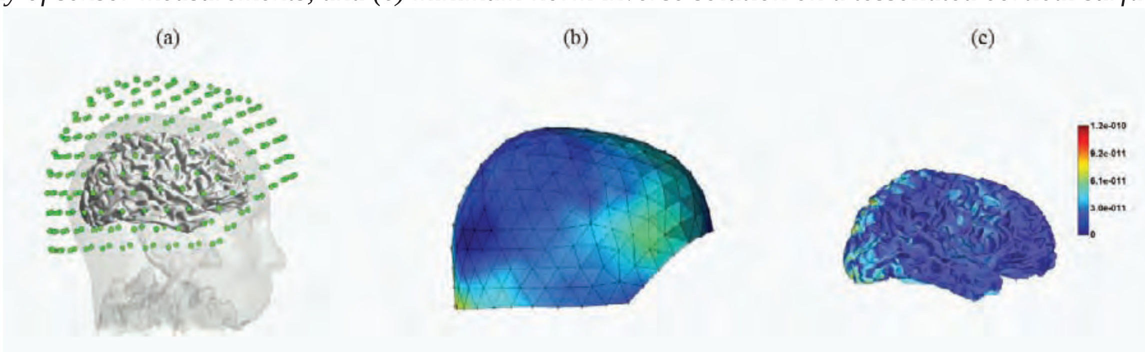
MODELING

To estimate the neural sources of magnetic fields, one must first solve the associated forward problem, that is, the forward model that maps sources of known location, strength, and orientation to the MEG sensors. The most common source model is the current dipole (Baillet, Mosher, & Leahy, 2001), used to approximate the flow of an electrical current in a small area of the brain. The typical strength of a current dipole, generated by the synchronous firing of thousands of neurons, is 10 nA-m. Alternatively, to avoid the identifiability problem that arises when too many small regions and their dipoles are required to represent a single large region of coherent activation, we can use multipolar models, consisting of dipoles, quadrupoles, octupoles, and so on (Mosher, Leahy, Shattuck, & Baillet, 1999).

Since the useful frequency spectrum for electrophysiological signals is largely below 100 Hz, the physics of MEG can be described with the quasistatic approximation of Maxwell's equations. The propagation of electromagnetic fields inside the head is estimated based on the conductivities of the scalp, skull, gray and white matter, cerebrospinal fluid, and other tissue types. Head models that consist of a set of nested concentric spheres with isotropic and homogeneous conductivities have closed-form solutions. Even though spherical head models work surprisingly well, more accurate solutions use realistic head models based on anatomical information from high-resolution magnetic resonance (MR) or x-ray computed tomography (CT) volumetric images. To estimate the parameters of these models, numerical solutions using boundary-element methods (BEMs), finite-element methods (FEMs), or finite-difference methods (FDMs) are necessary (Darvas, Pantazis, Kucukaltun-Yildirim, & Leahy, 2004).

To make inferences about the brain activity that gives rise to a set of MEG data, we must solve the inverse problem, that is, find a neuronal current-source configuration that explains the MEG measurements. Inverse methods for MEG can be roughly categorized into two classes: imaging methods and dipole-fitting or -scanning methods. The imaging approaches are based on the assumption that the primary sources are intracel-

Figure 3. MEG model depicting: (a) Sensor arrangement of a 275-channel CTF MEG system, (b) topography of sensor measurements, and (c) minimum-norm inverse solution on a tessellated cortical surface



lular currents in the dendritic trunks of cortical pyramidal neurons that are aligned normally to the cortical surface. Consequently, a tessellated representation of the cerebral cortex is extracted from a coregistered MR image, and the inverse problem is solved for a current dipole located at each vertex of the surface. In this case, since the position and orientation of the dipoles are fixed, image reconstruction is a linear problem and can be solved using standard techniques. The dipole-fitting or -scanning methods assume that the sources consist of only a few activated regions, each of which can be represented by an equivalent current dipole of unknown location and orientation. The standard approach to localization is to perform a least-squares fit of the dipole model to the data (Lu & Kaufman, 2003). More recently, scanning methods have been developed that are also based on the dipole model, but involve scanning a source volume or surface and detecting sources at those positions at which the scan metric produces a local peak (Baillet et al., 2001). Examples of these methods include the MUSIC (multiple signal classification) algorithm (Mosher, Leahy, & Lewis, 1992) and the LCMV (linearly constrained minimum variance) beamformer (VanVeen, van Drongelen, Yuchtman, & Suzuki, 1997).

Due to intrinsic spatial ambiguities of the electromagnetic principles that underlie MEG, the spatial resolution is lower than that of PET and fMRI. These ambiguities force a choice between low-resolution linear cortical imaging methods, or potentially higher resolution methods based on parametric models, or Bayesian or other nonlinear imaging methods incorporating physiological priors that reflect the expected characteristics of neural activation. A consensus is developing in the research community that no single method suits all MEG applications; each method has strengths and weaknesses, reflecting the ill pose of the inverse problem. The characteristics of expected neural activation, as well as model-fitting techniques, can facilitate the proper choice of inverse methodology.

STATISTICAL ANALYSIS

Given the large number of localization methodologies, it is important to perform validation and statistical analysis under different experimental settings, such as the number, location, and time series of neuronal sources. Furthermore, several methods require the fine-tuning of parameters, such as the subspace correlation threshold for the MUSIC algorithm. The receiver operating characteristic (ROC) curve is a standard tool to evaluate the trade-off between sensitivity and specificity, and to compare different inverse methods. By varying a threshold applied to localization maps, we can estimate two performance measures: the sensitivity or true positive fraction (TPF), and 1-specificity or false positive fraction (FPF). The ROC curve is a plot of the TPF vs. FPF as a detection threshold is varied. When comparing two detection methods, the one whose ROC curve gives higher sensitivity at matched specificity, and vice versa, for all points on the curve is the better detector. A simple metric to compare methods is the area under the ROC curve (AUC), where the method with the largest AUC is superior. The use of free-response ROC, an ROC variant that can handle the presentation and detection of multiple targets per image, is demonstrated in Yildirim, Pantazis, and Leahy (in press) for the evaluation of minimum-norm and scanning-inverse methods.

In addition to evaluating the relative performance of different methods, it is important to establish some degree of confidence in the results of real data analysis. Dipole-scanning methods often produce unstable solutions, and the reproducibility of the reconstructed dipoles is not guaranteed. A number of different approaches have been investigated for assessing dipole-localization accuracy, including Cramer Rao lower bounds, perturbation analysis, and Monte Carlo simulation. To avoid strict distributional assumptions, a resampling alternative based on bootstrap theory was proposed in Darvas, Rautiainen, et al. (2005). The principle underlying the bootstrap theory is that although the distribution of the data is unknown, it can be approximated by the empirical distribution of a set of independent trials. By sampling with replacement over independent trials collected during an event-related MEG study, the position, variance, and time series of current dipoles can be estimated reliably.

In contrast to dipole-scanning methods, imaging methods are hugely underdetermined, resulting in low-resolution localization maps; interpretation is further confounded by the presence of additive noise exhibiting a highly nonuniform spatial correlation. In this case, we need a mechanism to decide which features in the data are indicative of true activation vs. those that are noise artifacts. To determine a suitable threshold for detecting statistically significant activation, the familywise error rate (FWER), that is, the chance of any false positives under the null hypothesis of no activation (Type 1 error), is typically controlled. Parametric random-field methods and nonparametric permutation methods are used to estimate familywise-corrected thresholds in Pantazis, Nichols, Baillet, and Leahy (2005). Alternatively, the control of the false discovery rate (FDR), that is, the proportion of false positives among those tests for which the null hypothesis is rejected, can produce more sensitive thresholds.

Recent literature in MEG statistical analysis has been mostly limited to pairwise comparisons at each cortical surface element for event-related averages. However, extensions of this methodology to the investigation of multiple effects using analysis of variance (ANOVA) and analysis of covariance (ANCOVA) in individuals and groups is possible, as, for example, in Brookes et al. (2004).

APPLICATIONS

Applications in MEG include both basic and clinical research. One of the most important clinical applications is the detection, classification, and localization of abnormal neuronal activity in epilepsy patients. MEG has been successfully used to localize three different spontaneous interrictal signal components: epileptic spikes, slow-wave activity, and fast-wave activity (Lu & Kaufman, 2003). The neurosurgical planning of medically intractable epilepsy often includes the identification of epileptogenic lesions with MEG (Ossadtchi, Baillet, Mosher, Thyerlei, Sutherling, & Leahy, 2004; Stefan et al., 2003). Furthermore, recent literature investigates the possibility of seizure prediction based on a drop in the complexity of neural activity immediately before seizures (Maiwald, Winterhalder, Aschenbrenner-Scheibe, Voss, Schulze-Bonhage, & Timmer, 2004).

In addition to the diagnosis of epilepsy, MEG is currently used for functional brain mapping. Evoked response fields have been used to identify somatosensory-, motor-, and vision-related activity (Lu & Kaufman, 2003). Several MEG studies (Pantazis, Merrifield, Darvas, Sutherling, & Leahy, 2005) have localized language-specific cortical activity using either equivalent current dipoles or distributed cortical imaging, with promising results for clinical application in neurosurgery. Time-frequency analysis of MEG oscillatory-evoked responses (Pantazis, Weber, Dale, Nichols, Simpson, & Leahy, 2005) has detected networks of cortical interactions and determined the functional specificity of several frequency bands. A wide range of signal-processing techniques including image modeling and reconstruction, blind source separation, phase synchrony estimation, nonlinear analysis, and chaos theory are under investigation to reveal complex cognitive processes such as attention and working memory.

Recent literature investigates how evoked response fields relate to neuronal disorders, such as Alzheimer's disease, autism, dyslexia, brain tumors, and Parkinson's disease. Furthermore, MEG has been used in conjunction with trans-axial magnetic stimulation to ameliorate abnormal brain activity (Anninos, Tsagas, Sandyk, & Derpapas, 1991).

CONCLUSION

Magnetoencephalography is a relatively new medical imaging modality for the monitoring and imaging of human brain function. While spatial resolution is significantly lower than that of PET and fMRI, the ability to monitor neuronal activation at the millisecond time scale makes this modality, together with EEG, a unique window on the human brain. Recent developments in instrumentation have lead to the manufacture of whole-head MEG arrays with an excess of 300 magnetometers. Coupled with new data-analysis tools for mapping brain function from MEG data, these systems will lead to important new insights into the workings of the human brain with applications in both clinical and cognitive neuroscience.

REFERENCES

Anninos, P. A., Tsagas, N., Sandyk, R., & Derpapas, K. (1991). Magnetic stimulation in the treatment of partial seizures. *International Journal of Neuroscience, 60*, 141-171.

Baillet, S., Mosher, J. C., & Leahy, R. M. (2001). Electromagnetic brain mapping. *IEEE Signal Processing Magazine, 18*(6), 14-30.

Barone, A., & Paterno, G. (1982). *Physics and applications of the Josephson effect.* Wiley.

Brookes, M. J., Gibson, A. M., Hall, S. D., Furlong, P. L, Barnes, G. R., Hillebrand, A., et al. (2004). A general linear model for MEG beamformer imaging. *Neuroimage, 23*, 936-946.

Cohen, D. (1972). Magnetoencephalography: Detection of the brain's electrical activity with a superconducting magnetometer. *Science, 175*, 664-666.

Darvas, F., Pantazis, D., Kucukaltun-Yildirim, E., & Leahy, R. M. (2004). Mapping human brain function with MEG and EEG: Methods and validation. *Neuroimage, 23*(Suppl. 1), S289-S299.

Darvas, F., Rautiainen, M., Pantazis, D., Baillet, S., Benali, H., Mosher, J. C., et al. (2005). Investigations of dipole localization accuracy in MEG using the bootstrap. *Neuroimage, 25*(2), 355-368.

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography: Theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics, 65*(2), 413.

Lu, Z.-L., & Kaufman, L. (Eds.). (2003). *Magnetic source imaging of the human brain*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Maiwald, T., Winterhalder, M., Aschenbrenner-Scheibe, R., Voss, H. U., Schulze-Bonhage, A., Timmer, J. (2004). Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic. *Physica D, 194*, 357-368.

Mosher, J., Leahy, R., & Lewis, P. (1992). Multiple dipole modeling and localization from spatiotemporal meg data. *IEEE Transactions on Biomedical Engineering, 39,* 541-557.

Mosher, J. C., Leahy, R. M, Shattuck, D. W., & Baillet, S. (1999). MEG source imaging using multipolar expansions. *Proceedings of the 16th Conference on Information Processing in Medical Imaging, IPMI'99* (pp. 15-28).

Ossadtchi, A., Baillet, S., Mosher, J. C., Thyerlei, D., Sutherling, W., & Leahy, R. M. (2004). Automated interrictal spike detection and source localization in magnetoencephalography using independent components analysis and spatio-temporal clustering. *Clinical Neurophysiology, 115*(3), 508-522.

Pantazis, D., Merrifield, W., Darvas, F., Sutherling, W., & Leahy, R. M. (2005). Hemispheric language dominance using MEG cortical imaging and non-parametric statistical analysis. *WSEAS Transactions on Biology and Biomedicine, 2*(3), 318-325.

Pantazis, D., Nichols, T. E., Baillet, S., & Leahy, R. M. (2005). A comparison of random field theory and permutation methods for the statistical analysis of MEG data. *Neuroimage, 25*(2), 383-394.

Pantazis, D., Weber, D. L., Dale, C. C., Nichols, T. E., Simpson, G. V., & Leahy, R. M. (2005). Imaging of oscillatory behavior in event-related MEG studies. In C. A. Bouman & E. L. Miller (Eds.), *Proceedings of SPIE, Computational Imaging III, 5674* (pp. 55-63).

Stefan, H., Hummel, C., Scheler, G., Genow, A., Druschky, K., Tilz, C., et al. (2003). Magnetic brain source imaging of focal epileptic activity: A synopsis of 455 cases. *Brain, 126*(11), 2396-2405.

VanVeen, B., van Drongelen, W., Yuchtman, M., & Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering, 44*(9), 867-880.

Yildirim, E. K., Pantazis, D., & Leahy, R. M. (in press). Task-based comparison of inverse methods in MEG. *IEEE Transactions on Biomedical Engineering.*

Zimmerman, J. E., Thiene, P., & Harding, J. T. (1970). Design and operation of stable rf-based superconducting poinst-contact quantum devices and a note on the properties of perfectly clean metal contacts. *Journal of Applied Physics, 41*, 1572-1580.

KEY TERMS

ANOVA and ANCOVA: Analysis of variance or covariance is a collection of statistical models and their associated procedures that compare means by splitting the overall observed variance into different parts.

Current Dipole: Popular source model in MEG, representing a point's current source. It is a convenient representation for the coherent activation of a large number of pyramidal cells, possibly extending over a few square centimeters of gray matter.

EEG: Electroencephalography measures neuronal activity by recording electrical potentials with electrodes attached on the human scalp. The resulting waveforms are used to localize brain activity and assess brain damage, epilepsy, or even in some cases brain death.

fMRI: Functional magnetic resonance imaging uses powerful magnets to create a field that resonates the nuclei of atoms in the body. The oscillating atoms emit radio signals that are converted by a computer into 3-D images of the human body and cerebral blood flow.

LCMV Beamformer: Linearly constrained minimum-variance beamformer applies spatial filtering to sensor array data to discriminate between signals from a location of interest and those originating elsewhere. In the application to MEG, the goal is to find a spatial filter that minimizes the output power of the beamformer subject to a unity gain constraint at the desired location on the brain.

MUSIC: Multiple signal classification is a localization algorithm that uses the subspace correlation between the data and model subspace to identify the origin of signals. It is often used in MEG to estimate the location, orientation, and strength of current dipoles.

PET: Positron emission tomography is a non-invasive imaging modality that measures the distribution of radioactive-labeled molecules inside a biological system. By using molecular probes that have different rates of uptake depending on the type of tissue involved, PET can localize lesions, and detect regional blood flow and gene expression among others.

This work was previously published in Handbook of Research on Informatics in Healthcare and Biomedicine, edited by A. A. Lazakidou, pp.294-302, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.17
Picture Archiving and Communication System in Healthcare

Carrison KS Tong
Pamela Youde Nethersole Eastern Hospital and Tseung Kwan O Hospital, Hong Kong

Eric TT Wong
The Hong Kong Polytechnic University, Hong Kong

INTRODUCTION

Radiology is the branch of medicine that deals with the diagnostic and therapeutic applications of radiation. It is often used in X-rays in the diagnosis and treatment of a disease. Filmless radiology is a method of digitizing traditional films into electronic files that can be viewed and saved on a computer. This technology generates clearer and easier-to-read images, allowing the patient the chance of a faster evaluation and diagnosis. The time saved may prove to be a crucial element in the patient's treatment process. With filmless radiology, images taken from various medical sources can be manipulated to enhance resolution, increasing the clarity of the image. Images can also be transferred internally within departments and externally to other locations such as the office of the patient's doctor. This is made possible through the picture-archiving and communication system (PACS; Dreyer, Mehta, & Thrall, 2001), which electrsonically captures, transmits, displays, and saves images into digital archives for use at any given time. The PACS functions as a state-of-the-art repository for long-term archiving of digital images, and includes the backup and bandwidth to safeguard uninterrupted network availability. The objective of the picture-archiving and scommunications system is to improve the speed and quality of clinical care by streamlining radiological service and consultation. With instant access to images from virtually anywhere, hospital doctors and clinicians can improve their work processes and speed up the delivery of patient care. Besides making film a thing of the past, the likely benefits would include reduced waiting times for images and reports, and the augmented ability of clinicians since they can get patient information and act upon it much more quickly. The creation of a permanent, nondegradable archive will eliminate the loss of film and so forth. Today, the growing importance of PACS on the fight against highly infectious disease is also identified.

BACKGROUND

PACS (Huang, 2004) started with a teleradiology project sponsored by the U.S. Army in 1983. A follow-up project was the Installation Site for

Digital Imaging Network and Picture Archiving and Communication System (DIN/PACS) funded by the U.S. Army and administered by the MITRE Corporation in 1985. Two university sites were selected for the implementation—the University of Washington in Seattle and Georgetown University and George Washington University Consortium in Washington, DC—with the participation of Philips Medical Systems and AT&T. The U.S. National Cancer Institute funded one of UCLA's first PACS-related research projects in 1985 under the title Multiple Viewing Stations for Diagnostic Radiology.

The early installations of PACS in public health-care institutions were in Baltimore Veterans Administration Medical Center (United States), Hammersmith Hospital (United Kingdom), and Samsung Medical Center (Korea). In Hong Kong, there was no PACS-related project until the establishment of Tseung Kwan O Hospital (TKOH) in 1998. The TKOH was a newly built 600-bed acute hospital with a hospital PACS installed for the provision of filmless radiological service. The design and management of the PACS for patient care will be discussed in this article. The TKOH was opened in 1999 with PACS installed. At the beginning, due to immature PACS technologies, the radiology service was operating with film printing. A major upgrade was done in 2003 for the implementation of server clustering, network resilience, liquid crystal display (LCD), smart card, and storage-area-network (SAN) technologies. This upgrade has greatly improved the reliability of the system. Since November 2003, TKOH has started filmless radiology service for the whole hospital. It has become one of the first filmless hospitals in the Greater China area (Seto, Tsang, Yung, Ching, Ng, & Ho, 2003; Tsou, Goh, Kaw, & Chee, 2003).

MAIN FOCUS OF THE ARTICLE

It certainly goes without saying that most equipment is designed for reliability, but breakdowns can still occur, especially when equipment is used in a demanding environment. A typical situation is what could be called a "single-point failure." That is, the entire system fails if only one piece of equipment such as a network switch fails. If some of the processes that the system supports are critical or the cost of a system stop is too high, then building redundancy into the system is the way to overcome this problem. There are many different approaches, each of which uses a different kind of device, for providing a system with redundancy.

The continuous operation of a PACS in a filmless hospital for patient care is a critical task. The design of a PACS for such a system should be high speed, reliable, and user friendly (Siegel & Kolodner, 2001). The main frame of the design is avoiding the occurrence of any single point of failure in the system. This design includes many technical features. The technical features of the PACS installed in a local hospital include the archiving of various types of images, clustering of Web servers installed, redundancy provision for image distribution channels, and adoption of barcode and smart-card systems. All these features are required to be integrated for effective system performance and they are described below.

ARCHIVING OF MULTIPLE IMAGE TYPES

In order to make connections with different imaging modalities, a common international standard is important. The Digital Imaging and Communications in Medicine (DICOM) standard developed by the American College of Radiology (ACR) and the National Electrical Manufacturers' Association (NEMA) is the most common standard used today. The DICOM standard is extremely comprehensive and adaptable. It covers the specification image format, a point-to-point connection, network requirements, and the handling of information on networks. The adoption of DICOM by other specialties that generate images (e.g., pathology, endoscopy, dentistry) is also planned.

The fact that many of the medical imaging-equipment manufacturers are global corporations has sparked considerable international interest in DICOM. The European standards organization, the Comitâ Europâen de Normalisation, uses DICOM as the basis for the fully compatible MEDICOM standard. In Japan, the Japanese Industry Association of Radiation Apparatus and the Medical Information Systems Development Center have adopted the portions of DICOM that pertain to the exchange of images on removable media and are considering DICOM for future versions of the Medical Image Processing Standard. The DICOM standard is now being maintained and extended by an international, multispecialty committee. Today, the DICOM standard has become a predominant standard for the communication of medical imaging devices.

WEB TECHNOLOGY

The World Wide Web (WWW) began in March 1989 at CERN (CERN was originally named after its founding body, the Conseil Europeen pour la Recherche Nucleaire, that is now called the European Laboratory for Particle Physics.). CERN is a meeting place for physicists from all over the world who collaborate on complex physics, engineering, and information-handling projects. Thus, the need for the WWW system arose from the geographical dispersion of large collaborations and the fast turnover of fellows, students, and visiting scientists who had to get up to speed on projects and leave a lasting contribution before leaving.

Set off in 1989, the WWW quickly gained great popularity among Internet users. For instance, at 11:22 a.m. of April 12, 1995, the WWW server at the SEAS (School of Engineering & Applied Science) of the University of Pennsylvania responded to 128 requests in 1 minute. Between 10:00 and 11:00, it responded to 5,086 requests in 1 hour, or about 84 requests per minute. Even years after its creation, the Web is constantly maturing: In December 1994 the WWW was growing at roughly 1% a day—a doubling in a period of less than 10 weeks (Berners-Lee, 2000).

The system requirements for running a WWW server (Menasce & Almeida, 2001, 2004) are minimal, so even administrators with limited funds had a chance to become information providers. Because of the intuitive nature of hypertext, many inexperienced computer users were able to connect to the network. Furthermore, the simplicity of the hypertext markup language, used for creating interactive documents, has allowed many users to contribute to the expanding database of documents on the Web. Also, the nature of the World Wide Web provided a way to interconnect computers running different operating systems, and display information created in a variety of existing media formats. In short, the Web technology provides a reliable platform for the distribution of various kinds of information including medical images.

Another advantage of Web technology is its low demand on the Web client. Any computer running on a common platform such as Windows or Mac can access the Web server for image viewing just using Internet Explorer or Netscape. Any clinical user can carry out his or her duty anytime and anywhere within a hospital.

CLUSTERING OF DICOM WEB SERVERS

The advantage of clustering computers for high availability (Piedad & Hawkings, 2000) is that if one of the computers fails, another computer in the cluster can then assume the workload of the failed computer at a prespecified time interval. Users of the system see no interruption of access. The advantages of clustering DICOM Web servers for scalability include increased application performance and the support of a greater number of users for image distribution.

There is a myth that to provide high availability (Marcus & Stern, 2003), all that is required is to cluster one or more computer-hardware solutions. To date, no hardware-only solution has been able to deliver trouble-free answers. Providing trouble-free solutions requires extensive and complex software to be written to cope with the myriad of failure modes that are possible with two or more sets of hardware.

Clustering can be implemented at different levels of the system, including hardware, operating systems, middleware, systems management, and applications. The more layers that incorporate clustering technology, the more complex the whole system is to manage. To implement a successful clustering solution, specialists in all the technologies (i.e., hardware, networking, and software) are required. The authors used the clustering of Web servers by connecting all of the Web servers using a load-balancing switch. This method has the advantage of a low server overhead and requires no computer-processor power.

RAID TECHNOLOGY

Patterson, Gibson, and Katz (1988) at the University of California, Berkeley, published a paper entitled "A Case for Redundant Arrays of Inexpensive Disks (RAID)." This paper described various types of disk arrays, referred to by the acronym RAID. The basic idea of RAID was to combine multiple small, inexpensive disk drives into an array of disk drives, which yields performance exceeding that of a single large, expensive drive (SLED). Additionally, this array of drives appears to the computer as a single logical storage unit or drive.

The mean time between failure (MTBF) of the array will be equal to the MTBF of an individual drive divided by the number of drives in the array. Because of this, the MTBF of an array of drives would be too low for many application requirements. However, disk arrays can be made fault tolerant by redundantly storing information in various ways.

Five types of array architectures, RAID-1 through RAID-5, were defined by the Berkeley paper, each providing disk fault tolerance and each offering different trade-offs in features and performance. In addition to these five redundant array architectures, it has become popular to refer to a nonredundant array of disk drives as a RAID-0 array.

In PACS, RAID technology can provide protection for the availability of the data in the server. In RAID level 5, no data is lost even during the failure of a single hard disk within a RAID group. This is essential for a patient-care information system. Extra protection can be obtained by using spare global hard disks for automatic protection of data during the malfunctioning of more than one hard disk. Today, most SANs for high capacity storage are built on RAID technology.

STORAGE AREA NETWORK

A storage area network (Marcus & Stern, 2003; Toigo & Toigo, 2003) is a high-speed, special-purpose network (or subnetwork) that interconnects different kinds of data-storage devices with associated data servers on behalf of a larger network of users. Typically, a storage-area network is part of the overall network of computing resources for an enterprise. A storage-area network is usually clustered in close proximity to other computing resources such as SUN (SUN Microsystems) servers, but it may also extend to remote locations for backup and archival storage using wide-area-network carrier technologies such as ATM (Asynchronous Transfer Mode) or Ethernet.

Storage-area networks use fiber channels (FCs) for connecting computers to shared storage devices and for interconnecting storage controllers and drives. Fiber channel is a technology for transmitting data between computer devices at data rates of up to 1 or 2 Gbps and 10 Gbps in the near future. Since fiber channel is 3 times as fast, it has begun to replace the small computer system interface (SCSI) as the transmission interface between servers and clustered storage devices. Another advantage of fiber channel is its high flexibility; devices can be as far as 10 km apart if optical fiber is used as the physical medium. Standards for fiber channel are specified by the Fiber Channel Physical and Signaling standard, and the ANSI (The American National Standards Institute) X3.230-1994, which is also ISO (International Organization for Standardization) 14165-1.

Other advanced features of a SAN are its support of disk mirroring, backup, and restoring; archival and retrieval of archived data; data migration from one storage device to another; and the sharing of data among different servers in a network. SANs can also incorporate subnetworks with network-attached storage (NAS) systems.

REDUNDANT NETWORK FOR IMAGE DISTRIBUTION

Nevertheless, all of the PACS devices still need to be connected to the network, so to maximize system reliability, a PACS network should be built with redundancy (Jones, 2000). To build up a redundant network (Marcus & Stern, 2003), two parallel gigabit-optical fibers were connected between the PACS and the hospital networks as two network segments using four Ethernet switches. The Ethernet switches were configured in such a way that one of the network segments was in active mode while the other was in standby mode. If the active network segment fails, the standby network segment will become active within less than 300 ms to allow the system to keep running continuously.

BAR-CODE SYSTEM

Recognizing that manual data collection and keyed data entry are inefficient and error prone, bar codes evolved to replace human intervention. Bar codes are simply a method of retaining data in a format or medium that is conducive to electronic data entry. In other words, it is much easier to teach a computer to recognize simple patterns of lines, spaces, and squares than it is to teach it to understand written characters or the English language. Bar codes not only improve the accuracy of entered data, but also increase the rate at which data can be entered.

A bar-code system includes printing and reading the bar-code labels. In most hospital information systems, the bar-code system has commonly been adopted as a part of the information system for accurate and fast patient-data retrieval. In PACS, bar-code labels are mostly used for patient identification and DICOM accession. They are used to retrieve records on patient examinations and studies.

SMART-CARD SYSTEM

A smart card is a card that is embedded with either a microprocessor and a memory chip or only a memory chip with nonprogrammable logic. The microprocessor card can add, delete, and otherwise manipulate information on the card, while a memory-chip card, such as prepaid phone cards, can only undertake a predefined operation. Smart cards, unlike magnetic-stripe cards, can carry all necessary functions and information on the card. Smart cards can also be classified as contact and contactless types. The contactless smart card communicates with the reader using the radio frequency (RF) method.

In PACS, a contactless smart-card system was installed for the authentication of the user. The information about the user name, log-in time, and location are stored in a remote server through a computer network.

NO-FILM POLICY

No film was printed when the patients were still under hospital care. Film was printed only when the patient was transferred to another hospital. Under the no-film policy, the chance of spreading infectious diseases through film is reduced.

EMBEDDED LCD MONITOR

To display medical images in the hospital, LCD monitors were installed on the walls in ward areas adjacent to existing light boxes. LCD displays utilize two sheets of polarizing material with a liquid crystal solution between them. An electric current passed through the liquid causes the crystals to align so that light cannot pass through them. Each crystal, therefore, is like a shutter, either allowing light to pass through or blocking the light. Monochrome LCD images usually appear as blue or dark-grey images on top of a greyish-white background. Colour LCD displays use two basic techniques for producing colour: Passive matrix is the less expensive of the two technologies. The other technology, called thin film transistor (TFT) or active matrix, produces colour images that are as sharp as traditional CRT (Cathode Ray Tube) displays, but the technology is expensive. Recent passive-matrix displays using new colour super-twist nematic (CSTN) and double-layer super-twisted nematic (DSTN) technologies produce sharp colours rivaling active-matrix displays.

Most LCD screens used are transmissive to make them easier to read. These are a type of LCD screens in which the pixels are illuminated from behind the monitor screen. Transmissive LCDs are commonly used because they offer high contrast and deep colours, and are well suited for indoor environments and low-light circumstances. However, transmissive LCDs are at a disadvantage in very bright light, such as outdoors in full sunlight, as the screen can be hard to read. In PACS, the LCD monitors were installed in pairs for the comparison of a large number of medical images. They were also configured in portrait mode for the display of chest X-ray CR (computed radiography) images.

IMPLEMENTATION

In the design of the TKOH PACS (Figure 1), all computed tomographic (CT), magnetic resonance (MR), ultrasound (US) and computed radiographic images were archived in image servers of the PACS (Figure 2). During the diagnosis and monitoring of patients with highly infectious diseases, CT and CR scans were commonly used for comparison. A large storage capacity for the present and previous studies was required. The capacity of the image servers designed was about 5 terabytes using 2.3-terabyte SAN technology and a DICOM compression of 2.5. The image distribution to the clinicians was through a cluster of Web servers, which provided high availability of the service. The connection between the PACS and the hospital network was through a cluster of automatic fail-over switches as shown in Figure 3. Our users can use a Web browser for X-ray-image viewing for the diagnosis or follow-up of patients. The Web-based X-ray-image viewers were set up on the computers in all wards, intensive care

Figure 1. X-ray imaging modalities in the TKOH PACS

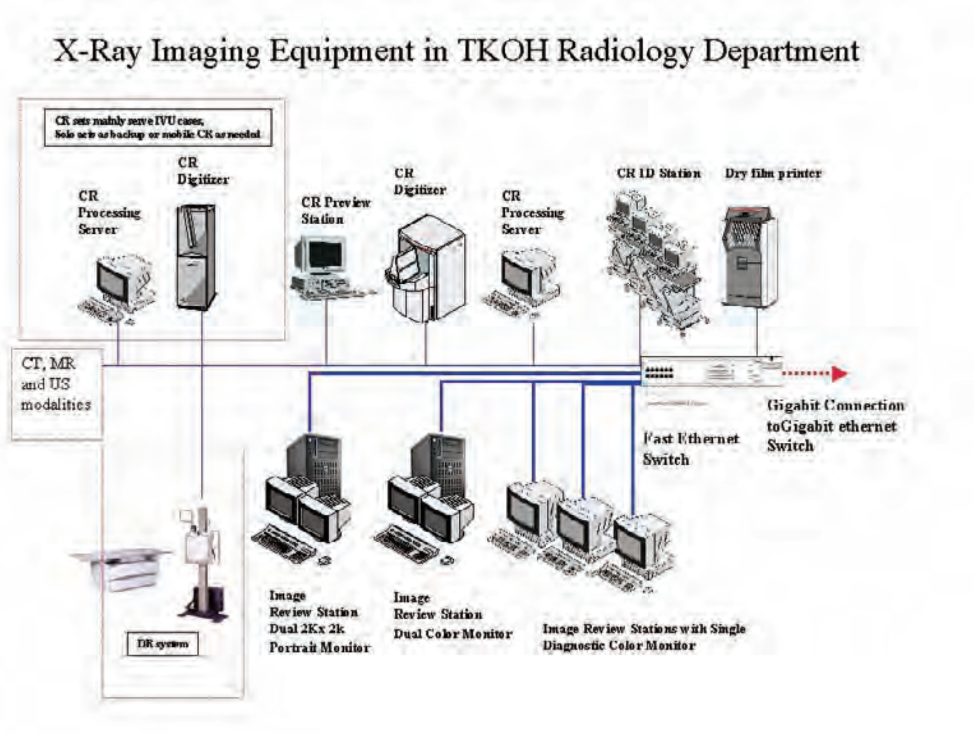


Figure 2. Design of the TKOH PACS

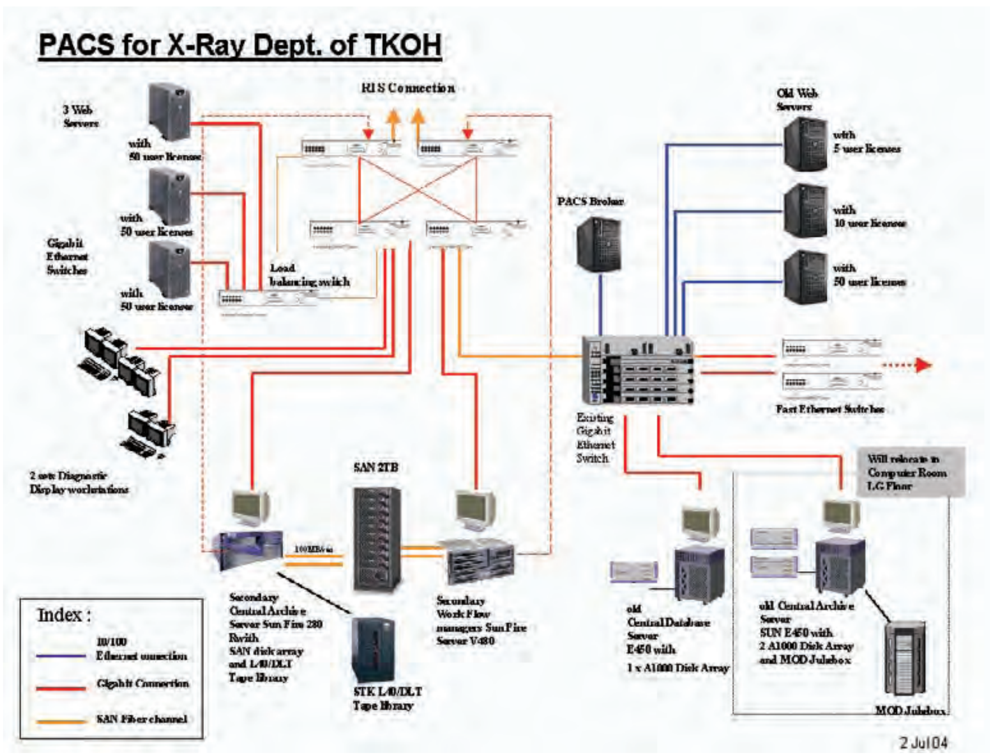


Figure 3. Design of a PACS and hospital network interface

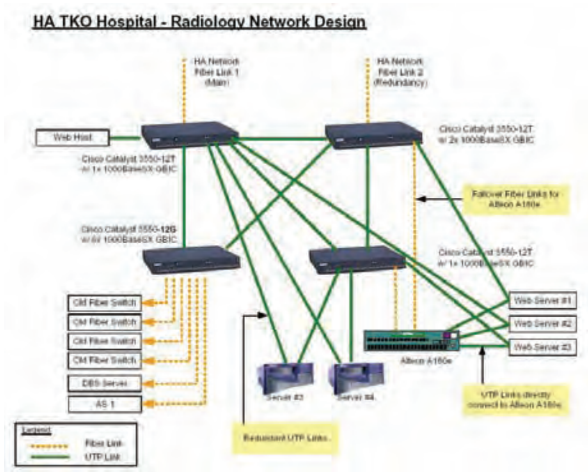
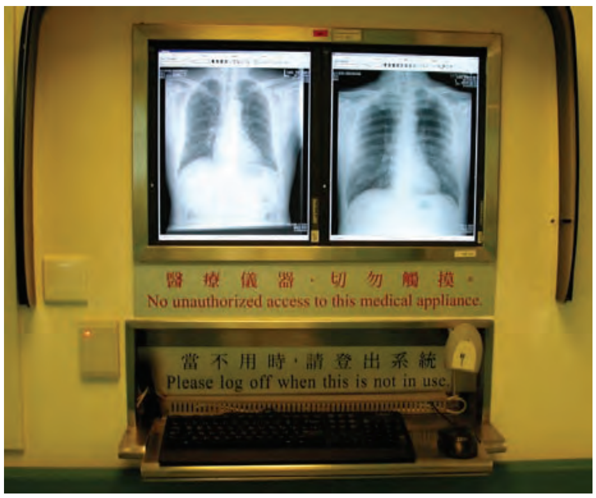


Figure 4. X-ray image viewer in wards



units, and specialist and outpatient departments to provide a filmless radiological service. The design of the computers for X-ray-image viewing in wards is shown in Figure 4. These computers were built using all the above technologies for performance and reliability.

After 10 months of filmless radiological operation in TKOH, less than 1% of the cases required special X-ray film for follow-up. Basically, X-ray-image viewing through a computer network was sufficient for the radiological diagnosis and monitoring of patients. Furthermore, filmless radiology (Siegel & Kolodner, 2001) service definitely reduced the chance for spreading highly infectious diseases through health-care staff. No staff member from the radiology department became infected during the outbreak of the severe acute respiratory syndrome (SARS) in 2003. No film-loss and film-waiting times were recorded.

FUTURE TRENDS

In PACS, most of the hard disks used in the RAID are expensive fiber-channel drives. Some RAID manufacturers are designing their RAID controllers using mixed ATA (Advanced Technology Attachment) and fiber-channel drives in the same array with 100% software compatibility. This design has many benefits. It can reduce the data backup and restore from seconds to hours, keep more information online, reduce the cost of the RAID, and replace the unreliable tap devices in the future. Another advanced development of PACS was in the application of voice recognition (Dreyer et al., 2001) in radiology reporting, in which the computer system was able to automatically and instantly convert the radiologist's verbal input into a textual diagnostic report. Hence, the efficiency of diagnostic radiologists can be further improved.

CONCLUSION

It has been reported (Siegel & Kolodner, 2001) that filmless radiological service using PACS could be an effective means to improve the efficiency and quality of patient care. Other advantages of filmless radiological service are infection protection for health-care staff and the reduction of the spreading of disease through the distribution of films. In order to achieve the above tasks, many computer and multimedia technologies such as the Web, SAN, RAID, high availability, LCD, bar code, smart card, and voice recognition were applied. In conclusion, the applications of computer and multimedia technologies in medicine for efficient and quality health care is one of the important areas of future IT development. There is no boundary and limitation in this application. We shall see doctors learning and using computers in their offices and IT professionals developing new medical applications for health care. The only limitation we have is our imagination.

REFERENCES

Berners-Lee, T. (2000). *Weaving the Web: The original design and ultimate destiny of the World Wide Web.* San Francisco, CA: HarperBusiness.

Dreyer, K. J., Mehta, A., & Thrall, J. H. (2001). *PACS: A guide to the digital revolution* (1st ed.). New York: Springer-Verlag.

Huang, H. K. (2004). *PACS and imaging informatics: Basic principles and applications* (2nd ed.). Hoboken, NJ: Wiley-Liss.

Jones, V. C. (2000). *High availability networking with Cisco* (1st ed.). Boston: Addison Wesley Longman.

Marcus, E., & Stern, H. (2003). *Blueprints for high availability* (2nd ed.). Indianapolis: Wiley.

Menasce, D. A., & Almeida, V. A. F. (2001). *Capacity planning for Web services: Metrics, models, and methods* (2nd ed., chap. 5). Upper Saddle River, NJ: Prentice Hall PTR.

Menasce, D. A., & Almeida, V. A. F. (2004). *Performance by design: Computer capacity planning by example* (chap. 6). Upper Saddle River, NJ: Pearson Education.

Patterson, D., Gibson, G., & Katz, R. H. (1988). A case for redundant arrays of inexpensive disks (RAID). *ACM SIGMOD Record*, *17*(3), 109-116.

Piedad, F., & Hawkings, M. (2000). *High availability: Design, techniques and processes* (chap. 8). Upper Saddle River, NJ: Prentice Hall PTR.

Seto, W. H., Tsang, D., Yung, R. W., Ching, T. Y., Ng, T. K., & Ho, M. (2003). Effectiveness of precautions against droplets and contact in prevention of nosocomial transmission of severe acute respiratory syndrome (SARS). *Lancet, 361*, 1519-1520.

Siegel, E. L, & Kolodner, R. M. (2001). *Filmless radiology* (chap. 5). New York: Springer-Verlag.

Toigo, J. W., & Toigo, M. R. (2003). *The holy grail of network storage management* (1st ed., chap. 3). Upper Saddle River, NJ: Prentice Hall PTR.

Tsou, I. Y. Y., Goh, J. S. K., Kaw, G. J. L., & Chee, T. S. G. (2003). Severe acute respiratory syndrome: Management and reconfiguration of a radiology department in an infectious disease situation. *FRCR* Radiology*, 229*, 21-26.

KEY TERMS

Clustering: A cluster is two or more interconnected computers that create a solution to provide higher availability, higher scalability, or both.

Computed Radiography (CR): Computed radiography is a method of capturing and converting radiographic images into a digital form. The medium for capturing the X-ray radiation passing through the patient and generated by a standard X-ray system is a phosphor plate that is placed in a standard-size cassette, replacing the regular radiographic film. The X-ray exposure forms a latent image on a phosphor plate that is then scanned (read or developed) using a laser-beam CR reader. The CR unit displays the resultant digital image on a computer-monitor screen. By the end of the short process, the phosphor plate is erased and ready for another X-ray image exposure.

Computed Tomography (CT): Computed tomography is a specialized radiology procedure that helps doctors see inside the body. CT uses X-rays and computers to create an image. The images show up as a cross-sectional image.

Digital Imaging and Communications in Medicine (DICOM): Digital Imaging and Communications in Medicine is a medical image standard developed by the American College of Radiology and the National Electrical Manufacturers' Association.

Picture-Archiving and Communication System (PACS): A picture-archiving and communication system is a system used for managing, storing, and retrieving medical image data.

Redundant Arrays of Inexpensive Disks (RAID): RAID is a method of accessing multiple individual disks as if the array were one larger disk, spreading data access out over these multiple disks, thereby reducing the risk of losing all data if one drive fails and improving access time.

Severe Acute Respiratory Syndrome (SARS): Severe acute respiratory syndrome is a newly emerged infectious disease with moderately high transmissibility that is caused by a coronavirus.

Storage-Area Network (SAN): A storage-area network is a networked storage infrastructure (also known as a fabric) that provides the any-to-any connectivity between servers and storage devices, such as RAID disk systems and tape libraries.

This work was previously published in Encyclopedia of Multimedia Technology and Networking, edited by M. Pagani, pp. 821-828, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.18
Imaging Technologies and Their Applications in Biomedicine and Bioengineering

Nikolaos Giannakakis
National and Kapodistrian University of Athens, Greece

Efstratios Poravas
National and Kapodistrian University of Athens, Greece

ABSTRACT

New developments are making the technology faster, more powerful, less invasive, and less expensive. While the technology evolves, new devices are developed, in purpose to be used in the hospitals. Many new imaging methods are used in biomedical applications today and can predict the growth of a tumor or detect a disease. The advantages are numerous, but the problems, during the acquisition and use by the staff, are also remarkable.

INTRODUCTION

We have come from the family doctor's signature black bag in the first half of the 20th century to the powerful scanning equipment of the modern medical center, from tens of thousands dying in influenza epidemics to hundreds of thousands of seniors receiving their annual flu shots, and from an average life expectancy of about 50 years to our present expectancy of 75 years. The biomedical community is taking advantage of the power of computing and technology so as to manage and analyse data. Imaging technologies save day to day more and more people.

X-rays, endoscopes, CT (computed tomography) scans, MRI (magnetic resonance imaging), digital mammography—these imaging technologies make it possible for medical scientists to peer into the body without cutting through the skin. With video monitors and robotic equipment, surgery becomes less invasive and less traumatic to the body (Sawchuck, 2000). Noninvasive means of looking into the human body are now being used to diagnose a wide variety of diseases, including cancer, Alzheimer's disease, stroke, heart failure, and vascular disease (President's

Committee of Advisors on Science and Technology, 2000). The first imaging technologies, the X-ray (discovered by W. K. Roentgen) and EEG (electroencephalogram), were primitive by today's standards, but both have been considerably improved and provided the conceptual base of the other amazing imaging technologies that have recently emerged.

The most common, CAT (computer-assisted tomography) scans, combine X-rays with computer technology to create cross-sectional images of the patient's body, which are then assembled into a three-dimensional picture that displays organs, bones, and tissues in great detail. MRI scanners use magnets and radio waves instead of X-rays to generate images that provide an even better view of soft tissues, such as the brain or spinal cord (President's Committee of Advisors on Science and Technology, 2000).

Much of today's imaging technology relies on microprocessors and software. In addition, the great advances in noninvasive sensing, tomography, and imaging technologies now allow repeated studies with minimal stress and damage (National Research Council, & Institute for Laboratory Animal Research, 2002).

Medical imaging is often thought of as a way of viewing anatomical structures of the body. Indeed, X-ray computed tomography and magnetic resonance imaging yield exquisitely detailed images of such structures. It is often useful, however, to acquire images of physiologic function rather than of anatomy. Such images can be acquired by imaging the decay of radioisotopes bound to molecules with known biological properties. This class of imaging techniques is known as nuclear medicine imaging.

Although the mathematical sciences were used in a general way for image processing, they were of little importance in biomedical work until the development in the 1970s of computed tomography for the imaging of X-rays (leading to the CAT scan) and isotope-emission tomography (leading to positron-emission tomography [PET] scans and single-photon-emission computed tomography [SPECT] scans). In the 1980s, MRI eclipsed the other modalities in many ways as the most informative medical imaging methodology (Webb, 1988).

Table 1 summarises some of the imaging methods used in biomedical applications.

Technologies such as those in Table 1 are all being investigated in small-animal models. The goal

Table 1. Imaging methods used in biomedical applications

- X-ray projection imaging (discovered in 1895)
- X-ray CT (1972)
- MRI (1980)
- Magnetic resonance spectroscopy (MRS)
- SPECT
- PET (1974)
- Gamma camera (1958)
- Nuclear magnetic resonance (NMR, 1946)
- Ultrasonics
- Electrical source imaging (ESI)
- Electrical impedance tomography (EIT)
- Magnetic source imaging (MSI)
- Medical optical imaging
- Micro computerised axial tomography (MicroCAT)
- Optical and thermal diagnostic imaging (OCT, DOT)

is to marry fundamental advances in molecular and cell biology with those in biomedical imaging to advance the field of molecular imaging (TA-Datenbank-Nachrichten, 2001). The two basic starting points in evaluating the overall utility of a medical technology are efficacy and safety. If a technology is not efficacious, it should not be used. In addition, efficacy and safety data are needed to evaluate the cost effectiveness of a technology (Banta, Clyde, & Williams, 1981).

Biomedical imaging devices have been used to obtain anatomical images and to provide localised biochemical and physiological analysis of tissues and organs. The ability of these devices to provide anatomical images and physiological information has provided unparalleled opportunities for biomedical and clinical research, and has the potential for important improvements in the diagnosis and treatment of a wide range of diseases (National Institute of Biomedical Imaging and Bioengineering [NIBIB], 2002).

Technological devices visualise and enlarge somatic space, rendering images of our most infinitesimal cells, molecules, and genetic structures, which allows for a more precise manipulation of our muscles, tissues, and bones (Sawchuck, 2000). Imaging tests now provide much clearer and more detailed pictures of organs and tissues. New imaging technology allows us to do more than simply view anatomical structures such as bones, organs, and tumours. Functional imaging—the visualisation of physiological, cellular, or molecular processes in living tissues—enables us to observe activity such as blood flow, oxygen consumption, or glucose metabolism in real time.

Imaging technology already has had lifesaving effects on our ability to detect cancer early and more accurately diagnose the disease (especially the PET device). Generally, the purpose of the biomedical imaging techniques is the early detection, clinical diagnosis, and staging of a disease, and therapeutic applications (*Biomedical Imaging Symposium: Visualizing the Future of Biology and Medicine*, 1999). Imaging technologies have many applications in biomedicine. Oncology, cardiology, and ophthalmology are only some of its sections that use these technologies, which everyday are developed more and more.

PROBLEMS AND DISADVANTAGES OF IMAGING TECHNOLOGIES

Despite all the promises, the use of imaging technologies in biomedicine and bioengineering evoke many problems. All biomedical imaging devices suffer from various limitations that can restrict their general applicability. Some major limitations are sensitivity, spatial resolution, temporal resolution, and the ease of the interpretation of data. One way to circumvent these limitations is to develop technological and methodological approaches that improve and extend the sensitivity and the information content of individual imaging techniques. Another way is to combine two or more complementary biomedical imaging techniques (like MRI and PET, MRI and MEG, and optical MRI).

Table 2 summarises some problems of the imaging technologies.

There is no crystal ball to predict the future of medical imaging technologies. New applications continue to be explored for both diagnosis and treatment (Canadian Institute for Health Information, http://www.cihi.ca). Biomedical imaging has seen truly exciting advances in recent years. New imaging methods can now reflect internal anatomy and dynamic body functions heretofore only derived from textbook pictures, and applications to a wide range of diagnostic and therapeutic procedures can be envisioned. Not only can technological advances create new and better ways to extract information about our bodies, but they also offer the promise of making some existing imaging tools more convenient and economical.

Advances based on medical research promise new and more effective treatments for a wide

Table 2. Problems and disadvantages of imaging technologies

- The high cost of equipment and their maintenance, which aggravates the national economy for medicine
- Wasteful expenditures because of bad usage by users and technical staff (20 to 40%)
- The technology changes rapidly and devices may become out of date
- Users need education to learn how to break the new technologies in
- Physicians and the nursing staff must continuously be acquainted through articles related to the new technologies and equipment
- New technologies cause disruption and disappointment for staff
- They are venturous for patients because the levels of radiation they are exposed to may be too high

variety of diseases. New noninvasive imaging techniques for the earlier detection and diagnosis of disease are essential to take full advantage of new treatments and to promote improvements in healthcare. The development of advanced genetic and molecular imaging techniques is necessary to continue the rapid pace of discovery in molecular biology. Several breakthrough imaging technologies, including MRI and CT, have been developed primarily abroad (American Institute for Medical and Biological Engineering, http://www.aimbe.org).

Key paradigms of emerging imaging technologies from different technological areas will be presented, and the engineering principles and research findings leading to the design of efficient bioimaging technologies will be introduced and analysed. Specifically, imaging technologies from space or aerospace research have been identified and successfully applied toward the development of novel high-resolution, multisensor medical imaging systems, with potential applications in digital radiography and CT. Similarly, experimental research findings for defence applications have been applied toward the development of multifusion optical sensing imaging systems and techniques for efficient disease detection (Giakos, 2003).

Today, as for all products and services in all sectors, there exists the DICOM (Digital Imaging and Communications in Medicine) Standards Committee. Its purpose is to create and maintain international standards that help the allocation of medical pictures (like radial tomographies, magnetic tomographies, etc.), and the communication of biomedical diagnostic and therapeutic information in disciplines that use digital images and associated data (DICOM, 2004). DICOM is used or will be used by every medical profession that utilises images within the healthcare industry.

CONCLUSION

The rapid progress in imaging technologies during the last decades has stimulated many developments and applications in medicine, biology, industry, aerospace, remote sensing, meteorology, oceanography, and the environment.

New developments are continually making the technology faster, more powerful, less invasive, and less expensive. Imaging technology was primarily used in medical diagnosis initially, but it is being increasingly used in pure neuroscience, psychological research, and many other fields. The quantitative nature of data will be relevant for the effective diagnosis as well as therapeutic management of patients, whichever disease they have (“Nuclear Medicine Sextet,” 1999).

REFERENCES

Banta, H. D, Clyde, J. B., & Williams, J. S. (1981). *Toward rational technology in medicine: Considerations for health policy*. New York: Springer Verlag.

Biomedical Imaging Symposium: Visualizing the Future of Biology and Medicine. (1999, June 25-26). Natcher Conference Center, National Institutes of Health, Bethesda, MD. Sponsored by the NIH Bioengineering Consortium (BECON), the American Institute for Medical and Biological Engineering (AIMBE), and the Radiological Society of North America (RSNA). Retrieved from http://www.becon.nih.gov/report_19990625.pdf

Digital Imaging and Communications in Medicine (DICOM). (2004). *Scope of DICOM: Strategic document, Version 4.0*. VA: National Electrical Manufacturers Association.

Giakos, G. (2003). *Emerging imaging technologies: Technology identification transfer and utilization for bioengineering applications.* McMaster University. Retrieved October 28, 2003, from http://www.ece.mcmaster. ca/news/seminars/giakos_sem.htm

National Institute of Biomedical Imaging and Bioengineering (NIBIB). (2002). *Improvements in imaging methods and technologies*. Retrieved from http://www.nibib1.nih.gov

National Research Council (U.S.), & Institute for Laboratory Animal Research (U.S.) (2002, April 17-19). *International perspectives: The future of non-human primate resources. Proceedings of the Workshop.* Retrieved from http://www.nap.edu/catalog/10774.html

Nuclear medicine sextet. (1999, August 21). *The Lancet*, 665.

President's Committee of Advisors on Science and Technology (2000, Spring). *Chapter 3. Biomedical technologies.* In *Wellspring of prosperity: Science and technology in the U.S. economy*. Washington, DC: Office of Science and Technology Policy. Retrieved from http://www.ostp.gov

Sawchuck, K. (2000). *Digibodies online exhibition and Synapse online forum*. Retrieved from http://www.digibodies.org/synapse/intro.html

TA-Datenbank-Nachrichten. (2001). *Nr. 1/10. Jahrgang: März, S. 13-22*. Retrieved from http://www.itas.fzk.de

Webb, S. (1988). The physics of medical imaging. *Magnetic Resonance Imaging and Biophysics and Medical Imaging.* Retrieved from http://newton.ex.ac.uk/teaching/modules/PHYM433.html

KEY TERMS

Assessment of Imaging Technology: Research on and development of methods for the evaluation and comparison of new and existing imaging technologies to establish their effectiveness, robustness, and range of applicability.

Bioengineering: The application of engineering principles to the fields of biology and medicine, as in the development of aids or replacements for defective or missing body organs. It is also called biomedical engineering.

Biomedicine: A branch of medical science concerned especially with the capacity of human beings to survive and function in abnormally stressful environments and with the protective modification of such environments. Broadly, it is medicine based on the application of the principles of the natural sciences, especially biology and biochemistry.

Development of Imaging Devices: Research and development of generic biomedical imaging technologies before specific applications are demonstrated.

Diagnostic Imaging: A study section reviews applications dealing with the development and evaluation of new technology for imaging, including instrumentation and software for producing, evaluating, storing, and transmitting images for anatomical, physiological, metabolic, diagnostic, and therapeutic information.

Image Exploitation: Development, design, and implementation of algorithms for image processing and information analysis, including advanced methods for the acquisition, storage, and display of images; research and development on image-guided procedures; and techniques for using multidimensional images to understand physiology and normal and abnormal function.

Medical Device: Any instrument, apparatus, appliance, material, or other article, whether used alone or in combination, including the software necessary for its proper application, intended for the purpose of the diagnosis, prevention, monitoring, treatment, alleviation, or investigation of a disease, injury, or handicap.

Medical Imaging: Term describing the various technologies that produce pictures or images of the body and its structures. Imaging technologies include X-ray, CT scanning, PET scanning, and ultrasound. This term also includes technology such as digital cameras, which produce digital images.

Medical Imaging Technologies: A study section reviews all modalities of medical imaging, including gamma ray; MRI; functional MRI; PET; SPECT; X-ray; CT; visible, infrared, and ultraviolet photons; and optical, photo-acoustic, microwave-acoustic, and exotic imaging methods.

Minimally Invasive Technologies: Basic research involving the use of robotics technologies for actuation, sensing, control, programming, and the human-machine interface, and the design of mechanisms to determine research end points such as diagnosis and the automated or remote treatment of disease.

This work was previously published in the Handbook of Research on Informatics in Healthcare and Biomedicine, edited by A. A. Lazakidou, pp. 271-276, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.19
Inertial Sensing in Biomechanics:
Techniques Bridging Motion Analysis and Personal Navigation

Angelo M. Sabatini
ARTS Lab, Scuola Superiore Sant'Anna, Italy

ABSTRACT

Sensing approaches for ambulatory monitoring of human motion are necessary in order to objectively determine a person's level of functional ability in independent living. Because this capability is beyond the grasp of the specialized equipment available in most motion analysis laboratories, body-mounted inertial sensing has been receiving increasing interest in the biomedical domain. Crucial to the success of this certainly not new sensing approach will be the capability of wearable inertial sensor networks to accurately recognize the type of activity performed (context awareness) and to determine the person's current location (personal navigation), eventually in combination with other biomechanical or physiological sensors — key requirements in applications of wearable and mobile computing as well. This chapter reviews sensor configurations and computational techniques that have been implemented or are considered to meet the converging requirements of a wealth of application products, including ambulatory monitors for automatic recognition of activity, quantitative analysis of motor performance, and personal navigation systems.

INTRODUCTION

At present, human motor performance can be accurately assessed with several measuring instruments, the use of which is spread in many motion analysis laboratories throughout the world. The most important technology used to detect and track human body motion is video-motion sensing. In common with other motion tracking technologies, such as infrared, electromagnetic, ultrasound, video-motion sensing is externally referenced, in the sense an external source — optical, magnetic, acoustic — is needed to determine position and orientation information concerning the moving object of interest (Meyer, Applewhite, & Biocca, 1992). Usually, this source is effective over a relatively small working space. In addition to the range restriction, interference, distortions and occlusions can easily result in erroneous location and orientation information, thereby leading, in critical situations, to a complete loss of track (You, Neumann, & Azuma, 1999). The availability of dedicated laboratory setups is, therefore, a prerequisite for the application of externally referenced sensing techniques. However, from a clinical viewpoint, motor performance measured in laboratory settings may not accurately reflect functional ability in daily-life environments, since behavior of patients in laboratory is not necessarily representative of their daily-life behavior. There is a need for ambulatory monitoring systems that are able to provide objective assessment of human functional ability in the absence of the behavioral modifications induced by performing within constraining laboratory settings.

The capability of inertial sensors of sensing their own motion is the sourceless feature that makes them so attractive for the development of ambulatory monitoring systems (Verplaetse, 1996). Body-mounted sensors of this kind make it possible to determine position and orientation information based on the measurement of physical quantities (acceleration, angular velocity), which are directly related to the motion of the body part where they are positioned. Being internally referenced, inertial sensors can then be proposed to detect and track body motion over a virtually unrestricted working space. Until recent years, inertial sensors have only found use to monitor the motion of man-made vehicles, including spaceships, planes, ships, submarines, cars, and, more recently, wheeled and legged robots. Recent advances in microelectromechanical systems (MEMS) technologies have led to the development of a new generation of inertial sensors (Bachmann, Yun, McKinney, McGhee, & Zyda, 2003), the specifications of which — in terms of encumbrance, robustness, power consumption, measuring performance and cost — seem to be appropriate for applications in the biomedical field.

In inertial systems, the main problem is that position and orientation are found by time-integrating the signals from accelerometers and gyroscopes, as well as any sensor drift and noise superimposed to them. As a result, position and orientation errors tend to grow unbounded. This problem is especially acute when low-cost MEMS inertial sensors are used. Their sensitivity and bias stability are, in fact, orders of magnitude less than the sensitivity and bias stability of the high-grade inertial sensors that are embedded in military and aviation navigation systems (Foxlin, 2002). Another drawback of inertial sensors is that they are not well-suited for determination of absolute position and orientation. In order to be accurate, the integration process needs to be started from accurately known initial conditions, which inertial sensors are unable to provide at all (position and velocity), or can provide to just a limited extent (orientation). Hence, the use of inertial systems is most effective in those applications which involve relative motion.

In this chapter, we are not interested in the viewpoint of those who aim at designing new and better sensors. Rather, we intend to survey the main computational techniques that have been investigated so far, in the effort to take measurements from available sensors and construct the best possible characterization of human body motion. The traditional approach to data processing for navigational purposes has involved the development of filtering algorithms to fuse measurements from inertial sensors and other sensors, such as global positioning system (GPS) receivers and Earth's magnetic field sensors. It has also involved the exploitation of suitable environment maps so as to deal with the person's location uncertainty during indoor and outdoor navigation. Other computational techniques have been designed with a stronger biomechanical inspiration. They specifically aim at obtaining valuable information about either absolute or relative motion of body parts from simple configurations of single or multiple inertial sensors, in the effort to gain a deeper understanding of how we humans perform functional activities, including gait, balance and postural transitions, in normal or pathological conditions.

The borderline between the use of inertial sensors as functional and navigational sensors in biomechanics is becoming increasingly fuzzy. As it will be seen, a better prediction of the errors incurred in the mathematical reconstruction of the walked path can be achieved by exploiting signatures which characterize how certain functional activities such as walking are performed. Conversely, the knowledge of temporo-spatial features of the walked path, reconstructed from navigational sensors, has relevant functional implications, when, for instance, an objective assessment of the person's physical activity in daily-life conditions is desired. Ideally, we would like to determine accurately the current person's location and recognize what he or she is doing and how, without incurring in any significant limitation in the extension of the measurement space and in the duration of the observation records.

At present, ambulatory monitoring systems based on sole inertial sensing cannot accomplish this ambitious goal, although they can already be used with remarkable success. In this chapter we will survey how inertial sensing has been applied in biomechanically-oriented works, with particular emphasis on the sensor configurations and adopted signal processing techniques. In doing that, we hope to give convincing support to the claim that another valuable analytical weapon is ready for inclusion in the repertoire of tools available to measure and study human motor performance.

BACKGROUND

Introductory Remarks

An inertial sensor is a motion sensor the references of which are internal, with the exception of initialization (Curey, Ash, Thielman, & Barker, 2004). It can be used for the determination of position and orientation by measuring physical quantities, such as linear accelerations and angular velocities. An inertial sensor assembly (ISA) is a structure that contains multiple inertial sensors (accelerometers and/or gyroscopes) in fixed orientations relative to one another. An inertial measurement unit (IMU) measures linear and angular motion in the three-dimensional space without external references. Using the outputs from an IMU, an inertial navigation system (INS) estimates a body's position and orientation, in connection with a gravitational field model and the operation of a reference clock.

Earth's magnetic field sensing is formally externally referenced, although it can be used in practice as though it is internally referenced. This is because the Earth's magnetic field is available anywhere, and no environmental modifications are required to operate an Earth's magnetic field sensor, or magnetic sensor in short. The importance of magnetic sensors in human-machine interface applications and human body motion tracking systems has been widely acknowledged (Foxlin, 2002), in part because of the recent market availability of miniaturized tri-axis inertial/magnetic sensors packages. Using a geomagnetic field model, magnetic sensors can help specify the location of an INS during initialization, and mitigate the INS error growth during flight. Magnetic sensors will be included in the sensor configurations of interest to our review in this chapter.

Inertial and Magnetic Sensors

Accelerometers are sensors that measure the applied acceleration acting along a sensitive axis. A range of different transducers are available to measure the acceleration. In virtually all applications in human movement science piezoresistive accelerometers are used (Verplaetse, 1996). Piezoresistive materials have the property of modifying their electrical resistance when they are strained or deflected. In terms of functionality, body-mounted piezoresistive accelerometers respond to either static accelerations, such as gravity, or dynamic accelerations, such as linear/rotational accelerations due to body movements (Figure 1). When exposed just to gravity, a body-mounted tri-axis accelerometer can be used to estimate the inclination relative to the vertical of the body segment where the device is positioned.

Gyroscopes, gyros in short, are sensors that measure the rate of rotation about a sensitive axis. The most popular design of a gyro for applications in the field of human movement science is of the vibrating type (Verplaetse, 1996). Vibrating gyros use the Coriolis acceleration effect to sense when they rotate. In the absence of rotation, the vibrating element embedded in the sensor continuously vibrate within a plane. In the presence of a rotation about an axis perpendicular to the plane of vibration, the vibrating element deviates from its plane of vibration under the Coriolis acceleration effect and the amplitude of the out-of-plane vibration is proportional to the rate of rotation (Abbott & Powell, 1999). Gyros with a vibrating element are more sensitive to temperature and shock than accelerometers, due to mechanical fastening of the vibrating beam inside the sensor case. Compared with accelerometers, gyros have also relatively higher cost and larger size.

Magnetic sensors resolve the components of the local magnetic field along a sensitive axis. A magnetic compass is an electronic device that measures its heading relative to magnetic north by measuring the direction of Earth's local magnetic field (Abbott & Powell, 1999). The Earth's magnetic field has a component parallel to the Earth's surface that always points toward magnetic north, hence its projection on the horizontal plane can be used to determine compass direction (Caruso, 1997). Common designs of magnetic sensors integrate three magnetic sensors with sensitive axes mutually perpendicular in the same package, so as to reconstruct the horizontal component, provided that the inclination of the sensor case is estimated from gravity sensors (tilt compensation) (Figure 2).

In applications in the field of human movement science, the capability to sense the Earth's magnetic field is the horizontally-sensing principle needed to complement the vertically-sensing principle of accelerometers, so as to enable three-dimensional monitoring of orientation (Kemp, Janssen, & van der Kamp, 1998).

Figure 1. A piezoresistive single-axis accelerometer measures the projection (in the direction of the sensitive axis) of an equivalent acceleration resulting from the sum of the acceleration due to the accelerometer's own motion and the equivalent gravity acceleration acting on its mechanical structure.

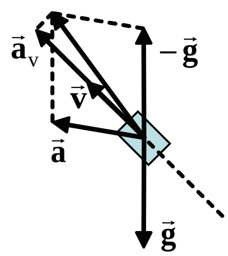
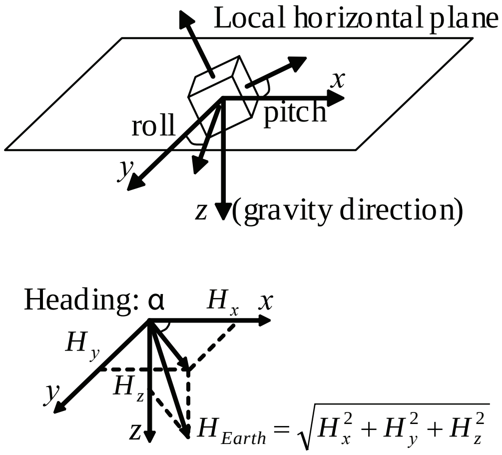


Figure 2. Inclination of the compass device relative to the Earth's local horizontal plane as defined by gravity direction (a). Earth's magnetic field resolved in (x, y, z) coordinates (b).



The Strapdown Approach to INS Design

Consider a rigid body in space and define a coordinate system attached to the body — the body frame, $\mathcal{B}$. In a strapdown INS, the sensitive axes of the ISA sensors are permanently aligned along the axes of $\mathcal{B}$ (Curey et al., 2004). Orientation determination means to specify the orientation of the axes of $\mathcal{B}$ with respect to the Earth's reference frame $\mathcal{E}$; position determination means to specify the origin of $\mathcal{B}$ relative to $\mathcal{E}$.

Suppose that a vector $\vec{\mathbf{x}}$ is resolved along the axes of $\mathcal{B}$ and $\mathcal{E}$, to yield the 3×1 column-vectors $\vec{\mathbf{x}}^B$ and $\vec{\mathbf{x}}^E$ respectively. The transformation between the 3×1 column-vector $\vec{\mathbf{x}}$\ as expressed in the body frame and in the reference frame is:

$$\vec{\mathbf{x}}^B = \overset{E\to B}{C}\ \vec{\mathbf{x}}^E \qquad (1)$$

where $\overset{E\to B}{C}$ is the direction cosine matrix for the transformation from $\mathcal{E}$ and $\mathcal{B}$. Furthermore, suppose a single-axis accelerometer is attached to the body at point $\vec{\mathbf{p}}(t)$, and the orientation of its sensitive axis is denoted by the unit vector $\vec{\mathbf{v}}(t)$. In principle, the accelerometer measures the projection of the acceleration of point $\vec{\mathbf{p}}(t)$, along $\vec{\mathbf{v}}(t)$.

$$\vec{\mathbf{a}}^E(t) = \left[\dot{\vec{\omega}}(t)\times\vec{\mathbf{p}}(t) + \vec{\omega}(t)\times\vec{\omega}(t)\times\vec{\mathbf{p}}(t) + \vec{\mathbf{a}}_o(t) + \vec{\mathbf{g}}\right]\cdot\vec{\mathbf{v}}(t) \qquad (2)$$

where $\vec{\mathbf{a}}_o(t)$ is the acceleration of the origin of $\mathcal{B}$ $\vec{\mathbf{o}}(t)$, is the gravitational acceleration, and $\vec{\omega}$ is the angular velocity of $\mathcal{B}$ relative to $\mathcal{E}$, resolved in $\mathcal{B}$. In (2), × and · denote the standard vector cross-product and dot-product, respectively.

The determination of the direction cosine matrix, also called attitude matrix, is usually performed by numerically integrating non-linear systems of first-order differential equations from initial conditions that are assumed to be known (initialization, or alignment) (Bortz, 1971). In this approach, the differential equations involve the components of $\vec{\omega}$, which are usually measured by a tri-axis gyro. Alternatively, the attitude can be determined by vector matching in $\mathcal{E}$ and $\mathcal{B}$ (Wertz, 1978).

Vector matching, also known as Wahba's problem, calculates the attitude solution by matching two non-zero, non-colinear vectors that are known in $\mathcal{E}$ and measured in $\mathcal{B}$. The Earth's gravitational field and the Earth's magnetic field are two these vectors. Vector matching solves the alignment process and forms, in principle, an alternative to the use of gyros in determining the attitude matrix (Gebre-Egziabher, Elkain, Powell, & Parkinson, 2000). While suitable to track slow movements, this gyro-free method is not suited for fast movements, yielding quite large orientation errors during the motion.

To understand why, let $\vec{\mathbf{o}}(t)$ be the point whose position in $\mathcal{E}$ is to be determined. It is apparent from (2) that the output of a body-mounted accelerometer depends on where it is placed, its orientation relative to the subject, the subject's posture and activity. In the absence of motion, the accelerometer output is determined by its orientation relative to the vertical. Knowing the accelerometer orientation relative to the subject, the accelerometer can be used to determine the subject's orientation relative to the vertical. In the presence of motion, the resulting signal depends on the subject's posture and activity.

The orientation estimate is used twice in the strapdown approach to position determination. First, the accelerometer senses the body's acceleration $\vec{\mathbf{a}}_o^B(t)$ as the superposition of the sensed acceleration in $\mathcal{B}$, $\vec{\mathbf{a}}^B(t)$ and the projection of the gravitational acceleration $\vec{\mathbf{g}}$ on $\mathcal{B}$:

$$\vec{\mathbf{a}}_o^B(t) = \vec{\mathbf{a}}^B(t) - \overset{E\rightarrow B}{C}(t)\,\vec{\mathbf{g}} \qquad (3)$$

The next step requires integration of $\vec{\mathbf{a}}_o^E(t) = \overset{B\rightarrow E}{C}(t)\vec{\mathbf{a}}_o^B(t)$ to derive the position estimate:

$$\vec{\mathbf{o}}^E(t) = \int_0^t dt' \int_0^{t'} \overset{B\rightarrow E}{C}(t')\,\vec{\mathbf{a}}_o^B(t'')dt''. \qquad (4)$$

Any numerical integration routine can be used to solve (4). The flow of information underlying the strapdown approach is illustrated in Figure 3.

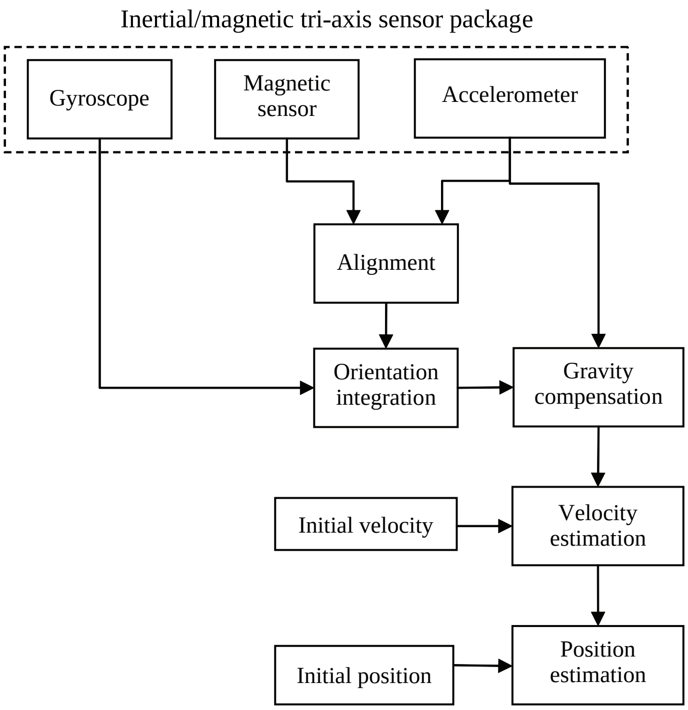
Drift in the orientation determination comes from gyro biases, defined as the output produced by gyros at rest. Uncompensated gyro biases lead to a constant rate of drift after integration. More than the biases in itself, their time-stability is a matter of concern. Time-stable biases can be measured when the INS is stationary and compensated for. There is also drift in the linear position determination which arises from several sources. First, there are the effects of accelerometer biases, the compensation of which is not so straightforward as the compensation of the gyro biases (Lötters, Schipper, Veltink, Olthius, & Bergveld, 1998). Second, since position is obtained by double integration, uncompensated accelerometer biases result in a position drift error which grows quadratically in time. Third, the orientation errors incurred by gyros lead to imperfect gravity cancellation during the double integration.

Additionally, the capability of geomagnetic compassing to correct drift is poor in many environments, because the Earth's magnetic field is weak and easily masked by magnetic disturbances within or near the measurement space.

In conclusion, it is critical to achieve high accuracy in position/orientation determination by strapdown INSs incorporating low-cost inertial/magnetic sensors, the stand-alone accuracy and run-to-run stability of which are poor. Different applications may involve different accuracy requirements relative to the duration of each observation run. In the absence of special precautions, the requirements of human motion tracking applications are shown to be violated when the duration of the observation run exceeds just several seconds (Foxlin, 2002).

There appear to be different means to deal with these problems, as stated below. An approach is to use externally referenced aids, such as GPS, and carry out the integration process underlying the combined use of GPS and INS technologies with Kalman filters (KFs). Another approach is

Figure 3. Strapdown approach to the design of an inertial navigation algorithm



to exploit idiosyncrasies of the human motion dynamics by designing algorithms that can keep the drift rate low (Foxlin, 2002; Sabatini, 2005). Alternatively, in some applications, the difficulty inherent in the strapdown approach can be circumvented by using signal features the construction of which do not require time integration from uncertain initial conditions.

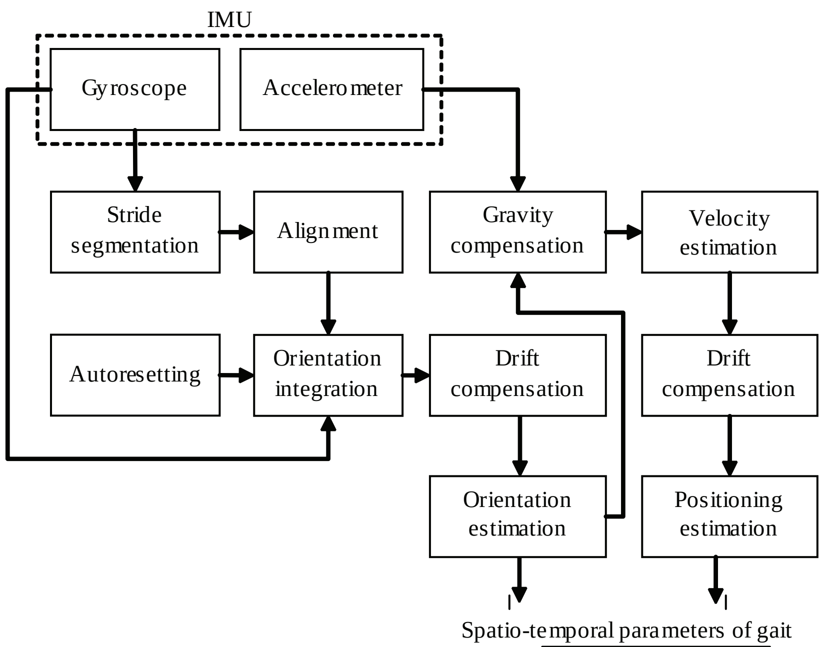
TAXONOMY OF APPLICATIONS

Ambulatory Monitoring of Human Motion

The use of accelerometry to perform ambulatory measurements of human motion is well established (Gage, 1964; Smidt, Arora, & Johnston, 1971; Morris, 1973). Morris (1973) suggests to calculate the general displacement of a rigid body with one point fixed—a rotation about an axis through the fixed point—from three acceleration measurements, followed by three additional acceleration measurements to resolve the motion of the fixed point with respect to the Earth's reference frame. Valuable considerations are made by this author concerning the problem of IMU alignment and drift correction. In the case of shank motion during level walking, the redundancy of the information on shank motion during the foot-flat phase of the walking cycle is exploited to calculate the initial orientation (alignment). Because of the cyclical nature of leg motion, the initial angular velocity and rotation, translation acceleration and velocity can be made equal to the values of these functions at the beginning of the next cycle (drift correction).

IMUs composed of several single-axis accelerometers with their sensitive axes properly positioned in space relative to one another have been proposed and lead, for the estimation of the components of angular velocity, to systems of nonlinear coupled differential equations (IMU with six accelerometers) and linear uncoupled

Figure 4. Strapdown approach to the design of an inertial navigation algorithm, specifically intended for gait analysis applications (Sabatini, 2005)



The need for alignment exists unless the change of orientation is the quantity of interest. This is the case when, for instance, the change of orientation is introduced in simple symmetric gait models of extended limbs to determine stride length (Miyazaki, 1997).

Assessment of Gait

A full kinematic analysis of the body motion based on body-mounted sensors would be impossible without incurring in the "Christmas tree" effect (Meijer, Westerkerp, Verhoeven, Koper, & ten Hoor, 1991), even in the limit of a simplified two-dimensional analysis (Mayagoitia, Nene, & Veltink, 2002). Fortunately, in clinical practice, the values of simple temporo-spatial parameters of gait, including stride time/length, cadence and walking speed, turn out to be good indicators of how well a person walks, and of how effective are different rehabilitation techniques or exercise programs in facilitating ambulatory recovery (Weir & Childress, 1997; Maki, 1997). Valuable information concerning a person's gait can also be obtained by measuring temporo-spatial features related to the displacement of particular anatomical points, such as the vertical displacement of the sacrum during walking (Weir & Childress, 1997).

Simple parameters of gait can be determined from waist, thigh, heel accelerations (Evans, Duncan, & Gilchrist, 1991; Auvinet, Chaleil, & Barrey, 1999; Aminian et al., 1999a; Zijlstra & Hof, 2003); from thigh, shank, heel, foot instep angular velocities (Miyazaki, 1997; Tong & Granat, 1999; Pappas, Popovic, Keller, Dietz, & Morari, 2001; Aminian, Najafi, Büla, Leyvraz, & Robert, 2002; Sabatini et al., 2005). The parameterization of gait requires the detection of subsequent foot contacts. The commonest approach to the detection of foot contacts with inertial sensors is based on a search for distinctive and stable signal features occurring at the time of heel-strike (when the foot first touches the floor) and toe-off (when it last takes off) during a gait cycle. Usually, the detection and time-localization of these features require (a combination of) peak or zero-crossing detectors applied to signals contaminated by relatively high-frequency noise components. In a number of papers (Aminian et al., 1999; 2002; Paraschiv-Ionescu, Buchser, Rutschmann, Najafi, & Aminian, 2004) the use of wavelet packages is proposed to perform the process of gradually focusing on the exact location of heel-strikes and toe-offs. In particular, the claim of these authors is that wavelet packages are well adapted for gait events identification, such as heel-strike and toe-offs, because they allow detection of a specified frequency at a specified time. In general, for biomechanical signal processing applications, the superiority over standard filtering techniques of wavelet denoising and smoothing of kinematic data related to complex motions is documented (Ismail & Asfour, 1999). Another good point in favour of wavelets is that the existence of fast algorithms for carrying the multiresolution analysis would allow to implement gait parameterization in real-time conditions, although this point has not been yet investigated.

The spatial parameters of gait can be determined by so-called indirect methods or by integrating the strapdown equations (direct method). Indirect methods are based on the recognition of few important features of human's walking patterns. These include the good correlation existing between the stride length and the cadence, or the influence of walking speed on the amplitudes of the acceleration signals at different point on the legs. Recent studies demonstrate that during walking trunk accelerations exhibit patterns with fixed relationships to spatio-temporal gait parameters. Zijlstra and Hof (2003) analyze the 3D-acceleration pattern of the lower trunk at different walking speeds, and develop algorithms to determine the instant of foot contacts during each step. Based on a simple inverted pendulum model of the body's center of mass (BCOM) trajectory, the step length is predicted by the peak-to-peak trunk vertical displacement, which is obtained from double integration of the high-pass filtered vertical component of trunk acceleration. An advantage of computing peak-to-peak displacements is that sensor calibration and initialization issues are insignificant, although inaccurate placement of the sensor device in the frontal plane is a matter of concern for achieving good test-retest reliability, as discussed by Moe-Nillsen (1998), who prefer to work in the domain of root mean square (RMS) values from 3D-waist accelerometer data.

In the attempt to include the incline of the walked surface in the number of variables that are known to correlate with walking speed, the parameterization of body accelerations at the waist and heel is performed in (Aminian, Robert, Jéquier, & Schutz, 1995) on the basis of simple features, such as means, standard deviations, correlation coefficients between various acceleration components, step time, and so forth. Two artificial neural networks (ANNs) are designed, one for predicting the walking speed (averaged over a single stride), one for predicting the incline of the walked surface, according to standard design procedures for training and testing multilayer perceptrons. Subject-specific training is performed by presenting examples of treadmill walking at different combinations of walking speed/incline, so as to account for individual walking styles. Testing is performed during overground walking. Results show that only marginal improvements are due to using ANN technology in the walking speed prediction, as compared with the results of a stepwise linear regression analysis. Conversely, owing to the possible presence of strong nonlinearities in the complex relationships between the acceleration parameters selected for the study and the incline, the ANN use is shown to remarkably improve the incline estimate. It is worthy noting that ANNs exhibit significant generalization abilities in this application, although the training phase must be necessarily subject-specific.

Common to all indirect methods, intra-individual physiological variability and environmental conditions heavily influence the accuracy of the relationships exploited to infer the quantities of interest. Hence, frequent calibration procedures may be needed, which requires additional sensors (Perrin, Terrier, Ladetto, Merminod, & Schutz, 2000; Terrier, Ladetto, Merminod, & Schutz, 2000). Another limitation of indirect methods is that they are approaches which suffer from changes in walking style. It is expected that, in particular situations, such as crowded streets and many indoor environments, the human displacements occur with frequent walking strategy adaptations. Finally, the generalization implied by indirect methods is critical when walking is performed over irregular terrain or in the face of frequent changes of inclines.

At the expense of greater computational costs, direct methods would be more appropriate than indirect methods. Anecdotal evidence highlights the remarkable accuracy of direct methods for the estimation of gait spatio-temporal parameters (Sagawa, Sato, & Inooka, 2000; Veltink et al., 2003). In support of this claim, a recent study analyzes the performance of a direct method during level/uphill/downhill treadmill walking (Sabatini et al., 2005).

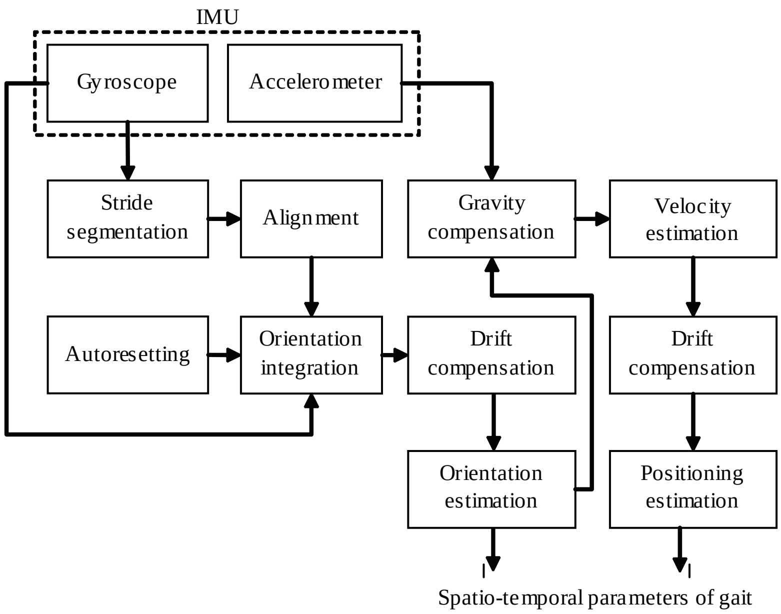
Assessment of Physical Activity and Energy Expenditure of Daily-Life Activities

Physical activity is an important part of human behavior and it may be related to various aspects of health and disease; intense research is therefore undergoing to find means for its objective assessment (Mejier et al.,1991). The complex nature of physical activity makes this goal very difficult to pursue. At present, the energy expenditure due to physical activity is widely accepted as the standard reference for physical activity. Together with heart rate monitoring, motion sensors such as accelerometers are considered suitable to obtain a quantitative estimation of the energy expenditure incurred in daily-life activities.

The use of accelerometers is based on relationships between accelerometer output and energy expenditure or oxygen uptake that are documented in several studies on gait analysis and ergonomics (Bouten, Koekkoek, Verduin, Kodde, & Janssen, 1997; Bussmann, Hartgerink, van der Woude, & Stam, 2000). Regardless where the right location of body-mounted accelerometers is for the best prediction of the energy expenditure, or which the effects of placement and orientation are (Bouten, Sauren, Verduin & Janssen, 1997), we cannot expect extremely accurate estimates of energy expenditure with the use of accelerometers (Bussmann et al.,2000). Albeit dynamic activities such as walking greatly contribute to daily-life physical activity, and are a prerequisite to many other activities, the contribution to energy expenditure due to static activities is also relevant, for instance weight bearing in the upright posture, work of isometric contractions, and so forth. Of course, motion sensors are totally blind to this part of the energy budget.

Since double integration of acceleration to retrieve position and compute external work is not believed accurate enough for long-term monitoring, the general consensus is that the summation of the time integrals of acceleration components in various measurement directions, and the time integral of the acceleration vector norm result in the most accurate prediction of energy expenditure during dynamic activities (Bouten et al.,1997). More precisely, the various acceleration components are high-pass filtered to subtract gravity influence; the numeric integration is carried out, then the integral(s) are normalized to the duration of the integration interval. Possibly, energy expenditure estimation can be improved by adopting nonlinear regression models to capture the relationship between the acceleration statistical feature and the energy expenditure (Chen & Sun, 1997). The improvement is especially important in estimating light-intensity activities, whose energy expenditure is generally underestimated by linear models (Bouten et al.,1997a).

Figure 4. Strapdown approach to the design of an inertial navigation algorithm, specifically intended for gait analysis applications (Sabatini, 2005)



Besides walking speed, the assessment of the energy expenditure during walking is influenced by many other parameters, including the incline. Without an external measurement of the slope, the standard method of analysis of body accelerations cannot accurately predict the energy cost of uphill/downhill walking (Terrier, Aminian, & Schutz, 2001). Use of additional processing techniques, such as ANNs (Aminian et al.,1995), have shown success in estimating the incline along which the subject is walking. Use of a direct method based on strapdown integration from a foot IMU is discussed in Sabatini, Martelloni, Scapellato and Cavallo (2004), in connection with the use of empirical models for determining the energy cost of walking.

Automatic Recognition of Activity

Another problem existing with the assessment of physical activity by motions sensors as previously described is that these methods do not provide information on the type of activity. In order to assist a clinician in diagnosis, choosing the most suitable therapy for a patient, monitoring patient progress and assessing the effects of treatment, the "quantity" of motion is certainly an important issue, but aspects of mobility and locomotion related to the specific activity where energy is spent ought to be elucidated as well. Ambulatory monitors have been designed with the goal to identify and classify sets of postures and activities (Table 1).

Most systems have used only accelerometers (Bussmann, Veltink, Koelma, van Lummel, & Stam, 1995; Foerster & Fahrenberg, 2000; Mathie, Celler, Lovell, & Coster, 2003), while other systems have used accelerometers together with another type of sensor to improve discrimination. Najafi et al. (2003) and Parischev-Ionescu et al. (2004) are interesting works where inertial sensors of different types are integrated into ambulatory monitors where activity recognition and some (limited) form of gait analysis are jointly differential equations (IMU with nine accelerometers) (Padgaonkar, Krieger, & King, 1975). The angular velocity components are then used to estimate the attitude matrix (Bortz, 1971). Double-integration of gravity-compensated accelerometer signals yields positioning information. Giansanti, Macellari, Maccioni, and Cappozzo (2003) analyze these IMU configurations in depth, and conclude that severe restrictions exist in the time duration over which motion tracking is feasible by accelerometry methods in routine biomechanical applications.

The determination of angular acceleration is probably the commonest computational task solved by accelerometry for the ambulatory monitoring of human motion. The main application is the clinical assessment of gait. Under the simplifying assumption of a gait motion planar model, a minimal configuration set composed of two leg-mounted single-axis accelerometers with parallel sensitive axes would be enough. In the attempt to circumvent the problem of integration drift, pairs of accelerometers on each segment are used to resolve the relative angle between two segments, namely the joint angle, without integration (Willemsen, van Alsté, & Boom, 1990). The potential of such an accelerometric angle sensor is highlighted in (Willemsen, Frigo, & Boom, 1991), where the most important error sources are analyzed in detail, namely not fulfilling the gait motion planar model and the rigid-body condition by external fixation of body-mounted sensors.

Luinge and Veltink (2004) propose to use one tri-axis accelerometer to measure inclination during dynamic tasks without requiring additional sensors. A Kalman-based filtering algorithm is designed to separate the different acceleration components: gravity, linear acceleration, bias, measurement noise, in accordance with a model of the linear acceleration components which relies on reasonable assumptions about the dynamics of human motion. The resulting inclination error is shown to be significantly smaller than the inclination error obtained by the method of low-pass filtering the accelerometer signals, especially as the speed of motion increases. Usually, low-pass filtering is used to separate the gravitational component from the linear acceleration components (Veltink, Bussmann, de Vries, Martens, & van Lummel, 1996; Foerster & Fahrenberg, 2000; Williamson & Andrews, 2001). Unfortunately, since the procedure of bias compensation works only in the direction of gravity, the bias estimate in all measurement directions is accurate only when the sensor is rotated over large angles.

As for the computation of joint angles, an alternative to the use of accelerometers is represented by the use of gyros (Miyazaki, 1997; Williamson & Andrews, 2001). Two approaches are possible to compensate the gyro bias during integration: a) high-pass filtering of the signals (Miyazaki, 1997) — this technique is not recommended for non-cyclical motions, for instance postural transitions, for whom a zero frequency component of angular velocity cannot be expected; (b) automatic nulling and resetting (Williamson & Andrews, 2001). The underlying basis for the implementation of automatic nulling is that, when the gyro is at rest, the angular velocity is null and the bias can then be estimated by averaging. The underlying basis for the implementation of automatic resetting is that, for cyclical motions, the values of angular rotation would be identical at each instance of a specific event occurring in the gait cycle, for instance foot-flat (Sabatini, Martelloni, Scapellato, & Cavallo, 2005). Automatic nulling and resetting imply the capability of discriminating rest and dynamic conditions by algorithmic means—a precursor to more sophisticated methods for automatic recognition of activity, as discussed later in this section. Another problem of gyroscopic sensing is that gyros measure changes of orientation rather than absolute orientation. Gravimetric tilt sensing and geomagnetic compassing are then required (Kemp et al.,1998), however alignment would be performed only in the time intervals between successive flights of the involved body segment (Figure 4) (Sabatini, 2005).

performed.

There are a number of difficulties in developing ambulatory monitors (Kiani, Snijders, & Gelsema, 1997; Foerster & Fahrenberg, 2000). The problems concern the algorithms that have to be designed to provide a reliable detection and discrimination of motions patterns and posture and the optimal sensor configuration set (type, number, placement). Algorithms for the detection of postures and motion patterns are a crucial aspect of this approach. The research is still intense on methods to achieve adequate data reduction and discrimination power, in the face of the generally ample variety of strategies existing in performing a particular daily-life motor activity and the large inter- and intra-subject variability. Optimality means that the highest discrimination ability would be achieved for a spectrum of motor activities as broad as possible, while keeping sensor number at minimum so as to increase subject's compliance and usability, and to reduce overall system complexity and cost.

An example of a decision tree for discriminating a set of postures and mobility-related activities by an accelerometry-based system is sketched in Figure 5.

Each node of the tree has multiple branches leading to all of the movements of interest at the next level of the hierarchy. At each node, all the possible classifications are considered, including the fallback case, namely the case that is accepted after that all other classifications have been discarded, and the most likely candidate is retained. Decision trees of this kind are presented in several works (Bussmann et al.,1995; Veltink et al.,1996; Kiani et al.,1997; Mathie, Celler, Lovell, & Coster, 2004), based on the particular mode of operation of the ambulatory system (clinical assessment or event monitoring), the specific hardware configuration, and the strategy upon which the development of the pattern recognition system is based.

Each node of the tree must have a classification algorithm associated to it. Only two approaches have been used to a greater extent in the design of these classifiers. In fixed-threshold classification, motion patterns are detected and discriminated by applying a threshold to the signal of interest. The threshold setting is critical, since intra- and inter-individual variations could easily lead to wrong classification.

In reference-pattern-based classification, the detection of motion patterns could be improved by developing a template pattern for each postural and activity condition and cross-correlating the signal with this (Veltink et al.,1996; Foerster & Fahrenberg, 2000). The precise placement of sensors is essential when thresholds are used for the classification of motion.

A classification that is based on individual reference patterns appears to be less susceptible than threshold-based classification, although careful sensor placement is critical for achieving good test-retest reliability.

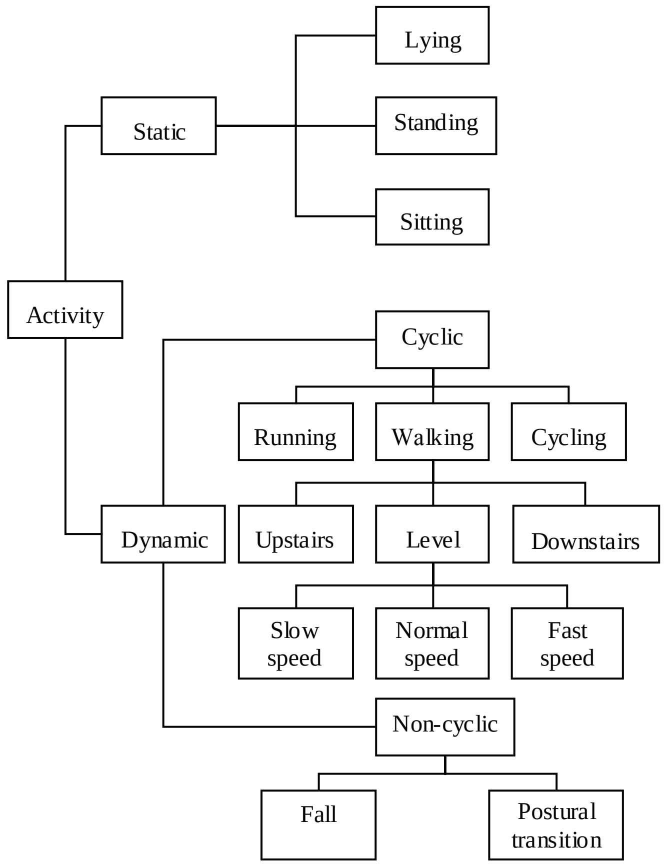
Posture detection (sitting, standing, lying and their subtypes) (see Figure 5) requires that the orientation of a number of body segments (trunk, thigh, and so forth) relative to the gravitational field is computed, which involves necessarily the creation of a suitable sensor configuration arrangement. Crucial to the success of posture detection is then the ability to discriminate between gravitational and motion-related components of the raw acceleration signal. With some abuse of terminology, these components are commonly referred to as DC and AC components, respectively. Underlying this definition is the approach of low-pass filtering the raw signals for their separation, which is usually achieved by threshold-based tests for significant deviations of acceleration norm from gravity (Veltink et al.,1996).

Postural transitions are usually detected by threshold-based methods, and classified according to the nature of the static activities which precede and follow them. The process is claimed to be more robust when motion patterns are analyzed in specific frequency bands via wavelet-based multi-

Table 1. Summary of a collection of relevant past and ongoing work on automatic recognition of activity using inertial sensors

Reference	Activity	Sensor number	Sensor type and placement	Subjects	Accuracy [%]
Aminian et al., 1999b	Dynamic/sitting/standing/lying	2	1D-acc: thigh, sternum	5	89.3
Bao and Intille, 2004	Activities of daily living	10	2D-acc: shank, thigh, hip upper leg, wrist	20	41-97
Bussmann et al., 1995	Sitting/standing/lying Walking level/upstairs/downstairs, cycling	6	2D-acc: shank, thigh 1D-acc: sternum, shoulder	5	N/A
Bussmann et al., 2004	Inactive Active: walking upstairs/downstairs, cycling, other movements Locomotion: walking level	3	2D-acc: shank 1D-acc: shank	12	98.5
Foerster et al., 1999	Sitting/standing/lying Walking level/upstairs/downstairs, cycling	4	1D-acc: shank, thigh sternum, wrist	24	95.8
Kiani et al., 1997	Sitting/standing/lying Walking level (at different speeds)	4	1D-acc: thigh(s) 2D-acc: sternum	11	98
Lee and Mase, 2002	Sitting/standing Walking level (at different speeds), walking upstairs/downstairs	4	(2D-acc, 1D-gyro): hip compass: waist	8	91.8
Mäntyjarvi et al., 2001	Walking level/upstairs/downstairs	6	3D-acc: hip(s)	6	83-90
Mathie et al., 2004	Sitting/standing/lying Walking level	3	3D-acc: lower back	26	97
Paraschiv-Ionescu et al., 2004	Sitting/standing/lying Walking level (with estimation of gait parameters)	5	(2D-acc, 1D-gyro): sternum 1D-gyro: thigh, shank	21	98
Najafi et al., 2003	Sitting/standing/lying Walking level	3	(2D-acc, 1D-gyro): sternum	15	> 90
Randell and Muller, 2000	Sitting/standing Walking level/upstairs/downstairs, running	2	2D-acc: hip	10	85-90
Uiterwal et al., 1998	Sitting/standing/lying Walking	3	3D-acc: waist	1	86-93
Van Laerhoven and Cakmacki, 2000	Sitting/standing Walking level/upstairs/downstairs, running, cycling	4	2D-acc: knee(s)	1	45-96
Veltink et al., 1996	Sitting/standing Walking level (at different speeds) Walking upstairs/downstairs , cycling (at different speeds)	3	2D-acc: sternum 1D-acc: thigh	10	83

Figure 5. Decision tree showing relationships between different movements of interest in a generic ambulatory monitor



With a few exceptions, the behaviors most ambulatory monitors attempt to discriminate are standing, sitting, lying, postural transitions and walking. Standing, sitting, lying are instances of static activities, subtypes of which may be: "standing in stooped position, in upright position, lateral bending to the left, lateral bending to the right" (Kiani et al. 1997), "sitting upright or flexed," "standing upright or flexed" (Paraschiv-Ionescu et al., 2004), "standing lying on the left, on the right, supine or with back supported" (Foerster & Fahrenberg, 2000). Motion patterns may be non-cyclical (postural transitions) or cyclical (walking). Postural transitions such as "sit-to-stand, stand-to-sit, upright-to-lying, lying-to-upright" are included in the decision tree, among others, by Mathie et al. (2004) and Paraschiv-Ionescu et al. (2004). Walking is the most common example of cyclic dynamic activity, subtypes of which are "level walking at slow, normal, fast speed" (Kiani et al., 1997) and "walking upstairs, downstairs" (Bussmann et al. 1995; Foerster & Fahrenberg, 2000). Other dynamic activities may include cycling (Bussmann et al., 1995).

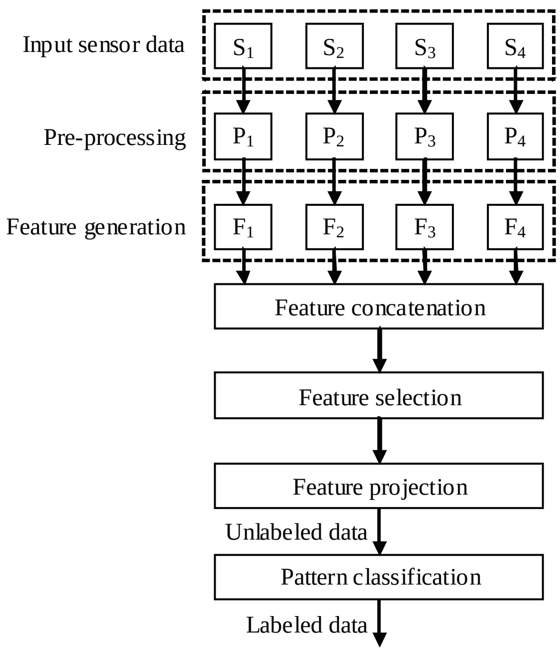
resolution analysis, which is also a better approach to deal with the discrimination between AC and DC components (Paraschiv-Ionescu, 2004).

Dynamic activity classification is based upon the observation that acceleration signals per cycle may differ in several aspects: morphology, statistical moments, and so forth, which calls for a suitable strategy upon which the development of the pattern recognition systems would be based.

The statistical approach to pattern recognition represents a pattern by features, viewed as points in suitably high-dimensional spaces (Figure 6). Machine-learning techniques are needed to turn raw data into useful contextual information, that is, information useful to provide knowledge about the state of the subject. To maximize the pattern recognition ability, these techniques would be able to generate and select data representations (features) with high discrimination ability. To perform classification, training patterns, captured from a large set of sampled movements and a range of conditions, can be used to determine the decision boundaries. There appears to be general consensus that a number of challenges exist in this regard. Different activities do not undergo abrupt changes, which makes their detection difficult. The variability in performing these activities by humans is generally high. The amount of information we can get from simple sensors is limited and noisy. Increasing the number of sensors, not necessarily inertial sensors, to achieve higher redundancy and robustness is not without practical (wearability, computational bottlenecks) or theoretical limits (curse of dimensionality, incurred by any machine learning approach we may favor). Numerous are the contact points between automatic recognition of activity as stated above and the research on methods for augmenting a computer with sensors which can make it aware of the context, as carried out in robotics and artificial intelligence.

Albeit identification of context without requiring external supervision seems to be better suited to become aware of context (Krause, Siewiorek, Smailagic, & Farringdon, 2003), most current approaches to the problem of automatic recogni-

Figure 6. The general structure of a pattern recognition system



tion of activity are based on using supervised machine learning techniques. The choice of good features is a fundamental step in statistical pattern recognition and a highly problem dependent task. A systematic approach to the search for features usable for analyzing daily-life motor activities is illustrated in Kiani et al. (1997), where simple statistics such as cumulative sums, means, standard deviations, RMS values computed over sliding windows with selectable size and overlap are shown to be generally quite effective. Other possible choices include frequency-domain features (energy distribution, entropy, and so forth), and features that specifically measure correlations among signals from different sites and measurement directions—the latter ones are especially useful in the attempt to improve recognition of activities involving movements of multiple body parts (Bao & Intille, 2004).

The design of the clustering layer reflects different choices and approaches to the problem of feature selection and projection. Feature selection, either driven by an expert or by step-wise methods of selection, is an important step to deal properly with the issue of curse of dimensionality. Sometimes, the generation of additional features is costly (more sensors are needed), and their actual use may even be harmful to achieve optimality (Loosli, Canu, & Rakotomamonjy, 2003; Bao & Intille, 2004).

The ANN approach is perceived as a good tool to cluster noisy data from several sensors (Kiani et al.,1997; Van Laerhoven, Aidoo, & Lowette, 2001; Mäntyjärvi, Himberg & Seppänen, 2001). ANNs are used in (Mäntyjärvi et al.,2001). After raw accelerometer data are submitted to wavelet analysis for computation of the statistical power at selected scales of analysis, the principal component analysis (PCA) and the independent component analysis (ICA) are used for dimensionality reduction. Clustering is based on using a multilayer perceptron with back-propagation learning. The rationale for using PCA and ICA is to find directions and scales where the sensor signals would be more informative for discrimination purposes. Van Laerhoven et al. (2001) discuss in depth advantages and disadvantages of using ANNs for clustering, in particular Kohonen self-organizing maps (KSOM). Their strong point is that activity contributes highly to context awareness, and adaptivity is indispensable in context awareness. Indisputably, the main disadvantage of many machine-learning algorithms is in their limited capability to properly handle the stability-plasticity dilemma, in the face of highly dynamic contexts (Van Laerhoven & Cakmakci, 2000). To overcome the limitations of KSOM in keeping up with the learning process over time, these authors consider the addition of a k-nearest neighbors clustering cascaded to the KSOM as a feasible way to stabilize its behavior.

Activity recognition on the selected features can be performed using any of different algorithms, including rule-based activity recognition, standard k-nearest neighbors search, naïve Bayes classifiers, and support vector machines (SVMs). Rule-based activity recognition seems to be good in capturing correlations between feature values that may lead to good recognition accuracy (Bao & Intille, 2004). Standard k-nearest neighbors search is shown to be a good work-horse in general problems of activity recognition (Van Laerhoven & Cakmakci, 2000; Bao & Intille, 2004). The problem with a naïve Bayes approach may be the excessive reliance on assumptions of conditional independence between different features in the Bayesian update (Golding & Lesh, 1999), and its greediness of training data to accurately model feature value distributions (Bao & Intille, 2004). In an attempt to overcome some limitations of neural network-based modeling, SVMs are increasingly recognized for their good performance on difficult classification tasks. Despite their emergence as a valuable tool for several applications in the biomedical field, including gait analysis (Begg & Kamruzzaman, 2005), only Loosli et al. (2003), to the best of our knowledge, have used them in the attempt to perform automatic recognition of activity. These authors claim that their implementation of SVMs is largely superior to the algorithms described in Van Laerhoven and Cakmakci (2000) and Van Laerhoven et al. (2001) over exactly the same datasets and selected features, although this claim has to be tempered by other findings of theirs, which show how SVMs are substantially equivalent to k-nearest neighbors clustering directly applied to the statistical features elaborated in Van Laerhoven and Cakmakci (2000).

An important step to improve upon the design of a pattern recognition system for analysis of daily-life motor activities is the addition of a supervision layer to introduce memory, so that individual movements are not classified in isolation (Mathie et al.,2004). Beside the classical approach to have a rule-based system to check sequences of classified movements and to correct those sequences which would be impossible, other approaches can be based on methods such as Markov modeling (Van Laerhoven & Cakmakci, 2000). The supervision layer on top of the classifier can supervise transitions from one context to another. A probabilistic finite state machine architecture is used where each context is represented by a state, and transitions are represented by arcs between states. A probability measure is built for each change of contexts, so every time a transition occurs, the supervision model checks if this is really likely (update of the probability). Observed state transitions that have a low probability to occur can thus be suspected to come from erroneous detections performed by the classification algorithms associated to the corresponding nodes of the decision tree.

Regarding a complex behavior as the juxtaposition of a sequence of elementary movements leads quite naturally to consider the hidden Markov model (HMM) analytical framework as a promising avenue of research in the field of automatic recognition of activity. An HMM is a doubly stochastic process with an underlying process of transitions between hidden states of the system and a process of emitting observable outputs, and, as such, it is ideal to implicitly model all of the many sources of spatio-temporal variability inherent in real movements. So far, HMMs have been used for representation and recognition of speech, handwriting and, of particular importance in the present context, gestures. Mäntyla, Mäntyjärvi, Seppänen, and Tuulari (2001) attempt to use hand gesture recognition, using cybergloves or accelerometer-based systems studied as a method of computer input for people with severe speech and motor impairment. Kallio, Kela and Mäntyjärvi (2003) develop a small wireless gesture-based input device based on accelerometers to provide methods to interact with different kinds of devices and environments. In both cases, trainable pattern recognition methods, namely HMMs, clustering and vector quantization are used to develop models for describing prototype gestures and to examine the gesture recognition performance. If we refer to gesture as a specific, intentional action by a human in which part of the body is moved in a predefined way indicating a specific event (Chambers, Venkatesh, West, & Bui, 2002), the way is paved to consider HMM-based gesture recognition a possible means to provide context knowledge, useful, in particular within occupational settings, not only to improve on standard clinical decision-making procedures (Uiterwal, Glerum, Busser, & van Lummel, 1998), but also to reduce the cognitive load on a worker during complex real world tasks (Lukowicz et al.,2004).

A final point is important to discuss. At first sight, it seems not difficult to achieve high values of accuracy, sensitivity and specificity in recognizing a relatively small set of postures and activities, although, sometimes, specific activities are resistant to robust classification and are frequently confused with one another, for instance level walking and upstairs walking (Veltink et al.,1996; Foerster & Fahrenberg, 2000). However, it should be pointed out that the highly specific system and methodologies used by each group make it difficult to directly compare different approaches. In most cases, these approaches are intended to meet the requirements of specific applications, and are tested in restricted environments and with small-size pools of subjects. Ironically, ambulatory systems intended to overcome the limitations of traditional laboratory-based systems for assessing motor performance are victims of their own validation being pursued, in the vast majority of cases, within laboratory environments which may artificially constrict and influence subject activity (Foerster, Smeja, & Fahrenberg, 1999). A thorough discussion about the need to use naturalistic data in training and testing ambulatory monitors is in Bao and Intille (2004). Needless to say, this need to cope with the idiosyncrasies of real-life worlds is precisely one of the driving factors behind the current interest in advanced methods for context-awareness and ambient intelligence among the artificial intelligence and robotic communities.

Functional Assessment of Specific Motor Disorders

Traditionally, accelerometers are applied to the quantitative analysis of tremor time series. The quantification of amplitude, frequency and occurrence time of tremor in patients affected with Parkinson's disease (PD) and its relation to posture and motion is described in Van Someren et al. (1998) and Foerster and Smeja (1999). The theory of spectral estimation is applied in Timmer, Lauk and Deuchl (1996) to assess whether a spectrum of accelerometric recordings exhibits multiple significant peaks and discuss different approaches to determine the amplitude and frequency of the tremor components from the spectrum. Another approach considers detection algorithms looking at tremor-related features in the time-domain, as in Van Someren et al. (1997), where the discrimination of tremor from intentional movements is based on the persistency of some features, such the temporal distances occurring between consecutive zero-crossings, the values of these periods of half waves, and the values of the peak-to-peak amplitudes in each half wave. Because of the time-varying nature of tremor-related components in accelerometric recordings, and the need for distinguishing the components that are due to voluntary movements and the components that are due to the movement disorder of interest, adaptive filtering techniques have also been considered, such as the Fourier Linear Combiner (Riviere, Rader, & Thakor, 1998) and the cascade learning architecture (Riviere & Khosla, 1997).

The assessment of postural stability is studied in Mayagoitia, Lötters, Veltink and Hermens (2002), where a tri-axis accelerometer-based system is developed for determining the long-term ability to maintain balance while standing. The acceleration measurement at the BCOM level allows one to introduce a number of performance parameters, similar to those that are extracted from force plates in classical posturographic studies. Two other approaches are interesting for their connection, in terms of the processing tools adopted, to the studies on assessment of physical activity (Cho & Kamen, 1998; Moe-Nilssen, 1998). More complex IMUs are those reported in Lee, Laprade and Fung (2001) and Wall and Weinberg (2003). Wall and Weinberg (2003) propose a balance prosthesis for postural control. In order to prevent falls in the balance impaired, body-tilt information is displayed to the subject via an array of tactile vibrators. The body-tilt information is obtained by integrating the information from a waist dual-axis accelerometer with the information from a single-axis gyro. Lee et al. (2001) deal with the problem of monitoring the lumbar spine motion, and to this aim they develop a portable system for three-dimensional motion analysis. A recent contribution towards the use of inertial sensing for the monitoring of specific motor disorders during postural transitions is illustrated in the work by Najafi, Aminian, Loew, Blanc and Robert (2002). They analyze the trunk tilt, corresponding to the angle between the vertical axis and anterior wall of the subject's thorax, and the trunk vertical displacement during stand-sit transfers, in the attempt to introduce good indicators of the risk of falls sustained by elderly people.

As for the assessment of gait, Smidt et al. (1971) apply Fourier series analysis to define a measure of smoothness of walking, which is found to provide an effective method for discriminating between normal gait patterns and gait patterns of subjects with gait defects. The approach by Sekine, Tamura, Togawa and Fukui (2000) and Sekine et al. (2002) is interesting for the connection established between wavelet properties and characteristics of motor fluctuations in normal and pathological conditions. In Sekine et al. (2000), waist 3D-accelerations are submitted to wavelet analysis in order to discriminate different types of walking in healthy subjects. Also in Sekine et al. (2002), the approach is further refined so that the distribution of wavelet coefficient power over selected scales of analysis is subject to a complexity analysis for the determination of a fractal dimension (the Hurst exponent) in different groups—healthy, elderly, subjects with PD. The Hurst exponent is shown to be significantly different among three types of walking (level/upstairs/downstairs) for individual subjects and show a great reproducibility. Moreover, the fractal dimensions are higher for elderly and PD patients as compared with healthy subjects.

The study of motor fluctuations in PD patients is reviewed in Keijsers, Hornstink and Gielen (2003). Their goal, in particular, is to assess levodopa-induced dyskinesia in these subjects. Dyskinesia are involuntary movements — jerky and characterized by sudden contractions followed by stretching, twisting and rotation — that usually occur after several years of using levodopa, a drug which is widely used to treat the symptoms of PD. The ambulatory monitor developed by Keijsers et al. (2003) is composed of several tri-axis accelerometes (at both upper arms, at both upper legs, at the wrist of the most dyskinetic side, and at the top of the sternum). Due to the complex, time-varying relations between voluntary movements and motor disorders, a simple supervised ANN is the selected classification technique exercised over a number of time- and frequency-domain features. The features with the highest discriminative ability are survivors of a pruning phase, which is part of the design procedure. At the end, the system is successfully trained to detect and assess the severity of dyskinesia, provided that the score of the severity of the symptoms is given by an experienced clinician.

Personal Navigation

The ability to locate the position/orientation of a person is of great importance in a number of applications, including electronic travel aids for the blind or visually impaired, integrated navigation systems for the dismounted infantry soldier, and, in general, context-aware applications in wearable and mobile computing.

The location-sensing techniques for personal navigation systems are based on either relative or absolute position measurements. Traditionally, inertial sensors have been employed for implementing dead-reckoning techniques, a variant of relative position measurements whose fundamental idea is to integrate incremental motion information over time. The problem of absolute positioning in outdoor environments is seemingly straightforward to solve by an externally referenced sensing technology such as GPS. However, the disadvantages with GPS—inability to work in indoors, unavailability of satellite signals in environments such as urban canyons, poor accuracy in relation to the needs of a specific application, inability to provide static heading information—suggest that dead-reckoning based on inertial sensing and GPS would have a better chance to work in practice (Ladetto, van Seeters, Sokolowski, Sagan, & Merminod, 2002). In the case of indoor environments, one approach for overcoming dead-reckoning limitations is to acquire location information by adding intelligence to the environment, so that it can supply location information to users via special infrastructures (Want, Hopper, Falcão, & Gibbons, 1992; Krumm, Williams, & Smith, 2002)—applications of externally referenced sensing techniques. The traditional computational techniques used to improve the performance of dead-reckoning systems aim at implementing either map matching or in-line sensor calibration procedures based on KFs which use external aids, such as GPS, in the attempt to estimate bias drifts of inertial sensors and compass disturbances.

The key problem in the design of a personal navigation system is to find a method to measure length and direction of displacement using step time as the basic unit of time, so as to determine the distance and heading from a known origin at an acceptable level of accuracy. Detecting step occurrences can be based on accelerometers or gyros, as stated above with regard to the problem of assessing temporal parameters of gait. A simple model-based approach to the problem of estimating step length hypothesizes that, once a method is available to determine step time, step length estimation can be based on cadence. Judd (1997) suggests that the step length could be estimated online based on a linear relationship between measured cadence and step length, whose validity, discussed in Ladetto et al. (2002), is however limited to level walking in open spaces. The approach by Ladetto et al. (2002) is an interesting example of so-called biokinematic navigation (Elwell, 1999), an inexpensive technique that matches an individual's gait to inertial measurements. Ladetto et al. (2002) exploit the relationship existing between walking speed and statistical features, such as the RMS values of waist 3D-accelerations, under the premise that the main goal of locomotion is to promote the BCOM displacement in space. The fact that step length is a time-varying process, with large environment-dependent variations, motivates the use of additional sensors for online model calibration, for instance GPS as in Perrin et al. (2000), Ladetto et al. (2002) and Jirawimut, Ptasinski, Garaj, Cecelia and Balachandran (2003) or to incorporate additional features of human walking dynamics into the algorithm (Vildjiounaite, Malm, Kaartinen, & Alahuhta, 2002; Lee & Mase, 2002). These authors note that, when the step length is longer, the accelerations in different parts of the legs tend to be higher. Hence, it may be hypothesized that the greater the vertical acceleration during a step, the greater the distance that has been traveled across the ground. This relationship can be exploited to adjust the pre-calculated step length and hence reduce the error associated with fixed values by using sensor look-up tables (Vildjiounaite et al.,2002) or a simple fuzzy-logic reasoning method (Lee & Mase, 2002).

Determining heading may require the implementation of gyro-compassing techniques. After INS alignment is performed, heading is estimated by gyro and magnetic compass data, provided both data are available. In contrast with a gyro, magnetic compass is not prone to drift, however its use is critical. Sometimes, serious inaccuracies in the magnetic compass readings are found which cannot be traced back to imperfect tilt compensation, but are likely due to perturbations in the magnetic field near or within the measurement space. The presence of magnetic disturbance can be checked by comparing the rate of change of the heading estimates from gyro and magnetic compass (Ladetto & Merminod, 2000). The conflicts in data interpretation existing about the existence and extent of turns indicated by gyro and magnetic compass warn not to consider magnetic compass reliable. Otherwise, its long-term low-frequency response can be combined with the high-frequency response of gyros. The use of GPS to dynamically recalibrate the heading error is appropriate in this context. In theory this combination should give the most accurate results. In particular, computed GPS heading can allow one to model the bias of the magnetic compass (Jirawimut et al.,2003) and the gyro bias and drift (Ladetto et al.,2002). In indoor applications, where GPS is unavailable, the performance of biokinematic navigation algorithms would be improved by map matching, as pointed out in Lee and Mase (2002) and Vildjiounaite et al. (2002). These authors show that the task of determining the person's location when the navigation space is constrained, such as within a building, can be greatly simplified by having some form of environment knowledge available to the navigation algorithm. This knowledge can be limited to indicate where some locations, or transition between locations, are in the map, such as the transition from one room to the next, the transition from level to upstairs, the presence of a corridor. When this knowledge is combined with an even relatively crude heading information, irregular walking styles or strong perturbations in the magnetic field can be accommodated and reasonably accurate estimates of person's location can still be obtained.

In our view, the important point is that there is something peculiar to the way humans move which makes personal navigation so different from vehicular navigation. Walking styles of humans imply frequent changes of speed and orientation, for reasons of stability, safety, and so forth, depending on the environment the activity is actually performed—the context. Because of this complexity, the prediction of the walked path from one point to another turns out to be de facto impossible. Moreover, inertial/magnetic sensors record movements depending on their actual placement on the body. Four examples point at the difficulties inherent with their use: changes of orientation that occur with minimal displacement changes (on-the-spot turning); changes of orientation not recorded, in spite that they occur over large angles (backward displacements, side-stepping); instantaneous changes of orientation that are recorded, in spite that the actual direction of displacement does not change (lateral bending of the trunk while walking straight ahead); steps which are not taken with normal rhythm or style, hence requiring special algorithms for their detection (side-stepping). Of course, these examples are not comprehensive of all situations taking place in practice. However, they point at the need that personal navigation systems would be capable of classifying "movements," and, for each classified movement, of reliably estimating length and direction of displacement in real-time conditions.

EMERGING TRENDS

In this chapter, the main computational approaches to use motion sensing based on inertial and magnetic sensors have been reviewed, in connection with a wealth of biomedical applications, including assessment of motor performance, automatic recognition of activity, and personal navigation.

Of utmost importance is the point that the answers to the problems raised by each application ought not to be considered in isolation, because of the relevant overlapping they have with each other:

1. Advanced signal processing and machine-learning techniques are used in the effort to identify signatures of human motion and perform motor pattern recognition;
2. Constraints in navigation space and dynamic features of unrestrained human motion concur to provide accurate long-term solutions to the navigation problem;
3. Topological information merged with an estimate of the traveled distance by dead-reckoning methods leads to improve the accuracy of methods for automatic recognition of activity.

The most promising avenue of research is in how artificial intelligence and biomechanics can reinforce each other, in the direction to pave a new way of inertial sensing that will not be hampered by the idiosyncrasies of current-generation sensing hardware. Impressive technological advances are about to turn wireless, wearable distributed multi-sensor systems into reality and further advances are also expected in the field of MEMS technologies. However, the capability of performing a robust, accurate real-time reconstruction of the trajectory of selected anatomical points, such as, for instance, the sacrum, the stern, the foot instep, by inertial and magnetic sensors is still the prerequisite for making body personal area networks attractive for long-term ambulatory monitoring of human subjects engaged in functional activities involving motion in unrestrained conditions. Interesting research is yet to be performed in regard to the signal processing algorithms the practitioners in the field will have to create in order to successfully cope with the difficult behavior of the motion sensors discussed in this chapter.

REFERENCES

Abbott, E., & Powell, D. (1999). Land-vehicle navigation using GPS. In *Proceedings of the IEEE, 87*(1), 145-162.

Aminian, K., Rezakhanlou, De Andres, Fritsch, Leyvraz, P.-F., & Robert, Ph. (1999). Temporal feature estimation during walking using miniature accelerometers: An analysis of gait improvement after hip arthroplasty. *Medical & Biological Engineering & Computing, 37*(6), 686-691.

Aminian, K., Robert, Ph., Jéquier, E., & Schutz, Y. (1995). Incline, speed, and distance assessment during unconstrained walking. *Medicine & Science in Sports & Exercise, 27*(1), 226-234.

Aminian, K., Robert Ph., Buchser, E. E., Rutschmann, B., Hayoz, D., & Depairon, M. (1999). Physical activity monitoring based on accelerometry: validation and comparison with video observation. *Medical & Biological Engineering & Computing, 37* (3), 304-308.

Aminian, K., Najafi, B., Büla, C., Leyvraz, P.-F., & Robert, Ph. (2002). Spatio-temporal parameters of gait measured by an ambulatory system using miniature gyroscopes. *Journal of Biomechanics, 35*(5), 689-699.

Auvinet, B., Chaleil D., & Barrey, E. (1999). Accelerometric gait analysis for use in hospital outpatients. *Rev. Rhum. (Engl. Ed.), 66*(7-9), 389-397.

Bachmann E. R., Yun, X., McKinney, D., McGhee, R. B., & Zyda, M. J. (2003, September 14-19). Design and implementation of MARG sensors for 3-DOF orientation measurement of rigid bodies. In *Proceedings of the IEEE International Conference on Robotics and Automation*, Taipei, Taiwan, (pp. 1171-1178).

Bao, L., & Intille, S. S. (2004). Activity recognition from user-annotated acceleration data. In A. Ferscha & F. Mattern (Eds.), *Proceedings of Pervasive 2004, LNCS* (Vol. 3001, pp. 1-17). Berlin Heidelberg: Springer-Verlag.

Begg, R., & Kamruzzaman, J. (2005). A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data. *Journal of Biomechanics, 38*(3), 401-408.

Bortz, J. E. (1971). A new mathematical formulation for strapdwon inertial navigation. *IEEE Trans. on Aerospace and Electronic Systems, 7*(1), 61-66.

Bouten, C. V. C., Koekkoek, K. T. M., Verduin, M., Kodde, R., & Janssen, J. D. (1997). A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Trans. on Biomedical Engineering, 44*(3), 136-147.

Bouten, C. V. C., Sauren, A. A. H. J., Verduin, M., & Janssen, J. D. (1997). Effects of placement and orientation of body-fixed accelerometers on the assessment of energy expenditure during walking. *Medical & Biological Engineering &*

Computing, 35(1), 50-56.

Bussmann, J. B., Hartgerink, I., van der Woude, L. H., & Stam, H. J. (2000). Measuring physical strain during ambulation with accelerometry. *Medicine & Science in Sports & Exercise, 32*(2), 1462-1471.

Bussmann, J. B., Veltink, P. H., Koelma, F., van Lummel, R. C., & Stam, H. J. (1995). Ambulatory monitoring of mobility-related activities: The initial phase of the development of an activity monitor. *European Journal Physical Medicine Rehabilitation, 5*(1), 2-7.

Caruso, M. J. (1997). Applications of magneto-resistive sensors in navigation systems. *Sensor and Actuators, 1220*, 15-21.

Chambers, G. S., Venkatesh, S., West, G. A. W., & Bui, H. H. (2002). Segmentation of intentional human gestures for sports video annotation. In *Proceedings of the International Conference Pattern Recognition* (pp. 1082-1085).

Chen, K. Y., & Sun, M. (1997). Improving energy expenditure estimation by using a triaxial accelerometer. *Journal of Applied Physiology, 83*(6), 2112-2122.

Cho, C. Y., & Kamen, G. (1998). Detecting balance deficits in frequent fallers using clinical and quantitative evaluation tools. *Journal of the Americal Geriatrics Society, 46*(4), 426-430.

Curey, R. K., Ash, M. E., Thielman, L. O., & Barker, C. H. (2004, April 26-29). Proposed IEEE inertial systems terminology standard and other inertial sensor standards. *IEEE Position Location and Navigation Symposium, PLANS 2004*, Monterey, CA, (pp. 83-90).

Elwell, J. (1999). Inertial navigation for the urban warrior. In *Proceedings of the SPIE International Society of Optical Engineering* (Vol. 3709, pp. 196-204).

Evans, A. L., Duncan, G., & Gilchrist, W. (1991). Recording accelerations in body movements. *Medical & Biological Engineering & Computing, 29*(1), 102-104.

Foerster, F., & Smeja, M. (1999). Joint amplitude and frequency analysis of tremor activity. *Electromyography & Clinical Neurophysiology, 39*(1), 11-19.

Foerster, F., Smeja, M., & Fahrenberg, J. (1999). Detection of posture and motion by accelerometry: A validation in ambulatory monitoring. *Computers in Human Behavior, 15*, 571-583.

Foerster, F., & Fahrenberg, J. (2000). Motion patterns and posture: correctly assessed by calibrated accelerometers. *Behavior Research Methods, Instruments, & Computers, 32*(3), 450-457.

Foxlin, E. (2002). Motion tracking requirements and technologies. In K. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications.* Mahwah, NJ: Lawrence Erlbaum.

Gage, H. (1964). *Accelerographic analysis of human gait* (Paper No. 64-WA/HUF 8). Washington, DC: American Society for Mechanical Engineers.

Gebre-Egziabher, D., Elkaim, G. H., Powell, J. D., & Parkinson, B. W. (2000, March 13-16). A gyro-free quaternion-based attitude determination system suitable for implementation using low cost sensors. In *Proceedings of the IEEE Position, Location and Navigation Symposium PLANS 2000*, San Diego, CA, (pp. 185-192).

Giansanti, D., Macellari, V., Maccioni, G., & Cappozzo, A. (2003). Is it feasible to reconstruct body segment 3D position and orientation using accelerometric data? *IEEE Trans. on Biomedical Engineering, 50*(4), 476-483.

Golding, A., & Lesh, N. B. (1999, October 18-19). Indoor navigation using a diverse set of cheap, wearable sensors. In *Proceeding of the3rd International Symposium on Wearable Computers,* San Francisico (pp. 29-36).

Ismail, A. R., & Asfour, S. S. (1999). Discrete wavelet transform: A tool in smoothing kinematic data. *Journal of Biomechanics, 32*(3), 360-368.

Kallio, S., Kela, J., & Mäntyjärvi, J. (2003, October 5-8). Online gesture recognition system for mobile interaction. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Washington, DC (Vol. 3, pp. 2070-2076).

Keijsers, N. L. W., Horstink, M. W. I. M., & Gielen, S. C. A. M. (2003). Online monitoring of dyskinesia in patients with Parkinson's disease. *IEEE Engineering in Medicine and Biology Magazine, 22*(3), 96-103.

Kemp, B., Janssen, A. J. M. W., & van der Kamp, B. (1998). Body position can be monitored in 3D using accelerometers and earth-magnetic field sensors. *Electroencephalography and Clinical Neurophysiology, 109*(6), 484-488.

Kiani, K., Snijders, C. J., & Gelsema, E. S. (1997). Computerized analysis of daily life motor activity for ambulatory monitoring. *Technology & Health Care, 5*(4), 307-318.

Krause, A., Siewiorek, D. P., Smailagic, A., & Farringdon, J. (2003, October 21-23). Unsupervised, dynamic identification of physiological and activity context in wearable computing. In *Proceedings of the 7th IEEE International Symposium Wearable Computers,* New York (pp. 88-97).

Krumm, J., Williams, L., & Smith, G. (2002). SmartMoveX on a graph - An inexpensive active badge tracker. In G. Borriello & L. E. Holmquist (Eds.), *UbiComp 2002, LNCS* (Vol. 2498, pp. 299-307). Berlin Heidelberg: Springer-Verlag.

Jirawimut, R., Ptasinski, P., Garaj, V., Cecelia, F., & Balachandran, W. (2003). A method for dead reckoning parameter correction in pedestrian navigation system. *IEEE Trans. on Instrumentation and Measurement, 52*(1), 209-215.

Judd, C. T. (1997, September). A personal dead reckoning module. In *Proceedings of the ION GPS '97,* Kansas City, MO, (pp. 47-51).

Ladetto, Q., & Merminod, B. (2002). In step with INS. *GPS World, 10*(10), 30-38.

Ladetto, Q., van Seeters, J., Sokolowski, S., Sagan, Z., & Merminod, B. (2002, October 14-16). Digital magnetic compass and gyroscope for dismounted soldier position and navigation. *NATO-RTO Meetings*, Istanbul, Turkey.

Lee, R. Y. W., Laprade, J., & Fung, E. H. K. (2003). A real-time gyroscopic system for three-dimensional measurement of lumbar spine motion. *Medical Engineering & Physics, 25*(10), 817-824.

Lee, S.-W., & Mase, K. (2002). Activity and location recognition using wearable sensors. *Pervasive Computing, 1*(3), 10-18.

Loosli, G., Canu, S., & Rakotomamonjy, A. (2003). *Détection des activités quotidiennes à l'aide des séparateurs à Vaste Marge*. RJCIA, France.

Lötters, J. C., Schipper, J., Veltink, P. H., Olthius, W., & Bergveld, P. (1998). Procedure for in-use calibration of triaxial accelerometers in medical applications. *Sensors and Actuators A, 68*(1-3), 221-228.

Luinge, H. J., & Veltink, P. H. (2004). Inclination measurement of human movement using a 3-D accelerometer with autocalibration. *IEEE Trans. on Neural Systems and Rehabilitation Engineering, 12*(1), 112-121.

Lukowicz, P., Ward, J. A., Junker, H., Stäger, M., Tröster, G., Atrash, A., & Starner, T. (2004).

Recognizing workshop activity using body worn microphones and accelerometers. In A. Ferscha & F. Mattern (Eds.), *Pervasive 2004, LNCS* (Vol. 3001, pp. 18-32). Berlin Heidelberg: Springer-Verlag.

Maki, B. E. (1997). Gait changes in older adults: Predictors of falls or indicators of fear? *Journal of the American Geriatrics Society, 45*(3), 313-320.

Mayagoitia, R. E., Lötters, J. C., Veltink, P. H., & Hermens, H. (2002). Standing balance evaluation using a triaxial accelerometer. *Gait & Posture, 16*(1), 55-59.

Mayagoitia, R. E., Nene, A. V., & Veltink, P. H. (2002). Accelerometer and rate gyroscope measurement of kinematics: An inexpensive alternative to optical motion analysis systems. *Journal of Biomechanics, 35*(4), 537-542.

Mäntyla, V.-M., Mäntyjärvi, J., Seppänen, T., & Tuulari, E. (2000). Hand gesture recognition of a mobile device user. In *Proceedings of the International IEEE Conference on Multimedia and Expo*, New York (pp. 281-284).

Mäntyjärvi, J., Himberg, J., & Seppänen, T. (2001, October 7-10). Recognizing human motion with multiple acceleration sensors. 747-752. In *Proceedings of the IEEE International Conference Systems, Man and Cybernetics, 2*, Tucson, AZ, (pp. 747-752).

Mathie, M. J., Celler, B. G., Lovell, N. H., & Coster, A. C. F. (2003). Classification of basic daily movements using triaxial accelerometer. *Medical & Biological Engineering & Computing, 42*(5), 679-687.

Meijer, G. A. L., Westerkerp, K. R., Verhoeven, F. M. H., Koper, H. B. M., & ten Hoor, F. (1991). Methods to assess physical activity with special reference to motion sensors and accelerometers. *IEEE Trans. on Biomedical Engineering, 38*(3), 221-229.

Meyer, K., Applewhite, H. L., & Biocca, F. A. (1992). A survey of position trackers. *Presence: Teleoperators and virtual environments, 1*(1), 173-200.

Miyazaki, S. (1997). Long-term unrestrained measurement of stride length and walking velocity utilizing a piezoelectric gyroscope. *IEEE Trans. on Biomedical Engineering, 44*(8), 753-759.

Moe-Nilssen, R. (1998). Test-retest reliability of trunk accelerometry during standing and walking. *Archives Physical Medicine Rehabilitation, 79*(11), 1377-1385.

Morris, J. R. W. (1973). Accelerometry: A technique for the measurement of human body movements. *Journal of Biomechanics, 6*(6), 729-736.

Najafi, B., Aminian, K., Loew, F., Blanc, Y., & Robert, Ph. (2002). Measurement of stand-sit and sit-stand transitions using a miniature gyroscope and its application in fall risk evaluation in the elderly. *IEEE Trans. on Biomedical Engineering, 49*(8), 843-851.

Najafi, B., Aminian, K., Paraschiv-Ionescu, A., Loew, F., Büla, C. J., & Robert, Ph. (2003). Ambulatory system for human motion analysis using a kinematic sensor: monitoring of daily physical activity in the elderly. *IEEE Trans. on Biomedical Engineering, 50*(6), 711-723.

Padgaonkar, A. J., Krieger, K. W., & King, A. I. (1975). Measurement of angular acceleration of a rigid body using linear accelerometers. *ASME Journal of Applied Mechanics, 42*(3), 552-556.

Pappas, I. P. I., Popovic, M. R., Keller, T., Dietz, V., & Morari, M. (2001). A reliable gait phase detection system. *IEEE Trans. on Neural Systems and Rehabilitation Engineering, 9*(2), 113-125.

Paraschiv-Ionescu, A., Buchser, E. E., Rutschmann, B., Najafi, B., & Aminian, K. (2004). Ambulatory system for the quantitative and qualitativa analysis of gait and posture in chronic pain patients treated with spinal cord stimulation. *Gait & Posture, 20*(2), 113-125.

Perrin, O., Terrier, P., Ladetto, Q., Merminod, B., & Schutz, Y. (2000). Improvement of walking speed prediction by accelerometry and altimetry, validated by satellite positioning. *Medical & Biological Engineering & Computing, 38*(2), 164-168.

Randell, C., & Muller, H. (2000, October 31-November 3). Context awareness by analyzing accelerometer data. In B. MacIntyre & B. Iannucci (Eds.), *The Fourth International Symposium on Wearable Computers,* Arlington, VA (pp. 175-176).

Riviere, C. N., & Khosla, P. K. (1997, April 20-25). Augmenting the human-machine interface: improving manual accuracy. In *Proceedings of the IEEE International Conference on Robotics and Automation*, Albuquerque, NM (Vol. 4, pp. 2346-2350).

Riviere, C. N., Rader, R. S., & Thakor, N. V. (1998). Adaptive canceling of physiological tremor for improved precision in microsurgery. *IEEE Trans. Biomedical Engineering, 45*(7), 839-846.

Sabatini, A. M. (2005). Quaternion based strap-down integration method for applications of inertial sensing to gait analysis. *Medical & Biological Engineering & Computing, 43*(1), 94-101.

Sabatini, A. M., Martelloni, C., Scapellato, S., & Cavallo, F. (2004). Energy expenditure rate in level and uphill treadmill walking determined from empirical models and foot inertial sensing data. *Electronics Letters, 40*(2), 95-96.

Sabatini, A. M., Martelloni, C., Scapellato, S., & Cavallo, F. (2005). Assessment of walking features from foot inertial sensing. *IEEE Trans. on Biomedical Engineering, 52*(3), 486-494.

Sagawa, K., Sato, Y., & Innoka, H. (2000, October 8-11). Non-restricted measurement of walking distance. In *Proceedings of the IEEE International Conference Systems, Man, and Cybernetics*, Nashville, TN (Vol. 3, pp. 1847-1852).

Sekine, M., Tamura, T., Akay, M., Fujimoto, T., Togawa, T., & Fukui, Y. (2002). Discrimination of walking patterns using wavelet-based fractal analysis. *IEEE Trans. on Neural Systems and Rehabilitation Engineering, 10*(3), 188-196.

Sekine, M., Tamura, T., Togawa, T., & Fukui, Y. (2000). Classification of waist-acceleration signals in a continuous walking record. *Medical Engineering & Physics, 22*(4), 285-291.

Smeja, M., Foerster, F., Fuchs, G., Emmans, D., Hornig, A., & Fahrenberg, J. (1999). 24 hr assessment of tremor activity and posture in Parkinson's disease by multichannel accelerometry. *Journal of Psychophysiology, 13*, 245-256.

Smidt, G. L., Arora, J.S., & Johnston, R. C. (1971). Accelerographic analysis of several types of walking. *American Journal of Physical Medicine, 50*(6), 285-300.

Terrier, P., Aminian, K., & Schutz, Y. (2001). Can accelerometry accurately predict the energy cost of uphill/downhill walking? *Ergonomics, 44*(1), 48-62.

Terrier, P., Ladetto, Q., Merminod, B., & Schutz, Y. (2000). High-precision satellite positioning system as a new tool to study the biomechanics of human locomotion. *Journal of Biomechanics, 33*(12), 1717-1722.

Timmer, J., Lauk, M., & Deuschl, G. (1996) Quantitative analysis of tremor time series. *Electroencephalography and Clinical Neurophysiology, 101*(5), 461-468.

Tong, K., & Granat, M. H. (1999). A practical gait analysis system using gyroscopes. *Medical Engineering & Physics, 21*(2), 87-94.

Uiterwal, M., Glerum, E. B. C., Busser, H.J., & van Lummel, R. C. (1998). Ambulatory monitoring of physical activity in working situations, a validation study. *Journal of Medical Engineering and Technology, 22*, 168-172.

Van Laerhoven, K., Aidoo, K., & Lowette, S. (2001). Real-time analysis of data from many sensors with neural networks. In *Proceedings of the 5th International Symposium on Wearable Computers*, Zurich, Switzerland (pp. 115-123).

Van Laerhoven, K., & Cakmakci, O. (2000). What shall we teach our pants? In B. MacIntyre & B. Iannucci (Eds.), *The Fourth International Symposium on Wearable Computers* (pp. 77-83).

Van Someren, E. J., Vonk, B. F., Thussen, W. A., Speelman, J. D., Schuurman, P. R., Mirmiran, M., & Swaab, D. F. (1998). A new actigraph for long-term registration of the duration and intensity of tremor and movement. IEEE Trans. on Biomedical Engineering, 45(3), 386-395.

Veltink, P. H., Bussmann, H. B. J., de Vries, W., Martens, W. L. J., & van Lummel, R. C. (1996). Detection of static and dynamic activities using uniaxial accelerometers. IEEE Trans. on Rehabilitation Engineering, 4(4), 375-385.

Veltink, P. H., Slycke, P., Hemssems, J., Buschman, R., Bulstra, G., & Hermens, H. (2003). Three dimensional inertial sensing of foot movements for automatic tuning of a two-channel implantable drop-foot stimulator. Medical Engineering & Physics, 25(1), 21-28.

Verplaetse, C. (1996). Inertial proprioceptive devices: self-motion-sensing toys and tools. IBM Systems Journal, 35, 639-650.

Vildjiounaite E., Malm, E.-J., Kaartinen, J., & Alahuhta, P. (2002). Location estimation indoors by means of small computing power devices, accelerometers, magnetic sensors and map knowledge. In F. Mattern & M. Naghshineh (Eds.), Pervasive 2002, LNCS (Vol. 2414, p. 211-224). Berlin Heidelberg: Springer-Verlag.

Wall C., & Weinberg, M.S. (2003). Balance prostheses for postural control. IEEE Engineering in Medicine and Biology Magazine, 2, 84-90.

Want, R., Hopper, A., Falcao, V., & Gibbons, J. (1992). The active badge location system. ACM Transactions on Information Systems, 10, 91-102.

Weir, R. F. ff., & Childress, D. S. (1997). A portable, low-cost, real-time, clinical gait analysis system. IEEE Trans. on Rehabilitation Engineering, 5(4), 310-321.

Wertz, J. R. (Ed.). (1984). Spacecraft attitude determination and control. Dordrecht, The Netherlands: D. Reidel.

Willemsen, A. Th. M., Bloemhof, F., & Boom, H. B. K. (1990). Automatic stance-swing phase detection from accelerometer data for peroneal nerve stimulation. IEEE Trans. on Biomedical Engineering, 37(12), 1201-1208.

Willemsen, A. Th. M., Frigo C., & Boom, H. B. K. (1991). Lower extremity angle measurement with accelerometers - Error and sensitivity analysis. IEEE Trans. on Biomedical Engineering, 38(12), 1186-1193.

Willemsen, A. Th. M., van Alsté, J. A., & Boom, H. B. K. (1990). Real-time gait assessment utilizing a new way of accelerometry. Journal of Biomechanics, 23(8), 859-863.

Williamson, R., & Andrews, B. J. (2001). Detecting absolute human knee angle and angular velocity using accelerometers and rate gyroscopes. Medical & Biological Engineering & Computing, 39(3), 294-302.

You, S., Neumann, U., & Azuma, R. (1999). Orientation tracking for outdoor augmented reality registration. IEEE Computer Graphics and Applications, 6(6), 36-42.

Zijlstra, W., & Hof, A. L. (2003). Assessment of spatio-temporal parameters from trunk acceleration during human walking. Gait & Posture, 18(2), 1-10.

This work was previously published in Computational Intelligence for Movement Sciences: Neural Networks and Other Emerging Techniques, edited by R. Begg; M. Palaniswami, pp. 70-100, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 3.20
Immersive Image Mining in Cardiology

Xiaoqiang Liu
Delft University of Technology, The Netherlands
Donghua University, China

Henk Koppelaar
Delft University of Technology, The Netherlands

Ronald Hamers
Erasmus Medical Thorax Center, The Netherlands

Nico Bruining
Erasmus Medical Thorax Center, The Netherlands

INTRODUCTION

Buried within the human body, the heart prohibits direct inspection, so most knowledge about heart failure is obtained by autopsy (in hindsight). Live immersive inspection within the human heart requires advanced data acquisition, image mining and virtual reality techniques. Computational sciences are being exploited as means to investigate biomedical processes in cardiology.

IntraVascular UltraSound (IVUS) has become a clinical tool in recent several years. In this immersive data acquisition procedure, voluminous separated slice images are taken by a camera, which is pulled back in the coronary artery. Image mining deals with the extraction of implicit knowledge, image data relationships, or other patterns not explicitly stored in the image databases (Hsu, Lee, & Zhang, 2002). Human medical data are among the most rewarding and difficult of all biological data to mine and analyze, which has the uniqueness of heterogeneity and are privacy-sensitive (Cios & Moore, 2002). The goals of immersive IVUS image mining are providing medical quantitative measurements, qualitative assessment, and cardiac knowledge discovery to serve clinical needs on diagnostics, therapies, and safety level, cost and risk effectiveness etc.

BACKGROUND

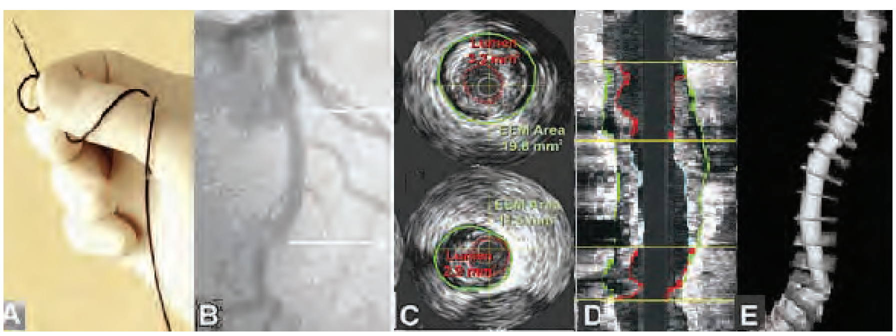
Heart disease is the leading cause of death in industrialized nations and is characterized by diverse cellular abnormalities associated with decreased ventricular function. At the onset of many forms of heart disease, cardiac hypertrophy and ventricular changes in wall thickness or chamber volume occur as a compensatory response to maintain cardiac output. These changes eventually lead to greater vascular resistance, chamber dilation, wall fibrosis, which ultimately impair the ability of the ventricles to pump blood and lead to overt failure. To diagnose the many possible anomalies and heart diseases is difficult because physicians can't literally see in the human heart. Various data acquisition techniques have been invented to partly remedy the lack of sight: non-invasive inspection including CT (Computered Tomography), Angiography, MRI (Magnetic Resonance Imaging), ECG signals etc. These techniques do not take into account crucial features of lesion physiology and vascular remodeling to really mine blood-plaque. IVUS, a minimal-invasive technique, in which a camera is pulled back inside the artery, and the resulting immersive tomographic images are used to remodel the vessel. This remodeling vessel and its virtual reality (VR) aspect offer interesting future alternatives for mining these data to unearth anomalies and diseases in the moving heart and coronary vessels at earlier stage. It also serves in clinical trials to evaluate results of novel interventional techniques, e.g. local kill by heating cancerous cells via an electrical current through a piezoelectric transducer as well as local nanotechnology pharmaceutical treatments. Figure 1 explains some aspects of IVUS technology.

However, IVUS images are more complicated than medical data in general since they suffer from some artifacts during immersed data acquisition (Mintz et al., 2001):

- Non-uniform rotational distortion and motion artifacts.
- Ring-down, blood speckle, and near field artifacts.
- Obliquity, eccentricity, and problems of vessel curvature.
- Problems of spatial orientation.

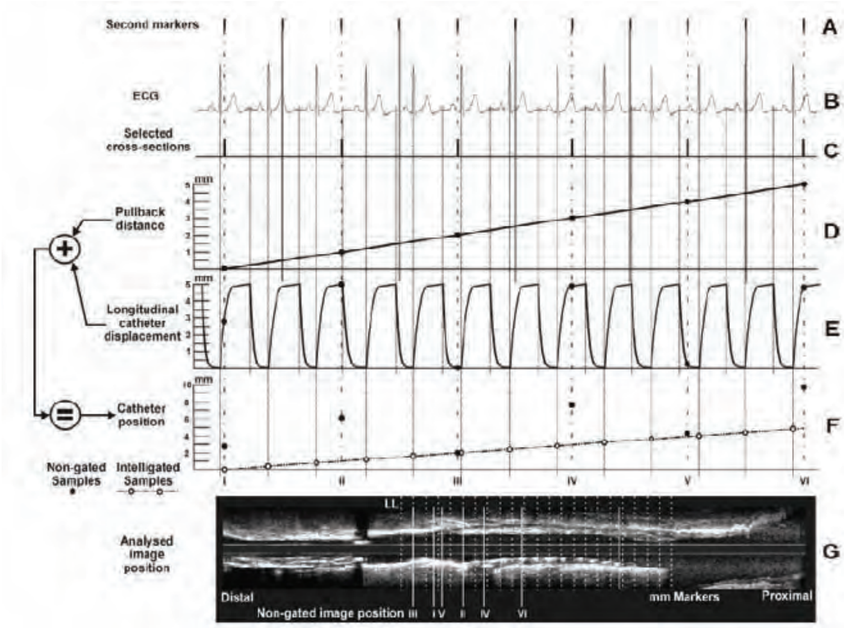
The second type of artifacts is treated by image processing and therefore falls outside the scope of this paper. The pumping heart, respiring lungs and

Figure 1. IVUS immersive data acquisition, measurements and remodeling



A: IVUS catheter/endosonics and the wire. B: angiograophy of a contrast-filled section of coronary artery. C: two IVUS cross-sectional images and some quantitative measurements on them. D: Virtual L-View of the vessel reconstruction. E: Virtual 3D-Impression of the vessel reconstruction.

Figure 2. Trajectory position anomalies of the invasive camera



moving immersed catheter camera cause the other three types of artifacts. These cause distortion on longitudinal position, x-y position, and spatial orientation of the slices on the coronary vessel. For example, it has been reported that more than 5 mm of longitudinal catheter motion relative to the vessel may occur during one cardiac cycle (Winter et al., 2004), when the catheter was pulled back at 0.5 mm/sec and the non-gated samples were stored on S-VHS videotape at a rate of 25 images/sec. Figure 2 explains the longitudinal displacement caused by cardiac cycles during a camera pullback in a segment of coronary artery. The catheter position equals to the sum of the pullback distance and the longitudinal catheter displacement. In F, the absolute catheter positions of solid dots are in disorder, which will cause a disordered sequence of camera images. The consecutive image samples selected in relation to the positions of the catheter relative to the coronary vessel wall are highlighted in G. In conclusion, these samples used for analysis are anatomically dispersed in space (III, I, V, II, IV, and VI).

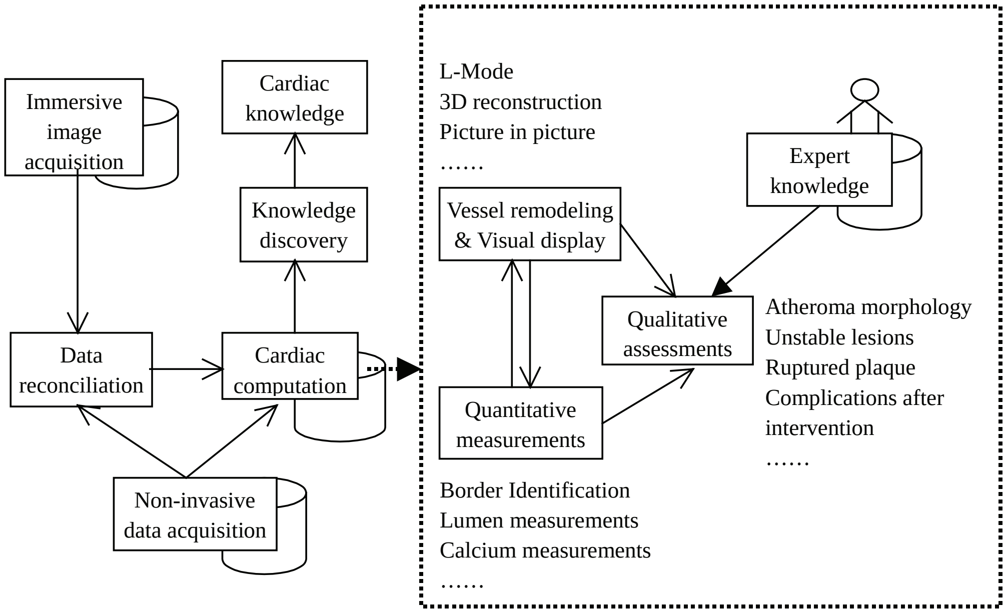
The data of IVUS sources are voluminous separated slices with artifacts, and the quantitative measurements, the physicians' interpretations are also essential components to remodel the vessel, but the accompanying mathematical models are poorly characterized. Mining in these datasets is more difficult and it inspires interdisciplinary work in medicine, physics and computer science.

MAIN THRUST

Immersive Image Mining Processes in Cardiology IVUS

IVUS image mining includes a series of complicated mining procedures due to the complex immersive data: data reconciliation, image mining, remodeling and VR display, and knowledge discovery. Figure 3 shows the function-driven processes.

Figure 3. Immersive IVUS image mining



The data reconciliation compensates or decreases artifacts to improve data, usually fused with non-invasive inspection cardiology data such as angiography and ECG etc. This data mining extracts features in the individual dataset to use for data reconciliation.

Cardiac computation is taken on the reconciled individual dataset. Quantitative measurement calculates features such as lumen, atheroma, calcium and stent on every slice. Vessel remodeling based on these measurements forms a straight 3D volume, and the fusion with the pullback path determined from angiography yields an accurate spatio-temporal (4D) model of the coronary vasculature. This 4D model is used for further cardiac computation to gain VR display and volume measurements on the vessel. IVUS images are fundamentally different from histology and cannot be used to detect and quantify specific histologic contents directly. But based on quantitative measurements and VR display, combined cardiac knowledge, the qualitative assessment such as atheroma morphology, unstable lesions and complications after intervention may be gained semi-automatically or artificially by the physicians.

These quantitative measurements and qualitative assessments may be organized and stored in a database or data warehouse. Statistics-based mining on colony data and mining on individual history data lead to knowledge discovery of heart diseases.

Individual Data Mining to Reconcile the Data

Data reconciliation is the basis and most important step to cardiac computation, and it is expecting more effective methods to attack the data artifacts. Three types of methods are applied: parsimonious data acquiring, data fusion of invasive and non-invasive data, and hybrid motion compensation.

- **Parsimonious Phase-Dependent Data Acquisition:** Cardiac knowledge dictates systolic/diastolic timing features. It has been suggested to select IVUS images recorded in the end-diastolic phase of the cardiac cycle, in which the heart is motionless and blood flow has ceased so that their influences on the catheter can be neglected. Online ECG-gated pullback has been used to acquire phase-dependent data but the technology is expensive and prolongs the acquisition procedure. Instead, a retrospective image-based gating mining method has been studied (Winter et al, 2004; Zhu, Oakeson, & Friedman, 2003). In this method, the different features are mined from sampled IVUS images over time by transforming the data with spectral analysis, to discover the most prominent repetition frequencies of appearance of these image features. From this mining, the near images at the end-diastolic phases can be inferred. The selection of images is parsimonious: only about 5% of the dataset are selected, wherein about 10% of the selections are mispositioned.
- **Data Fusion:** The motion vessel courses provide helpful information to identify space and time of the IVUS camera in the coronary of the moving heart. Fusion of complementary information from two or more differing modalities enhances insights into the underlying anatomy and physiology. Combining non-invasive data mining techniques to battle measurement errors is a preferred method. The positioning for the camera could be remedied if from angiograms the outer form of a vessel is available, as a path-road for the camera. Fusing the route and IVUS data, a simulator generates a VR reconstruction of the vessel (Wahle, Mitchell, Olszewski, Long, & Sonka, 2000; Sarry & Boire, 2001; Ding & Friedman, 2000; Rotger, Radeva, Mauri, & Fernandez-Nofrerias, 2002; Weichert, Wawro, & Wilke, 2004). This should help to detect the absolute catheter spatial positions and orientations, but usually the routes are static and data are parsimonious, phase-dependent, or without exhibiting the distortion. There are few papers on the consecutive motion tracking of the coronary tree (Chen & Carrol, 2003; Shechter, Devernay, Coste-Maniere, Quyyumi, & McVeigh, 2003), but they omitted the stretch of the arteries that is an important property for accurate positioning analyses.
- **Hybrid Motion Compensation:** Physical prior knowledge would predict space and time of the global position of IVUS camera, if it would be fully modeled. The many mechanical and electrical mechanisms in a heart make a full model intractable, but if most of its mechanisms could be dispensed with, the prediction model would become considerably simplified. A hybrid approach could solve the problem if a full 'Courant-type' model would be available, coined by the term "Computational Cardiology" (Cipra, 1999). An effective and pragmatic model is futuristic yet. Timinger reconstructed the catheter position on 3D roadmap by a motion compensation algorithm based on an affine model for compensating the respiratory motion and ECG gating method for the catheter positions acquired using a magnetic tracking system (Timinger, Krueger, Borgert, & Grewer, 2004). Fusing the empirical features of the coronary longitudinal movement with a motion compensation model is a novel way to resolve the longitudinal distortion of the IVUS dataset (Liu, Koppelaar, Koffijberg, Bruining, & Hamers, 2004).

Cardiac Computation to Quantitative and Qualitative Analysis

Cardiac computation aims at quantitative measurements, remodeling and qualitative assessments on the vessel and the lesion zones. The technologies of border detection of image processing and pattern recognition of the vessel layers are important in the process of cardiac calculation (Koning, Dijkstra, von Birgelen, Tuinenburg, et al., 2002). For qualitative assessments, expert knowledge of physicians must be fused within reasoning to get the assessments.

Quantitative Measurements

- **Cross-Sectional Slices Calculation:** The normal coronary artery consists of the lumen surrounded by intima, media, and adventitia of the vessel wall (Halligan & Higano, 2003). The innermost layer consists of a complex of three elements: intima, atheroma (diseased arteries), and internal elastic membrance. After gaining vessel layers and their attributes through edge identification and pattern recognition of image processing, every slice is calculated separately and the measurements such as lumen area, EEM (external elastic membrance) area, and maximum atheroma thickness can be reported in detail.
- **Vessel Remodeling and Virtual Display:** Based on the above calculation results, vessel layers including plaques can be remodeled and visualized in longitude. VR display is an important way to assist inspecting the artery and the distribution of plaque and lesions, to help navigate in guided surgery facilities for minimally invasive surgery. For examples, L-Mode display sets of slices taken from a single cut plane in longitude (Figure 1, panel D); 3D reconstruction display a shaded or wire-frame image of the vessel to give an entire view (Figure 1, Panel E).
- **Derived Measurements:** Calculating on the virtual remodeled vessel, the derived measurements can be obtained such as hemodynamics, length, volumes of the vessel and special lesion zones.
- **Qualitative Assessments:** IVUS images are fundamentally different from histology and cannot be used to detect and quantify specific histologic contents directly. However, based on quantitative measurements and a virtual model, as well as combined cardiac knowledge, the qualitative assessment such as atheroma morphology, unstable lesions and complications after intervention may be gained semi-automatically or artificially to serve physicians.

Knowledge Discovery in Cardiology

Once a large amount of quantitative and qualitative features are in hand, mining on these data can reveal knowledge about the forming and regression of some heart diseases, and can simulate the different stages of the disease as well as assess surgical treatment procedure and risk level.

Some papers are reported on medical knowledge discovery in cardiology using data mining. Pressure calculation by using fluid dynamic equations on 3D IVUS volumetric measurements predicts the physiologic severity of coronary lesions (Takayama & Hodgson, 2001). Artificial intelligence methods: a structural description, syntactic reasoning and pattern recognition method are applied on angiography to recognize stenosis of coronary artery lumen (Ogiela & Tadeusiewicz, 2002). Logical Analysis of Data is a method based on combinatory, optimization and theory of Boolean functions is used on 21 variables dataset to predict coronary risk, but these dataset do not include any medical images information (Alexe, 2003). Mining in single proton emission computed tomography images accompanied by clinical information and physician interpretation using inductive machine learning and heuristic approaches to mimic cardiologist's diagnosis (Kurgan, Cios, Tadeusiewicz, Ogiela, & Goodenday, 2001).

Immersive data mining or fusing the mined data with other cardiac data is a challenge for improving medical knowledge in cardiology. Mining will contribute to some difficult cardiology applications, for example: interventional procedures and healthcare; mining coronary vessel movement anomalies; highlight local abnormal cellular growth; accurate heart and virtual vessel reconstruction; adjust ferro-electric materials to monitor heart movement anomalies; prophylactic patient monitoring etc. Some technicalities should be considered in this mining field.

- **Volume and complexity in data:** First, IVUS data have a large volume in one acquisition procedure for one person. Second, usually tracing a patient needs a long time: more than several years. Third, the data from patients maybe are incomplete. Finally, immersive data acquisition is a complicated procedure depending on body condition, immersive mechanical system, and even operating procedure. The immersive complexity may even bias the historical data from the same heart.
- **Data standards and semantics:** IVUS images, especially the qualitative assessments, need a consistent standard and semantics to support mining. IVUS documents (Mintz, et al., 2001) and DICOM IVUS standards may be the base to follow. It is necessary to consider the importance of relative parameters, spatial positions, multi-interpretations of image quantitative measurements, which also need to be addressed in the IVUS standards.
- **Data fusion:** Since heart diseases are complicated, IVUS is one of the preferred diagnose tools, so mining IVUS needs considering other clinical information including physician's interpretation at the same time.

FUTURE TRENDS

Immersive image mining in cardiology is a new challenge for medical informatics. In mining our body physicians thus inspire interdisciplinary work in medicine, physics and computer science which will improve monitoring heart data to many deep applications serving clinical needs in diagnostics, therapies, safety level, cost and risk effectiveness. For instance, in due course of time nanotechnology will mature to a degree of immersive medical mining equipment, physicians will directly control medicine by instruction via mobile communication because of a transducer inside the human body.

CONCLUSION

Over the last several years, IVUS has developed into an important clinical tool in the assessment of atherosclerosis. The limitations on data artifacts and difficult discerning between entities with similar echodensities (Halligan & Higano, 2003) are waiting for better solution. Hospitals have stored a large amount of immersive data, which may be mined for effective application in cardiac knowledge discovery.

Data mining technologies play crucial roles in the whole procedures in IVUS application: from data acquisition, data reconciliation, image processing, vessel remodeling and virtual display, knowledge discovery, disease diagnosing, clinical treatment, etc. Reconciling coronary longitudinal movement with a motion compensation model is a novel way to resolve the longitudinal distortion of the IVUS dataset. This fusion with other cardiac datasets and online processing are very important for future application, which means more effective and efficient mining methods on the complicated datasets need to be studied.

REFERENCES

Alexe, S. (2003). Coronary risk prediction by logical analysis of data. *Annals of Operations Research, 119*(1-4), 15-42.

Chen, S-Y. J. & Carroll, J. D. (2003). Kinematic and deformation analysis of 4-D coronary arterial trees reconstructed from cine angiograms. *IEEE Transactions on Medical Imaging, 22*(6), 710-720.

Cios, K. J. & Moore, W. (2002). Uniqueness of medical data mining. *Artificial Intelligence in Medicine Journal, 26*(1-2), 1-24.

Cipra, B. A. (1999). Failure in sight for a mathematical model of the heart. *SIAM News, 32*(8).

Ding, Z. & Friedman, M.H. (2000). Quantification of 3D coronary arterial motion using clinical biplane cineangiograms. *The International Journal of Cardiac Imaging, 16*(5), 331-346.

Halligan, S. & Higano, S. T. (2003). Coronary assessment beyond angiography. *Applications in Imaging Cardiac Interventions, 12*, 29-35.

Hsu, W., Lee, M. L., & Zhang, J. (2002). Image mining: Trends and developments. *Journal of Intelligent Information System: Special Issue on Multimedia Data Mining, 19*(1), 7-23.

Koning, G., Dijkstra, J., von Birgelen, C., Tuinenburg, J. C. et al. (2002). Advanced contour detection for three-dimensional intracoronary ultrasound: A validation—in vitro and in vivo. *The International Journal of Cardiovascular Imaging, 18*, 235-248.

Kurgan, L. A., Cios, K. J., Tadeusiewicz, R., Ogiela, M., & Goodenday, L. S. (2001). Knowledge discovery approach to automated cardiac SPECT diagnosis. *Artificial Intelligence in Medicine, 23*(2), 149-169.

Liu, X., Koppelaar, H., Koffijberg, H., Bruining, N., & Hamers, R. (2004, October). Data reconciliation of immersive heart inspection. Paper presented at the *IEEE International Conference on Systems, Man and Cybernetics*, Hague, Netherlands.

Mintz, G. S., Nissen, S. E., Anderson, W. D., Bailey, S. R., Erbel, R., Fitzgerald, P. J., Pinto, F. J., Rosenfield, K., Siegel, R. J., Tuzcu, E. M., & Yock, P. G. (2001). ACC clinical expert consensus document on standards for the acquisition, measurement and reporting of intravascular ultrasound studies: A report of the American college of cardiology task force on clinical expert consensus documents. *Journal of the American College of Cardiology, 37,* 1478–1492.

Ogiela, M. R. & Tadeusiewicz, R. (2002). Syntactic reasoning and pattern recognition for analysis of coronary artery images. *Artificial Intelligence in Medicine, 26,* 145-159.

Rotger, B., Radeva, P., Mauri, J., & Fernandez-Nofrerias, E. (2002). Internal and external coronary vessel images registration. In M. T. Escrig, F. Toledo, & E. Golobardes (Eds.), *Topics in Artificial Intelligence, 5th Catalonian Conference on AI, Castellón* (pp. 408-418). Spain.

Sarry, L. & Boire, J. Y. (2001). Three-Dimensional tracking of coronary arteries from biplane angiographic sequences using parametrically deformationable models. *IEEE Transactions on Medical Imaging, 20*(12), 1341-1351.

Shechter, G., Devernay, F., Coste-Maniere, E., Quyyumi, A., & McVeigh, E. R. (2003). Three-Dimentional motion tracking of coronary arteries in biplane cineangiograms. *IEEE Transaction on Medical Imaging, 2,* 1-16.

Takayama, T. & Hodgson, J. M. (2001). Prediction of the physiologic severity of coronary lesions using 3D IVUS: Validation by direct coronary pressure measurements. *Catheterization and Cardiovascular Interventions, 53*(1), 48-55.

Timinger, H., Krueger, S., Borgert, J., & Grewer, R. (2004). Motion compensation for interventional navigation on 3D static roadmaps based on an affine model and gating. *Physics in Medicine and Biology*, *49*, 719-732.

Wahle A., Mitchell, S. C., Olszewski, M. E., Long R. M., & Sonka, M. (2000). Accurate visualization and quantification of coronary vasculature by 3-D/4-D fusion from biplane angiography and intravascular ultrasound. *EBiOS 2000, Biomonitoring and Endoscopy Technologies* (pp. 144-155). Amsterdam NL, SPIE Europto, *4158*.

Weichert, F., Wawro, M., & Wilke, C. (2004). A 3D computer graphics approach to brachytherapy planning. *The International Journal of Cardiovascular Imaging, 20*, 173-182.

Winter, S. A. de, Hamers, R. Degertekin, M., Tanabe, K., Lemos, P.A., Serruys, P.W., Roelandt, J. R. T. C., & Bruining, N. (2004). Retrospective image-based gating of intracoronary ultrasound images for improved quantitative analysis: The intelligate method. *Catheterization and Cardiovascular Interventions*, *61*(1), 84-94.

Zhu, H., Oakeson, K. D., & Friedman, M. H. (2003). Retrieval of cardiac phase from IVUS sequences. In F. William, M.F. Walker, & Insana (Eds.), *Medical Imaging 2003: Ultrasonic Imaging and Signal Processing, Proceedings of SPIE 5035* (pp. 135-146).

KEY TERMS

Cardiac Data Fusion: Fusion of complementary information from two or more differing cardiac modalities (IVUS, ECG, CT, physician's interpretation etc.) enhances insights into the underlying anatomy and physiology.

Computational Cardiology: Using mathematic and computer model to simulate the heart motion and its properties as a whole.

Immersive IVUS Images: The real-time cross-sectional images obtained from a pullback IntraVascular UltraSound transducer in human arteries. The dataset is usually a volume with artifacts caused by the complicated immersed environments.

IVUS Data Reconciliation: Mining in the IVUS individual dataset to compensate or decrease artifacts to get improved data using for further cardiac calculation and medical knowledge discovery.

IVUS Standards and Semantics: It refers to the standard and semantics of IVUS data and their medical quantitative measurements, qualitative assessments. The consistent definition and description improve medical data management and mining.

Medical Image Mining: It involves extracting the most relevant image features into a form suitable for data mining for medical knowledge discovery; or generating image patterns to improve the accuracy of images retrieved from image databases.

Virtual Reality Vasculature Reconstruction: For effective applications of intravascular analyses and brachytherapy, reconstruct and visualize the vessel-wall's interior structure in a single 3D/4D model by fusing invasive IVUS data and non-invasive angiography on?

This work was previously published in Encyclopedia of Data Warehousing and Mining, edited by J. Wang, pp. 586-592 copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.21
Application of Text Mining Methodologies to Health Insurance Schedules

Ah Chung Tsoi
Monash University, Australia

Phuong Kim To
Tedis P/L, Australia

Markus Hagenbuchner
University of Wollongong, Australia

ABSTRACT

This chapter describes the application of a number of text mining techniques to discover patterns in the health insurance schedule with an aim to uncover any inconsistency or ambiguity in the schedule. In particular, we will apply first a simple "bag of words" technique to study the text data, and to evaluate the hypothesis: Is there any inconsistency in the text description of the medical procedures used? It is found that the hypothesis is not valid, and hence the investigation is continued on how best to cluster the text. This work would have significance to health insurers to assist them to differentiate descriptions of the medical procedures. Secondly, it would also assist the health insurer to describe medical procedures in an unambiguous manner.

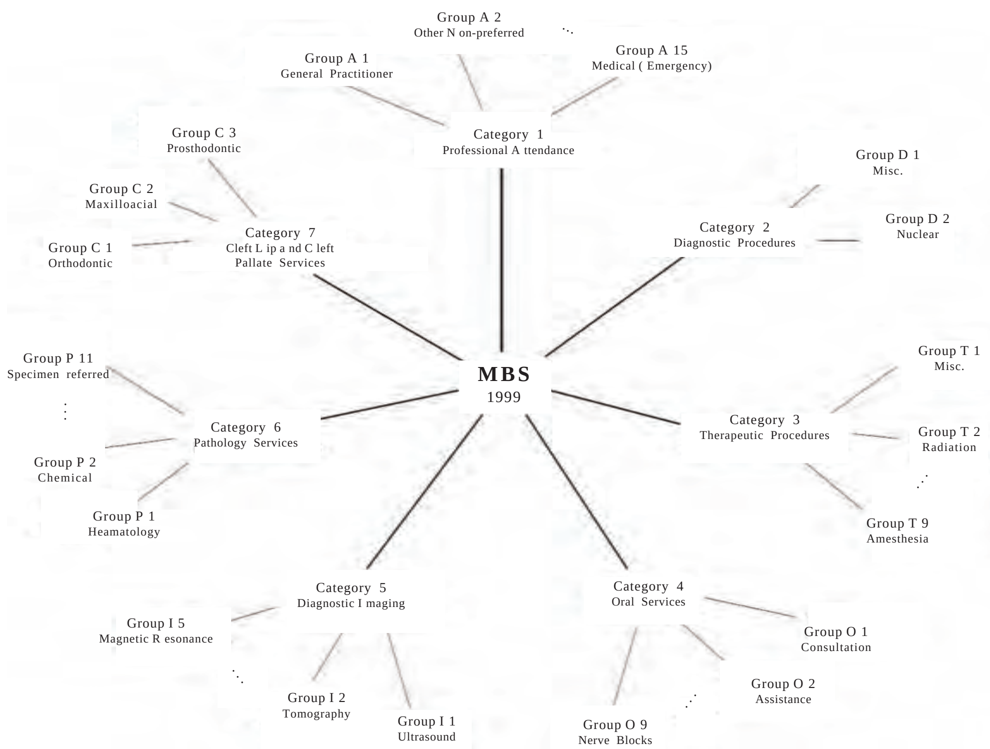
AUSTRALIAN HEALTH INSURANCE SYSTEM

In Australia, there is a universal health insurance system for her citizens and permanent residents. This publicly-funded health insurance scheme is administered by a federal government department called the Health Insurance Commission (HIC). In addition, the Australian Department of Health and Ageing (DoHA), after consultation with the medical fraternity, publishes a manual called Medicare Benefit Schedule (MBS) in which it details each medical treatment procedure and its associated rebate to the medical service providers who provide such services. When a patient visits a medical service provider, the HIC will refund or pay the medical service provider at the rate published in the MBS[1] (the MBS is publicly available online from http://www.health.gov.au/pubs/mbs/mbs/css/index.htm).

Therefore, the description of medical treatment procedures in the MBS should be clear and unambiguous to interpretation by a reasonable medical service provider as ambiguities would lead to the wrong medical treatment procedure being used to invoice the patient or the HIC. However, the MBS has developed over the years, and is derived through extensive consultations with medical service providers over a lengthy period. Consequently, there may exist inconsistencies or ambiguities within the schedule. In this chapter, we propose to use text mining methodologies to discover if there are any ambiguities in the MBS.

The MBS is divided into seven categories, each of which describes a collection of treatments related to a particular type, such as diagnostic treatments, therapeutic treatments, oral treatments, and so on. Each category is further divided into groups. For example, in category 1, there are 15 groups, A_1, A_2, …, A_{15}. Within each group, there are a number of medical procedures which are denoted by unique item numbers. In other words, the MBS is arranged in a hierarchical tree manner, designed so that it is easy for medical service providers to find appropriate items which represent the medical procedures provided to the patient.[2] This underlying MBS structure is outlined in Figure 1.

Figure 1. An overview of the MBS structure in the year of 1999



This chapter evaluates the following:

- **Hypothesis** — Given the arrangement of the items in the way they are organised in the MBS (Figure 1), are there any ambiguities within this classification? Here, ambiguity is measured in terms of a confusion table comparing the classification given by the application of text mining techniques and the classification given in the MBS. Ideally, if the items are arranged without any ambiguities at all (as measured by text mining techniques), the confusion table should be diagonal with zero off diagonal terms.
- **Optimal grouping** — Assuming that the classification given in MBS is ambiguous (as revealed in our subsequent investigation of the hypothesis), what is the "optimal" arrangement of the item descriptions using text mining techniques (here "optimal" is measured with respect to text mining techniques)? In other words, we wish to find an "optimal" grouping of the item descriptions together such that there will be a minimum of misclassifications.

The benefits of this work are as follows:

- From the DoHA point of view, it will allow the discovery of any existing ambiguities in the MBS. In order to make procedures described in the MBS as distinct as possible, the described methodology can be employed in evaluating the hypothesis in designing the MBS such that there would not be any ambiguities from a text mining point of view. This will lead to a better description of the procedures so that there will be little misinterpretation by medical service providers.
- From a service provider's point of view, the removal of ambiguities would allow efficient computer-assisted searching. This will limit misinterpretation, and allow the implementation of a semi-automatic process for the generation of claims and receipts.
- While the "optimal grouping" process is mainly derived from a curiosity point of view, this may assist the HIC in re-grouping some of their existing descriptions of items in the MBS, so that there will be less opportunities for misinterpretation.

Obviously, the validity of the described method lies in the validity of text mining techniques in unambiguously classifying a set of documents. Unfortunately, this may not be the case, as new text mining techniques are constantly being developed.

However, the value of the work presented in this paper lies in the ability to use existing text mining techniques and to discover, as far as possible, any ambiguities within the MBS. This is bound to be a conservative measure, as we can only discover ambiguities as far as possible given the existing tools. There will be other ambiguities which remain uncovered by current text mining techniques. But at least, using our approach will clear up some of the existing ambiguities. In other words, the text mining techniques do not claim to be exhaustive. Instead, they will indicate ambiguities as far as possible, given their limitations.

The structure of this chapter is as follows: In the next section, we describe what text mining is, and how our proposed techniques fall into the general fabric of text mining research. In the following section, we will describe the "bag of words" approach to text mining. This is the simplest method in that it does not take any cognizance of semantics among the words; each word is treated in isolation. In addition, this will give an answer to the hypothesis as stated above. If ambiguities are discovered by using such a simple text mining technique, then there must exist ambiguities in the set of documents describing the medical procedures. This will give us a repository of results to compare with those when we use other text mining techniques.

In the next section, we describe briefly the latent semantic kernel (LSK) technique to pre-process the feature vectors representing the text. In this technique, the intention is that it is possible to manipulate the original feature vectors representing the documents and to shorten them so that they can better represent the "hidden" message in the documents. We show results which do not assume the categories as given in the MBS.

TEXT MINING

In text mining, there are two main issues: retrieval and classification (Berry, 2004).

- **Retrieval techniques** — used to retrieve the particular document:
  - *Keyword-based search* — this is the simplest method in that it will retrieve a document or documents which matches a particular set of key words provided by the user. This is often called "queries".
  - *Vector space-based retrieval method* — this is often called a "bag of words" approach. It represents the document in terms of a set of feature vectors. Then, the vectors can be manipulated so as to show patterns, for example, by grouping similar vectors into clusters (Nigam, McCallum, Thrun, & Mitchell, 2000; Salton, 1983).
  - *Latent semantic analysis* — this is to study the latent or hidden structure of the set of documents with respect to "semantics". Here "semantics" is taken to mean "correlation" within the set of documents; it does not mean that the technique will discover the "semantic" relationships between words in the sense of linguistics (Salton, 1983).
  - *Probabilistic latent semantic analysis* — this is to consider the correlation within the set of documents within a probabilistic setting (Hofmann, 1999a).
- **Classification techniques** — used to assign data to classes.
  - *Manual classification* — a set of documents is classified manually into a set of classes or sub-classes.
  - *Rule-based classification* — a set of rules as determined by experts is used to classify a set of documents.
  - *Naïve Bayes classification* — this uses Bayes' theorem to classify a set of documents, with some additional assumptions (Duda, 2001).
  - *Probabilistic latent semantic analysis classification* — this uses the probabilistic latent semantic analysis technique to classify the set of documents (Hofmann, 1999b).
  - *Support vector machine classification* — this is to use support vector machine techniques to classify the set of documents (Scholkopf, Burges, & Smola, 1999).

This chapter explores the "bag of words" technique to classify the set of documents into clusters and compare them with those given in the MBS. The chapter also employs the latent semantic kernel technique, a technique from kernel machine methods (based on support vector machine techniques) to manipulate the features of the set of documents before subjecting them to clustering techniques.

BAG OF WORDS

If we are given a set of m documents $D = [d_1, d_2, ..., d_m]$, it is quite natural to represent them in terms of vector space representation. From this set of documents it is simple to find out the set of vocabularies used. In order that the set of vocabularies would be meaningful, care is taken by using the stemmisation technique which regards words of the same stem to be one word. For example, the words "representation" and "represent" are considered as one word, rather than two distinct words, as they have the same stem. Secondly, in order that the set of vocabularies would be useful to distinguish documents, we eliminate common words, like "the", "a", and "is" from the set of vocabularies. Thus, after these two steps, it is possible to have a set of vocabularies w_1, w_2,..., w_n which represents the words used in the set of documents *D*. Then, each document can be represented as an n-vector with elements which denote the frequency of occurrence of the word in the document d_i, and 0 if the word does not occur in the document d_i. Thus, from a representation point of view, the set of documents *D* can be equivalently represented by a set of vectors *V* - $[\mathbf{v}_1, \mathbf{v}_2,..., \mathbf{v}_m]$, where $\mathbf{v}_i$ is an n-vector. Note that this set of vectors *V* may be sparse, as not every word in the vocabulary occurs in the document (Nigam et al., 2000). The set of vectors *V* can be clustered together to form clusters using standard techniques (Duda, 2001).

In our case, we consider each description of an MBS item as a document. We have a total of 4030 documents; each document may be of varying length, dependent on the description of the particular medical procedure. Table 1 gives a summary of the number of documents in each category.

After taking out commonly occurring words, words with the same stem count, and so on, we find that there are a total of 4569 distinct words in the vocabulary.

We will use 50% of the total number of items as the training data set, while the other 50% will be used as a testing data set to evaluate the generalisability of the techniques used. In other words, we have 2015 documents in the training data set, and 2015 in the testing data set. The content of the training data set is obtained by randomly choos-

Table 1. An overview over the seven categories in the MBS

Category	Number of items
1	158
2	108
3	2734
4	162
5	504
6	302
7	62
Total	4030

Table 2. A confusion table showing the classification of documents (the actual classifications as indicated in the MBS are given horizontally; classifications as obtained by the naïve Bayes method are presented vertically)

Category	1	2	3	4	5	6	7	Total	% Accuracy
1	79	0	0	0	0	0	0	79	100.00
2	1	25	9	0	12	7	0	54	46.30
3	12	3	1323	15	10	3	1	1367	96.78
4	1	0	62	18	0	0	0	81	22.22
5	0	3	18	0	229	1	1	252	90.87
6	0	2	0	0	1	148	0	151	98.01
7	3	0	1	1	2	0	24	31	77.42

Table 3. Category-4 Items 52000, 52003, 52006, and 52009 misclassified by the naïve Bayes method as Category-3 items

Item No	Item Description
52000	Skin and subcutaneous tissue or mucous membrane, repair of recent wound of, on face or neck, small (not more than 7 cm long), superficial
52003	Skin and subcutaneous tissue or mucous membrane, repair of recent wound of, on face or neck, small (not more than 7 cm long), involving deeper tissue
52006	Skin and subcutaneous tissue or mucous membrane, repair of recent wound of, on face or neck, large (more than 7 cm long), superficial
52009	Skin and subcutaneous tissue or mucous membrane, repair of recent wound of, on face or neck, large (more than 7 cm long), involving deeper tissue

Table 4. Some items in Category 3 which are similar to items 52000, 52003, 52006, and 52009

Item No	Item Description
30026	Skin and subcutaneous tissue or mucous membrane, repair of wound of, other than wound closure at time of surgery, not on face or neck, small (not more than 7cm long), superficial, not being a service to which another item in Group T4 applies
30035	Skin and subcutaneous tissue or mucous membrane, repair of wound of, other than wound closure at time of surgery, on face or neck, small (not more than 7cm long), involving deeper tissue
30038	Skin and subcutaneous tissue or mucous membrane, repair of wound of, other than wound closure at time of surgery, not on face or neck, large (more than 7cm long), superficial, not being a service to which another item in Group T4 applies
30041	Skin and subcutaneous tissue or mucous membrane, repair of wound of, other than wound closure at time of surgery, not on face or neck, large (more than 7cm long), involving deeper tissue, not being a service to which another item in Group T4 applies
30045	Skin and subcutaneous tissue or mucous membrane, repair of wound of, other than wound closure at time of surgery, on face or neck, large (more than 7cm long), superficial
30048	Skin and subcutaneous tissue or mucous membrane, repair of wound of, other than wound closure at time of surgery, on face or neck, large (more than 7cm long), involving deeper tissue

ing items from a particular group so as to ensure that the training data set is sufficiently rich and representative of the underlying data set.

Once we represent the set of data in this manner, we can then cluster them together using a simple clustering technique, such as the naïve Bayes classification method (Duda, 2001). The results of this clustering are shown in Table 2.

The percentage accuracy is, on average, 91.61%, with 1846 documents out of 2015 correctly classified. It is further noted that some of the categories are badly classified, for example,

Table 5. Some correctly classified Category-1 items

Item No	Item description
3	Professional attendance at consulting rooms (not being a service to which any other item applies) by a general practitioner for an obvious problem characterised by the straightforward nature of the task that requires a short patient history and, if required, limited examination and management -- each attendance
4	Professional attendance, other than a service to which any other item applies, and not being an attendance at consulting rooms, an institution, a hospital, or a nursing home by a general practitioner for an obvious problem characterised by the straightforward nature of the task that requires a short patient history and, if required, limited examination and management -- an attendance on 1 or more patients on 1 occasion -- each patient
13	Professional attendance at an institution (not being a service to which any other item applies) by a general practitioner for an obvious problem characterised by the straightforward nature of the task that requires a short patient history and, if required, limited examination and management -- an attendance on 1 or more patients at 1 institution on 1 occasion -- each patient
19	Professional attendance at a hospital (not being a service to which any other item applies) by a general practitioner for an obvious problem characterised by the straightforward nature of the task that requires a short patient history and, if required, limited examination and management -- an attendance on 1 or more patients at 1 hospital on 1 occasion -- each patient
20	Professional attendance (not being a service to which any other item applies) at a nursing home including aged persons' accommodation attached to a nursing home or aged persons' accommodation situated within a complex that includes a nursing home (other than a professional attendance at a self contained unit) or professional attendance at consulting rooms situated within such a complex where the patient is accommodated in a nursing home or aged persons' accommodation (not being accommodation in a self contained unit) by a general practitioner for an obvious problem characterised by the straightforward nature of the task that requires a short patient history and, if required, limited examination and management -- an attendance on 1 or more patients at 1 nursing home on 1 occasion -- each patient

Table 6. Some correctly classified Category-5 items

Item No	Item description
55028	Head, ultrasound scan of, performed by, or on behalf of, a medical practitioner where: (a) the patient is referred by a medical practitioner for ultrasonic examination not being a service associated with a service to which an item in Subgroups 2 or 3 of this Group applies; and (b) the referring medical practitioner is not a member of a group of practitioners of which the first mentioned practitioner is a member (R)
55029	Head, ultrasound scan of, where the patient is not referred by a medical practitioner, not being a service associated with a service to which an item in Subgroups 2 or 3 of this Group applies (NR)
55030	Orbital contents, ultrasound scan of, performed by, or on behalf of, a medical practitioner where: (a) the patient is referred by a medical practitioner for ultrasonic examination not being a service associated with a service to which an item in Subgroups 2 or 3 of this Group applies; and (b) the referring medical practitioner is not a member of a group of practitioners of which the first mentioned practitioner is a member (R)
55031	Orbital contents, ultrasound scan of, where the patient is not referred by a medical practitioner, not being a service associated with a service to which an item in Subgroups 2 or 3 of this Group applies (NR)
55033	Neck, 1 or more structures of, ultrasound scan of, where the patient is not referred by a medical practitioner, not being a service associated with a service to which an item in Subgroups 2 or 3 of this Group applies (NR)

category-2 and category-4. Indeed, it is found that 62 out of 81 category-4 items are misclassified as category-3. Similarly, 12 out of 54 category-2 items are misclassified as category-5 items.

This result indicates that the hypothesis is not valid; there *are* ambiguities in the description of the items in each category, apart from category-1, which could be confused with those in other categories. In particular, there is a high risk of confusing those items in category-4 with those in category-3.

A close examination of the list of the 62 category-4 items which are misclassified as category-3 items by the naïve Bayes classification method indicates that they are indeed very similar to those in category-3. For simplicity, when we say items in category-3, we mean that those items are also correctly classified into category-3 by the classification method. Tables 3 and 4 give an illustration of the misclassified items. It is noted that misclassified items 52000, 52003, 52006, and 52009 in Table 3 are very similar to the category-3 items listed in Table 4.

It is observed that the way items 5200X are described is very similar to those represented in items 300YY. For example, item 52000 describes a medical procedure to repair small superficial cuts on the face or neck. On the other hand, item 30026 describes the same medical procedure except that it indicates that the wounds are *not* on the face or neck, with the distinguishing feature that this is not a service to which another item in Group T4 applies. It is noted that the description of item 30026 uses the word "not" to distinguish this from that of item 52000, as well as appending an extra phrase "not being a service to which another item in Group T4 applies". From a vector space point of view, the vector representing item 52000 is very close[3] to item 30026, closer than other items in category-4, due to the few extra distinguishing words between the two. Hence, item 52000 is classified as "one" in category-3, instead of "one" in category-4. Similar observations can be made for other items shown in Table 3, when compared to those shown in Table 4.

On the other hand, Tables 5 and 6 show items which are correctly classified in category-1 and category-5 respectively. It is observed that items shown in Table 5 are distinct from those shown in Table 6 in their descriptions. A careful examination of correctly-classified category-1 items, together with a comparison of their descriptions with those correctly-classified category-5 items confirms the observations shown in Tables 5 and 6. In other words, the vectors representing correctly-classified category-1 items are closer to other vectors in the same category than other vectors representing other categories.

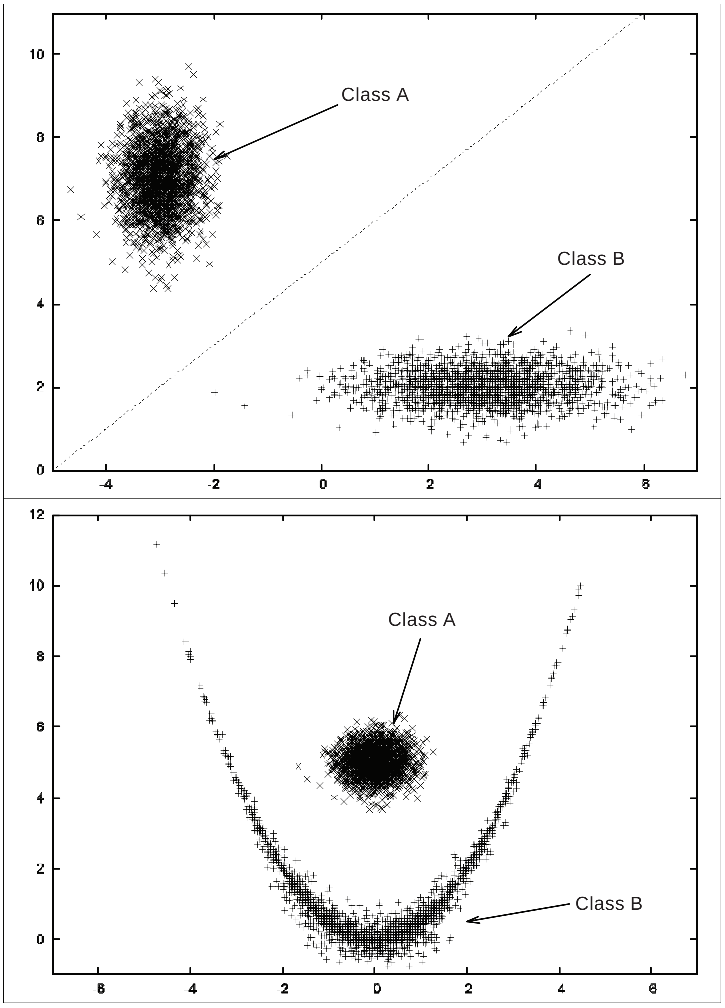
SUPPORT VECTOR MACHINE AND KERNEL MACHINE METHODOLOGIES

In this section, we will briefly describe the support vector machine and the kernel machine techniques.

Support Vector Machine and Kernel Machine Methodology

In recent years, there has been increasing interest in a method called support vector machines (Cristianni & Shawe-Taylor, 2000; Guermeur, 2002; Joachims, 1999; Vapnik, 1995). In brief, this can be explained quite easily as follows: Assume a set of (n-dimensional) vectors $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$. Assuming that this set of vectors is drawn from two classes, 1 and -1. If these classes are linearly separable, then there exists a straight line dividing these two classes as shown on the left of Figure 2. In Figure 2, it is observed that the vectors are well separated. Now if the two classes cannot be separated by a straight line, the situation becomes more interesting. Traditionally, in this case we use a non-linear classifier to separate the classes as shown on the right of Figure 2. In general terms, any two collections of n-dimensional vectors are

Figure 2. Illustration of the linear separability of classes (the two classes at top are separable by a single line, as indicated; for the lower two classes there is no line that can separate them)



said to be linearly separable if there exists an (n-1)-dimensional hyper-plane that separates the two collections.

One intuition is inspired by the following example: In the exclusive-OR case, we know that it is not possible to separate the two classes using a straight line, when the problem is represented in two dimensions. However, we know that if we increase the dimension of the exclusive-OR example by one, then in three dimensions one can find a hyper-plane which will separate the two classes. This can be observed in Tables 7 and 8, respectively.

Here it is observed that the two classes are easily separated when we simply add one extra dimension. The support vector machine uses this insight, namely, in the case when it is not possible to separate the two classes by a hyper-plane; if we augment the dimension of the problem sufficiently, it is possible to separate the two classes by a hyper-plane. $f(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b$, where $\mathbf{w}$ is a set of weights, and b a constant in this high-dimensional space. The embedding of the vectors $\mathbf{x}$ in the high-dimensional plane is to transform them equivalently to $\phi(\mathbf{x})$, where $\phi(\cdot)$ is a coordinate

Table 7. Exclusive-OR example

x	y	class
0	0	1
1	1	1
0	1	0
1	0	0

Table 8. Extended exclusive-OR example

x	y	z	class
0	0	0	1
1	1	1	1
0	1	0	0
1	0	0	0

transformation. The question then becomes: how to find such a transformation $\phi(\cdot)$?

Let us define a kernel function as follows:

$$K(\mathbf{x}, \mathbf{z}) \le \phi(\mathbf{x}), \phi(\mathbf{z}) >\equiv \phi(\mathbf{x})^T \phi(\mathbf{z}) \quad (1)$$

where ϕ is a mapping from X to an inner product feature space F. It is noted that the kernel thus defined is symmetric, in other words $K(\mathbf{x}, \mathbf{z}) = K(\mathbf{z}, \mathbf{x})$. Now let us define the matrix $\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_3]$. It is possible to define the symmetric matrix:

$$X^T X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix} \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \quad (2)$$

In a similar manner, it is possible to define the kernel matrix:

$$\mathbf{K} = [\phi(\mathbf{x}_1)\ \phi(\mathbf{x}_2)\ \ldots\ \phi(\mathbf{x}_n)]^T\ [\phi(\mathbf{x}_1)\ \phi(\mathbf{x}_2)\ \ldots\ (\mathbf{x}_n)] \quad (3)$$

Note that the kernel matrix $\mathbf{K}$ is symmetric. Hence, it is possible to find an orthogonal matrix $\mathbf{V}$ such that $\mathbf{K} = \mathbf{V}\Lambda\mathbf{V}^T$, where Λ is a diagonal matrix containing the eigenvalues of $\mathbf{K}$. It is convenient to sort the diagonal values of Λ such that $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_n$. It turns out that one necessary requirement of the matrix $\mathbf{K}$ to be a kernel function is that the eigenvalue matrix Λ must contain all positive entries, in other words, $\lambda_i \ge 0$. This implies that in general, for the transformation $\phi(\cdot)$ to be a valid transformation, it must satisfy some conditions such that the kernel function formed is symmetric. This is known as the Mercer conditions (Cristianni & Shawe-Taylor, 2000).

There are many possible such transformations; some common ones (Cristianni & Shawe-Taylor, 2000) being:

Power kernel:

$$K(\mathbf{x}, \mathbf{z}) = (K(\mathbf{x}, \mathbf{z}) + c)^p \text{ where } p = 2, 4, \ldots$$

Gaussian kernel:

$$K(\mathbf{x}, \mathbf{z}) = \exp\left(\frac{K(\mathbf{x},\mathbf{x}) + K(\mathbf{z},\mathbf{z}) - 2K(\mathbf{x},\mathbf{z})}{\sigma^2}\right).$$

There exist quite efficient algorithms using optimisation theory which will obtain a set of support vectors and the corresponding weights of the hyper-plane for a particular problem (Cristianni & Shawe-Taylor, 2000; Joachims, 1999). This is based on re-formulating the problem as a quadratic programming problem with linear constraints.

Once it is thus re-formulated, the solutions can be obtained very efficiently.

It was also discovered that the idea of a kernel is quite general (Scholkopf, Burges, & Smola, 1999). Indeed, instead of working with the original vectors **x**, it is possible to work with the transformed vectors $\phi(\mathbf{x})$ in the feature space, and most classic algorithms, for example, principal component analysis, canonical correlation analysis, and Fisher's discriminant analysis, all have equivalent algorithms in the kernel space. The advantage of working in the feature space is that the dimension is normally much lower than the original space.

Latent Semantic Kernel Technique

The latent semantic kernel method follows the same trend as the kernel machine methodology (Cristianini, Lodhi, & Shawe-Taylor, 2002). The latent semantic kernel is the kernel machine counterpart of the latent semantic technique, except that it operates in a lower dimension feature space, and hence is more efficient.

In latent semantic analysis, we have a set of documents represented by D, in terms of $V = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m]$. This set of vectors can be concatenated into a matrix **D**, then we may apply a singular value decomposition on the matrix **D** as follows:

$$\mathbf{D} = \mathbf{U}\sum\mathbf{V}^T \qquad (4)$$

where **D** is a n m matrix, **U** is an ortho normal n n matrix such that $\mathbf{U}\mathbf{U}^T = \mathbf{I}$, **V** is an ortho normal m m matrix, such that $\mathbf{V}^T\,\mathbf{V} = \mathbf{I}$, and $\sum$ is a n m matrix, with diagonal entries $\sigma_1, \sigma_2, ..., \sigma_n$, if $n > m$ or $\sigma_1, \sigma_2, ..., \sigma_m$ if $m > n$.

Often the singular values σ are arranged so that $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_n$. Thus, the singular values σ give some information on the "energy" of each dimension. It is possible that some of the σ may be small or negligible. In this case, it is possible to say that there are only a few significant singular values. For example, if we have $\sigma_1 \geq \sigma_2 \geq \sigma_i >> \sigma_{i+1} \geq \sigma_n$, then it is possible to approximate the set by $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_i$. In this case, it is possible to perform a dimension reduction on the original data so that it conforms to the reduced order data. By reducing the order of representation, we are "compressing" the data, thus inducing it to have a "semantic" representation.

The idea behind the latent semantic kernel is, instead of considering the original document matrix **D**, to consider the kernel matrix $\mathbf{K} = \mathbf{D}^T\,\mathbf{D}$. In this case, the feature space dimension is smaller, as we normally assume that there are more words in the vocabulary than the number of documents, in other words, $n \gg m$. Thus, by operating on m m matrix $\mathbf{D}^T\,\mathbf{D}$, it is a smaller space than the original n dimensional space. Once it is recognised that the kernel is K, we can then operate on this kernel, for example, performing singular value decomposition on the matrix **K**, and find the corresponding singular values.

One particular aspect of performing a singular value decomposition is to find a reduced order model such that, in the reduced order space it will approximate the model in the original space in the sense that the approximated model contains most of the "energies" in the original model. This concept can also be applied to the latent semantic kernel technique. It is possible to find a reduced order representation of the original features in such a manner that the reduced order representation contains most of the "energies" in the original representation.

The latent semantic kernel algorithm (Cristianini, Lodhi, & Shawe-Taylor, 2002) can be described as follows:

Given a kernel K, training set $d_1, ..., d_m$ and a number T:

for $i = 1$ to m do:

$norm2[i] = K(d_i, d_i)$;

for $j = 1$ to T do:

$i_j = \arg\max_i (norm2[i])$;

$index[j] = i_j$;

$size[j] =$;

for $i = 1$ to m do:

;

$norm2[i] = norm2[i]$ - $feat(i, j)$* $feat(i, j)$;

end;

end;

return $feat[i, j]$ as the j^{th} feature of input i;

To classify a new example d:

for $j = 1$ to T do:

;

end;

return $newfeat[j]$ as the j^{th} feature of the example d.

In our work, we use the latent semantic kernel method as a pre-processing technique in that we have a set of documents represented by the matrix **D**, and as mentioned previously, the matrix is quite sparse in that there are many null values within the matrix. Hence, our aim is to project this representation onto a representation in a reduced order space so that most of the "energies" are still retained. Once we obtain this set of reduced order representation, we can then manipulate them in the same way as we manipulate the full feature vectors.

Application of Latent Semantic Kernel Methodology to the Medical Benefit Schedule

We wish to apply the latent semantic kernel technique to the set of descriptions of items as contained in the MBS. The vector representation as described in in the *Bag of Words* section is used in the latent semantic kernel (LSK) approach, and it is known as the full feature vectors or, in short, full features.

We apply the LSK algorithm to the *full features* and produce reduced features which have a dimension determined by a variable T ($T \leq n$, the number of words in the corpus). In this paper, we will use the term LSK reduced features to represent such features.

We use the following procedures in our experiments with the latent semantic kernel:

- Run **bag of words method** to produce full features.
- Run **LSK algorithm** to produce LSK reduced features.
- Experiments with both **LSK** and **full features** including:
  - *Binary classification* — this will allow only binary classification using support vector machine techniques
  - *Multi-classification* — this will allow multi-class classification using support vector machine techniques
  - *Clustering the items in the MBS* using both full features and reduced features
  - *Compare the clustering result* using multi-classification results

EXPERIMENTS USING BINARY CLASSIFICATION

In this section we report the results of using a binary classification support vector machine technique for the classification of the items (Joachims, 1999). This is interesting in that it shows us the results of assuming one class, and the other items are assumed to be in a different class.

Originally, items in the MBS are classified into seven categories: 1, 2,..., 6 and 7. We have trained a binary classifier for both the full features and for the reduced features regarding each category versus the others. For example, we use category-1 and assume all the other categories are grouped as another class.

We run experiments on reduced features for each category where the dimension *T* of the reduced features was chosen from a set of values within the range [16; 2048].

We show the results of category-1 versus other categories first, and then summarise other situations and draw some general conclusions concerning the experiments.

For the experiment about category-1 versus others (see Table 9), we observe that the accuracy climbs from T = 16 rapidly to a stable position of a maximum at T= 32. Note that even though the accuracies oscillate about this maximum value for later values of T, this can be observed to be minor, and can be attributed to noise. From this, we can conclude that if we classify category-1 versus all the other categories, then using T= 32 is sufficient to capture most of the gain. What this implies is that most of the energies in the document matrix are already captured with 32 reduced order features. This shows that the LSK algorithm is very efficient in compressing the features from 4569 words in the vocabulary to only requiring 32 reduced features. Note that it is not possible to *interpret* the reduced features, as by nature they consist of a transformed set of features.

For experiments involving other sets, for example, using category-2 versus all the other categories, and so on, similar conclusions to that shown for category-1 are observed. It is found that in general a small value of T is sufficient to capture most of the energies in the document matrix. Each category versus the rest peaks at a particular value of T. On average it is found that T = 128 would capture most of the gain possible using the reduced features.

These experiments show us that the LSK algorithm is a very efficient pre-processing unit. The observed efficiency is likely due to the sparseness

Table 9. Accuracy on the training data and testing data set for Category 1 with various values of T and full features

T	**Train**		**Train**	
	Accuracy	**Correct (out of 2015)**	**Accuracy**	**Correct (out of 2015)**
16	99.60	2007	99.31	2001
32	100	2015	99.75	2010
64	100	2015	99.75	2010
128	100	2015	99.80	2011
256	100	2015	99.80	2011
400	100	2015	99.70	2009
512	100	2015	99.75	2010
800	100	2015	99.75	2010
1024	100	2015	99.75	2010
1200	100	2015	99.80	2011
1400	100	2015	99.75	2010
1600	100	2015	99.75	2010
1800	100	2015	99.75	2010
2000	100	2015	99.75	2010
2048	100	2015	99.75	2010
full features	100	2015	99.75	2010

of the full feature space. It can capture most of the energies contained in the document matrix using a small reduced feature set, in other words, with a value of T = 128.

EXPERIMENTS USING MULTIPLE CLASSIFICATIONS

In this section, we report on experiments with multi-classification using support vector machine (SVM) methodology. We first discuss the generalisation of SVM's to multi-class classification, then we describe the experimental results.

SVM, as a method proposed in Vapnik (1995), is suitable for two-class classifications. There are a number of extensions which extend this method to multi-class classification problems (Crammer & Singer, 2001; Guermeur, 2002; Lee, Lin, & Wahba, 2002). It turns out that it is possible that, instead of weighing the cost function as an equal cost (indicating that both classes are equally weighed in a two-class classification problem), one can modify the cost function and weigh the misclassification cost as well (Lee, Lin & Wahba, 2002). Once formulated in this manner, the usual formulation of SVM's can be applied.

In the *Experiments Using Binary Classification* section, we showed that by using a small reduced feature such as T = 128, we can capture most of the energies in the document matrix. The following result is obtained by running multi-class classification using support vector machine on reduced features with the same training and testing data sets as previously. The average accuracy is 88.98%, with 1793 correctly classified out of 2015 (Table 10).

Note that once again some of the HIC categories are poorly classified. For example, out of 81 HIC category-4 items, 71 are classified as category-3, while only 9 are classified as category-4. This further confirms the results obtained in the *Bag of Words* section, namely that there are ambiguities in the HIC classifications of item descriptions.

EXPERIMENTS WITH CLUSTERING ALGORITHMS

So far, we have experimented on MBS items which are classified by categories as contained in the MBS. We have shown that the MBS contained ambiguities using the bag of words approach. In this section, we ask a different question: If we ignore the grouping of the items into categories as contained in the MBS, but instead we apply clustering algorithms to the item descriptions, how many clusters would we find? Secondly, how

Table 10. A confusion table showing the classification of documents (the row gives the actual classification as indicated in the MBS, while the column shows figures which are obtained by using the support vector machine)

Category	1	2	3	4	5	6	7	total	% accuracy
1	74	0	0	0	0	4	1	79	93.67
2	0	28	13	0	8	5	0	54	51.85
3	1	6	1310	16	20	2	12	1367	95.83
4	0	0	71	9	0	0	1	81	11.11
5	0	11	25	0	211	4	1	252	83.73
6	0	5	5	1	4	136	0	151	90.07
7	0	0	4	0	0	2	25	31	80.65

efficient would these clusters be in classifying the item descriptions? Efficiency is measured in terms of the confusion matrix in classifying unseen data.

The methodology which we used is as follows:

- Use a clustering algorithm to cluster the document matrix into clusters, and label the clusters accordingly.
- Evaluate the efficiency of the clustering algorithm using a support vector machine. The efficiency is observed by examining the resulting confusion matrix.

The main reason why we need to take a two-step process is that the clustering algorithm is an unsupervised learning algorithm and that we do not have any *a priori* information concerning which item should fall into which cluster. Hence, it is very difficult to evaluate the efficiency of the clusters produced. In our methodology, we evaluate the efficiency of the clusters from the clustering algorithm by assuming that the clusters formed are "ideal", label them accordingly, and use the SVM (a supervised training algorithm) to evaluate the clusters formed.

Clustering Using Full Features: Choice of Clustering Method

The first experiment was performed on the full features (in other words, using the original document matrix). Different clustering methods were evaluated using various criteria; we found that the repeated bisection method of clustering gives the best results. Hence, we choose to use this clustering method for all future experiments.

In the clustering algorithm (Karypis, 2003; Zhao & Karypis, 2002), the document matrix is first clustered into two groups. One of the groups is selected and bisected further. This process continues until the desired number of clusters is obtained. During each step, the cluster is bisected so that the resulting two-way clustering solution optimises a particular clustering criterion. At the end of the algorithm, the overall optimisation function is minimised. There are a number of optimising functions which can be used. We use a simple criterion which measures the pair-wise similarities between two documents S_i and S_j as follows:

$$\sum_{d_q \in D_i, d_r \in D_j} \cos(d_q, d_r) = \sum_{d_q \in D_i, d_r \in D_j} d_q^T d_r \quad (5)$$

For a particular cluster S_r of size n_r, the entropy is defined as:

$$E(S_r) = -\frac{1}{\log q} \sum_{i=1}^{q} \frac{n_r^i}{n_r} \log \frac{n_r^i}{n_r} \quad (6)$$

where q is the number of classes in the data set, and n_r^i is the number of documents of the i-th class that were assigned to the r-th cluster. The entropy of the entire clustering solution is defined as:

$$E = \sum_{r=1}^{k} \frac{n_r}{n} E(S_r) \quad (7)$$

Perfect clustering means that each cluster will contain only one type of document, that is, each document in the same class belong to the same type. In this case, the entropy will be zero. In general, this is impossible. The clustering result which provides the lowest entropy would be the best clustering result. In this chapter, we made use of the Cluto software (Karypis, 2003) for performing the clustering.

Clustering Using Reduced Features into k-Clusters and Use of Support Vector Machine to Evaluate Cluster Accuracy

For experiments in this section, we used reduced features with T = 128, together with the repeated bisection clustering method. Our aim was to use a clustering algorithm with the repeated bisection method to group item descriptions into clusters irrespective of their original classification in the MBS. We divided the item descriptions into *k* (constant) clusters.

The detailed experimental procedures are as follows:

- Use reduced features of all MBS items obtained using the latent semantic kernel method as inputs into the clustering algorithm in order to group items into k-clusters, where k is a variable determined by the user.
- The output from the clustering algorithm gives *k* clusters. We perform clustering, do a classification on MBS items using the reduced features and the SVM methodology, however in this case cluster the item belonging to results from the clustering method, not the MBS category.
- The classification output can be displayed in a confusion table which can inform us on how well the clustering algorithm has performed.

In this section, we use the following notations:

- **Cluster category** — this is the cluster obtained using the clustering algorithm.
- **HIC category** — this is the category which is provided by the MBS.

Clusters (k = 7)

First, we ran a clustering algorithm to create seven cluster categories. From Table 11, we observe how items in HIC categories are distributed into cluster categories.

From this it is observed that the clusters as obtained by the clustering algorithm are quite different from those indicated in the HIC category.

We then validated the classification results from the clustering algorithm using SVM's. This informed us of the quality of the clustering algorithm in grouping the item descriptions. A classification accuracy of 93.70% on the testing data set (that is, 1888 correct out of 2015) was found.

Note that for these experiments, we used the 50% training and 50% testing data sets as in all previous experiments. The distribution of items in the training and testing data sets for each cluster category was selected using a random sampling scheme within each identified category.

Clusters (k = 8)

In this section, we experimented with eight clusters instead of seven. This provided the clustering algorithm with more freedom to choose to cluster the item descriptions. When we chose seven clusters, we implicitly tell the cluster algorithm that *no matter how the underlying data look like*, we nevertheless only allow seven clusters to be found. In this manner, even though the underlying data may be more conveniently clustered into a higher number of clusters, by choosing seven clusters we force the cluster algorithm to merge the underlying clusters into seven. On the other hand, if we choose eight clusters, this provides more freedom for the clustering algorithm to cluster the underlying data. If it is truly seven clusters, then the algorithm will report that there are seven clusters found. On the other hand, if the data is more conveniently clustered into eight clusters, then by choosing to allow for the possibility of eight clusters, it will allow the clustering algorithm to find one.

First, we ran a clustering algorithm to create eight cluster categories (Table 13).

Table 11. Distribution of MBS items into cluster categories (the cluster categories are given horizontally, and HIC categories are presented vertically)

Class	1	2	3	4	5	6	7
1	0	3	10	0	230	0	0
2	10	39	442	16	107	69	6
3	0	0	1269	62	9	1	3
4	148	13	65	2	7	17	18
5	0	4	297	34	20	0	12
6	0	12	505	47	18	35	21
7	0	37	146	1	113	180	2

Table 12. A confusion table obtained using the support vector machine method

Class	1	2	3	4	5	6	7	total	% accuracy
1	112	0	1	0	1	0	6	120	93.33
2	0	312	2	4	11	5	4	338	92.31
3	0	3	661	2	8	6	3	683	96.78
4	2	4	0	119	3	2	1	131	90.84
5	0	1	1	5	173	1	4	185	93.51
6	1	1	6	2	9	282	4	305	92.46
7	0	13	0	4	6	1	229	253	90.51

Table 13. Distribution of MBS items into eight cluster categories (the cluster categories are given horizontally, and HIC categories are presented vertically)

Class	1	2	3	4	5	6	7
1	0	4	10	0	244	0	0
2	10	39	326	17	104	70	6
3	0	0	1264	61	8	1	3
4	0	2	223	3	9	1	0
5	148	12	53	2	7	16	18
6	0	4	290	32	20	0	12
7	0	13	501	46	15	30	21
8	0	34	67	1	97	184	2

From Table 13, it is observed that the clusters as obtained by the clustering algorithm are quite different from those indicated in the HIC category.

We then validated the classification results from the clustering algorithm using SVM's. This informed us of the quality of the clustering algorithm in grouping the item descriptions. A classification accuracy is 94.14% (in other words, 1897 correct out of 2015 resulted).

Summary of Observations

It was observed that:

1. The categories as given by the clustering algorithm are good in grouping the medical item descriptions together. The evaluation using the SVM method shows that it is accurate in grouping them together, as

Table 14. A confusion table using support vector machine to validate the quality of cluster categories

Class	1	2	3	4	5	6	7	8	total	% accuracy
1	127	0	1	0	0	1	0	3	132	96.21
2	0	259	2	0	3	5	1	5	275	94.18
3	0	3	657	1	1	11	3	4	680	96.62
4	0	0	0	106	0	7	0	2	115	92.17
5	2	2	0	2	108	4	5	2	125	86.40
6	0	2	1	3	2	170	2	0	180	94.44
7	0	3	5	4	2	8	280	1	303	92.41
8	0	3	0	4	2	5	1	190	205	92.68

the confusion table has dominant diagonal elements, with few misclassified ones. The SVM is a supervised method. Once it is trained, it can be used to evaluate the generalisation capability of the model on testing results with known classifications. Thus, by examining the confusion table produced, it is possible to evaluate how well the model classifies unseen examples. If the confusion table is diagonally dominant, this implies that there are few misclassifications. On the other hand, if the confusion table is not diagonally dominant, this implies that there are high numbers of misclassifications. In our experiment, we found that the confusion table is diagonally dominant, and hence we can conclude that the grouping as obtained by the clustering algorithm is accurate. In other words, the clustering algorithm was able to cluster the underlying data together into reasonably homogeneous groupings.

2. The classification given by the clustering algorithm increases with the number of clusters, in other words, the degree of freedom in which the clustering algorithm is provided. Obviously there will be an upper limit to a reasonable number of clusters used, beyond which there will not be any further noticeable increase in the classification accuracy. It is observed that the accuracy of assuming seven clusters (93.70%) and the accuracy of using eight clusters (94.14%) are very close to one another. Hence we can conclude that the "optimum" number of clusters is around seven or eight.
3. It is noted that the items in the HIC categories are distributed in the cluster categories. This confirms our finding in the *Bag of Words* section that the HIC categories are not "ideal" categories from a similarity point of view, in that it can induce confusion due to their similarity.

CONCLUSION

We have experimented with classification and clustering on full features and reduced features using the latent semantic kernel algorithm. The results show that the LSK algorithm works well on Health Insurance Commission schedules. It has been demonstrated that the HIC categories are ambiguous in the sense that some item descriptions in one category are close to those of another. This ambiguity may be the cause of misinterpretation of the Medicare Benefits Schedule by medical service providers, leading to wrong charges being sent to the HIC for refund. It is shown that by using clustering algorithms, the item descriptions can be grouped into a number of clusters, and moreover, it is found that seven or eight clusters would be sufficient. It is noted, however, that the item descriptions as grouped using the clustering algorithm are quite different to those of the HIC categories. This implies that if the HIC wishes to re-group item descriptions, it would be beneficial to consider the clusters as grouped by using clustering algorithms.

Note that one may say that our methodology is biased towards the clustering algorithm or classification methods because we only use the simple top HIC categories — categories 1 through 7 as shown in Figure 1. This is only a very coarse classification. In actual fact, the items are classified in the MBS according to a three-tiered hierarchical tree as indicated in Figure 1. For example, an item belongs to category-x group-y item number z, where x ranges from 1 to 7, y ranges from 1 to y_i (where *i* indicates the category that it is in), and *z* ranges from 1 to z_j (where *j* depends on which group the item is located). This is a valid criticism in that the top HIC categories may be too coarse to classify the item descriptions. However, at the current stage of development of text mining methods, it is quite difficult to consider the hierarchical coding of the algorithms, especially when there are insufficient numbers of training samples. In the MBS case, a group may contain only a few items. Thus there is an insufficient number of data to train either the clustering algorithm or SVM. Hence our approach in considering the top HIC categories may be one possible way to detect if there are any inconsistencies in the MBS given the limitations of the methodology.

Even though in this chapter we have concentrated on the medical procedure descriptions of a health insurer, the techniques can be equally applicable to many other situations. For example, our developed methodology can be applied to tax legislation, in identifying whether there are any ambiguities in the description of taxable or tax-exempted items. Other applications of our methodology include: description of the social benefit schedule in a welfare state, and the description of degree rules offered by a university.

ACKNOWLEDGMENT

Financial support is gratefully acknowledged from both the Australian Research Council and the Health Insurance Commission through an ARC Linkage Project grant.

REFERENCES

Australian Department of Health and Ageing. The Medicare Benefits Schedule Online. Retrieved from http://www.health.gov.au/pubs/mbs/mbs/css/index.htm

Berry, M. W. (Ed.). (2004). *Survey of text mining: Clustering, classification, and retrieval*. New York: Springer Verlag.

Cristianini, N., & Shawe-Taylor, J. (2000). *Support vector machines*. Cambridge, UK: Cambridge University Press.

Crammer, K., & Singer, Y. (2001). On the algorithmic implementation of multi-class kernel- based vector machines. *Journal of Machine Learning Research*, *2*, 265-292.

Cristianini, N., Lodhi, H., & Shawe-Taylor, J. (2002). Latent semantic kernels. *Journal of Intelligent Information Systems*, *18*(2), March, 127-152.

Duda, R. O., Hart, P. E., & Stock, D. G. (2001). *Pattern classification*. New York: John Wiley & Sons.

Guermeur, Y. (2002). *A simple unifying theory of multi-class support vector machines* (Research Rep. No. RR-4669). Lorraine, France: Institut National de Recherche Informatique et en Automatique (INRIA), December.

Hofmann, T. (1999a). Probabilistic latent semantic analysis. *Proceedings of the 15th Conference on Uncertainties in Artificial Intelligence*. Stockholm, Sweden, July 30-August 1 (pp. 289-296).

Hofmann, T. (1999b). Probabilistic latent semantic indexing. *Proceedings of the 22nd ACM International Conference on Research and Development in Information Retrieval (SIGIR).* University of Berkeley, CA, August 15-19 (pp. 50-57).

Joachims, T. (1999). Making large-scale SVM learning practical. In B. Scholkopf, C. J. C. Burges, & A. J. Smola (Eds.), *Advances in kernel methods.* Boston: MIT Press.

Karypis, G. (2003). *Cluto, a clustering toolkit* (Tech. Rep. No. 02-017). Minneapolis, MN: Department of Computer Science and Engineering, University of Minnesota, November.

Lee, Y., Lin, Y., & Wahba, G. (2002). *Multicategory support vector machines, theory and application to the classification of microarray data and satellite radiance data* (Tech. Rep. No. 1064). Madison, WI: Department of Statistics, University of Wisconsin.

Nigam, K., McCallum, A., Thrun, S., & Mitchell, T. (2000). Text classification from labelled and unlabelled documents using EM. *Machine Learning. 39*, 103-134.

Salton, G., & McGill, M .J. (1983). *Introduction to modern information retrieval*. New York: McGraw Hill.

Scholkopf, B., Burges, C. J. C., & Smola, A. J. (1999). *Advances in kernel methods.* Boston: MIT Press.

Vapnik, V. (1995). *The nature of statistical learning theory.* New York: Springer Verlag.

Zhao, Y., & Karypis, G. (2002). *Criterion function for document clustering, experiments and analysis* (Tech. Rep. No. 01-40). Minneapolis, MN: Department of Computer Science and Engineering, University of Minnesota, February.

ENDNOTES

1. Note that this scheduled cost is somewhat different from those set by the medical service provider associations. A medical service provider can elect to charge at the scheduled cost set by the medical service provider associations, or a fraction of it, and any gaps in between the charge and the refund by the HIC will need to be met either by the patient, or through additional medical insurance cover specifically designed to cover the gap payment.
2. Note that the Medical Benefit Schedule is a living document in that the schedules are revised once every few years with additional supplements once every three months. The supplements contain minor modifications to particular sections of the schedule, while the major revisions may contain re-classification of the items, deletion or addition of items, mainly for newly introduced medical services, due to technological advances, or clarification of the intent of existing items. The version of MBS used in this chapter is based on the November, 1998, edition with supplements up to and including June 1, 2000.
3. Close here means the cosine of the angle between the two vectors is close to 1.

This work was previously published in Advances in Applied Artificial Intelligence, edited by J. Fulcher, pp. 29-51, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 3.22
A Software Tool for Reading DICOM Directory Files

Ricardo Villegas
Universidad de Carabobo, Venezuela

Guillermo Montilla
Universidad de Carabobo, Venezuela

Hyxia Villegas
Universidad de Carabobo, Venezuela

ABSTRACT

DICOMDIR directory files are useful in medical software applications because they allow organized access to images and information sets that come from radiological studies that are stored in conformance with the digital imaging and communication in medicine (DICOM) standard. During the medical application software development, specialized programming libraries are commonly used in order to solve the requirements of computation and scientific visualization. However, these libraries do not provide suitable tools for reading DICOMDIR files, making necessary the implementation of a flexible tool for reading these files, which can be also easily integrated into applications under development. To solve this problem, this work introduces an object-oriented design and an open-source implementation for such reading tool. It produces an output data tree containing the information of the DICOM images and their related radiological studies, which can be browsed easily in a structured way through navigation interfaces coupled to it.

INTRODUCTION

The digital imaging and communications in medicine (DICOM) standard (National Electrical Manufacturers Association (NEMA0, 2004a; Revet, 1997) was published in 1993. Its main goal was to establish norms for handling, storing, and interchanging medical images and associated digital information within open systems. Also it was to facilitate the interoperability among acquisition equipments and other medical devices, as well as their integration within specialized information systems in the medical and health care area.

Since then, the appearance and use of computer-assisted medical applications have increased, as a result of the accelerated technological development and the standardization process of medical information representation and handling, which generated a greater demand of development tools for those applications.

These applications range from health care information systems and picture archiving and communication systems (PACS) () solutions, to technological support systems for medical procedures, such as image-based diagnosis and surgical planning, which previously depended on the knowledge and expertise of the physicians.

In such applications, the handling of images coming from different acquisition modalities is essential. These images generated from radiological studies and stored according to the specifications of parts 10, 11 and 12 of the DICOM standard (NEMA, 2004e, f,& g) must be retrieved from storage media as a bidimensional display or in tridimensional reconstructions and other special processes, such as fusion and segmentation of images. The use of DICOMDIR directory files is almost mandatory for searching, accessing, and browsing medical images because they index the files belonging to the patient on whom the studies were performed, thus making it easier to access to those images and their associated medical information.

During the medical application software development, the use of programming interfaces (APIs) or class libraries is frequent in order to solve the computation and visualization needs, as well as for providing DICOM support to the applications. In that sense, there exist numerous public domain applications that can be used by radiologists and other specialists for reading and displaying DICOM images files and even for reading DICOMDIR index files, which cannot be integrated into applications under development because of their proprietary code.

Companies, such as Lead Technologies, ETIAM, Merge, Laurel Bridge, and DeJarnette, have commercial software development kits (SDKs) that provide complete implementations of the DICOM standard, but the acquisition costs for these SDKs are high. Open-source libraries are an alternative choice for integrating DICOM support into applications. Regarding this matter, libraries, such as visualization tool kit (VTK) (), insight segmentation and registration tool kit (ITK), DICOM tool kit (DCMTK), and virtual vision machine (VVM), allow the reading of DICOM images, but they do not provide mechanisms for reading DICOMDIR files. Like in the DCMTK case, there are other libraries that provide tools for a basic and low-level access to the information contained in the files. However, they have disadvantages, such as troublesome information retrieving process and reading tools, which are difficult to integrate into the applications.

Due to the lack of an adequate tool for reading and handling DICOMDIR files in a structured and simple way, which could be also easily coupled to browsing interfaces and attached to medical application under development, we introduce in this article the design and implementation of a DICOMDIR files reader. This tool has been successfully integrated into an application for neurosurgery preoperative planning (Montilla, Bosnjak, Jara, & Villegas, 2005), but it also can be attached to any other software under development that requires the handling of DICOM images and DICOMDIR directory files.

The next sections include the revision of related works, the essential theoretical background that frames this work within the DICOM standard context; the description of the methodology used for the implementation of the tool; and, finally, the discussion and conclusions obtained from the integration and test of the implemented reader into a medical application.

ANTECEDENTS AND RELATED WORKS

The creation of the American College of Radiology (ACR)-NEMA committee in 1983 was the product of earlier attempts by the American College of Radiology (ACR) and the National Electrical Manufacturers Association (NEMA) to establish a normative for exchanging, transmitting, and storing medical images and their associated information. The version 1.0 of the standard was published by this joint committee in 1985, under the document ACR-NEMA No. 300-1985, followed in 1988 by the document ACR-NEMA No. 300-1988 of version 2.0. Previous to these normatives, medical images were stored in files by the acquisition devices under their own proprietary formats and transferred through point-to-point communication or by removable storage media. Versions 1.0 and 2.0 established a standardized terminology and information structure, as well as hardware interfaces, software commands sets, and consistent data formats.

The most recent version, known as DICOM 3.0 (NEMA, 2004a) was published in 1993, and it was structured in parts, or documents, to facilitate its support and extension. In the last version, objects for the representation of patients, studies, reports, and other data sets were added, as well as unique identifiers for these objects, enabling the transmission of information through communication networks using the TCP/IP protocol. The DICOM 3.0 standard facilitates the transfer of images and related information within open systems containing different medical equipments and the integration with medical information systems and applications.

DICOM parts 3, 6, 10, and 12 (NEMA, 2004b, d, 4e, & g) were of particular interest for our tool design. They define and describe the DICOMDIR directory object and other information objects; attributes and representation values of DICOM information model entities; and specifications for files formats and information storage in physical media.

We have not found formal research papers related to the design and implementation of open-source tools for reading DICOMDIR files and their integration within medical applications. Nevertheless, there exist documented development libraries that include support for this kind of file. On the other hand, regarding the complexity of searching and decoding information contained in DICOM files and the fact that the tool had to be integrated into a medical application developed with C++ language, we decided to search for open-source development libraries, based upon C++ and with DICOM support in order to use them as a base for the tool development.

Just a few development libraries, besides the expensive commercial ones, enable the reading and analysis of DICOMDIR files. Due to their structures or features only three open-source libraries, based upon C++ language, were deemed appropriate to be used as reference for the tool design and implementation; the remaining APIs were found either to have a complicated structure or were based upon Java language.

GDCM (GDCM, 2005) is an API supported by Centre de Recherche et d'Applications en Traitement de l'Image et du Signal (CREATIS), which provides a fairly complete support for reading DICOMDIR files and a simple access to the information extracted from them. However, it implements only part 5 of the DICOM standard; therefore, it would be of little use as a base library for medical applications that require the implementation of other features defined by the standard.

Dicomlib library (DicomLib, 2005) provides a fuller implementation of the DICOM standard, and it also features the reading of DICOMDIR files. Although this library tries to ease the huge intrinsic complexity in the use of the DICOM standard, its access and presentation of the DICOMDIR compiled information is not the best one to be integrated into the applications.

Finally, DICOM tool kit (DCMTK) (DCMTK, 2005) from Oldenburger Forschungs und Entwicklungsinstitut für Informatik-Werkzeuge und Systeme (OFFIS) is another complete library, having several years of evolution and continous use in medical applications development. Although DCMTK provides support for the creation, modification, and opening of DICOMDIR files, it does not offer structured and simple access to the information gathered from the files. However, we selected DCMTK as the base library for the reader development due to its robustness, flexibility, and ability to handle of DICOM images.

Thus, starting from the basic functions for information searches and decoding that provides DCMTK, it was possible to develop the tool for reading the information contained in DICOMDIR files and attach this tool to medical applications that require organized access to DICOM image sets from medical studies.

THEORETICAL BACKGROUND

A lot has been written about the DICOM standard ever since it was published, including revisions and extensions. The scope of DICOM is so wide that the researcher certainly gets overwhelmed by the amount and complexity of the information contained in the standard documents. In our case, as in any other case of software development that involves the handling of information within the DICOM scope, the revision of basic fundamentals for real-world information representation according to the DICOM standard was necessary.

Within DICOM's structure, the term "information" refers to medical images coming from different acquisition modalities, signals, curves, look-up tables, and structured reports, as well as to other information gathered during patient visits to the healthcare specialist and the studies derived from them. This information and its generating agents are represented by the standard through models.

DICOM Application and Information Models

DICOM structures and organizes medical data and information through models that emulate the real-world hypothetical situation, where a patient visits a health care specialist, who later orders a set of radiological studies as shown in Figure 1.

Figure 1. Correspondence between the radiological exam environment and the DICOM Information Model

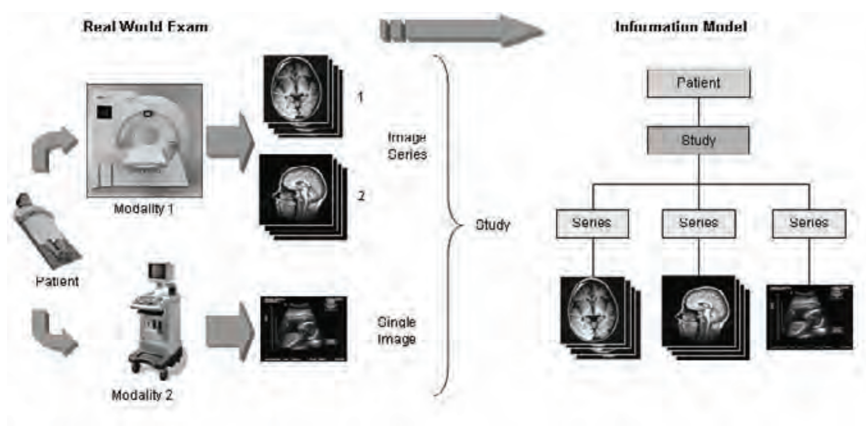
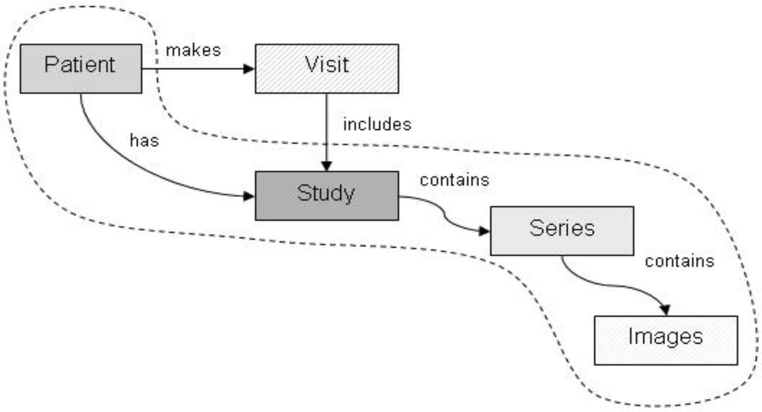


Figure 2. Simplified DICOM application model showing information entities and relationships among them



An application model is defined as an entity/relationship diagram (see Figure 2) that relates real-world objects inside the standard scope. Its diagram derives from the way hospitals' radiology departments handle images from one or more acquisition modalities. Those images are ordered in a series, according to some spatial or temporal relationship, and then stored in a folder for each patient.

Although the application model contains entities for representing several DICOM objects, such as exams results, medical reports, and study procedures, in this work only the patient, study, series, and image entities were considered because they are directly associated to data contained within the images. This consideration produces the simplified version of the DICOM application model shown in Figure 2, where the dashed region contains the entities of interest.

The DICOM information model derives from the application model and its entities are known as information object description (IOD). In an abstract way, an IOD describes real-world objects that share the same properties (NEMA, 2004b). This abstraction keeps a close relationship to the object-oriented design and programming paradigm.

Information model entities are featured by attributes whose types, multiplicities, and contents change depending upon the entity to which they belong. Attributes or data elements are defined in part 5 of the standard (NEMA, 2004c), and they are cataloged in part 6 of the data dictionary (NEMA, 2004d). An attribute is classified according to its presence in obligatory (types 1 and 2), conditional (types 1C and 2C), and optional (type 3).

Attributes are identified in the data dictionary through a tag composed from an ordered 16-bit number duple (*gggg,eeee*) expressed in hexadecimal form. These numbers represent the group and the element number within that group. The standard attributes have an even group number, different from 0000, 0002, 0004, and 0006, whereas private attributes, not contained in the data dictionary, have an odd group number, different from 0001, 0003, 0005, 0007, and FFFF. All aforementioned group numbers are reserved by the standard.

In addition to the tag, there are other data fields that also belong to the attributes structure, such as value representation (VR), value multiplicity (VM), length, and contained value. The VR describes, through a 2-byte character string, the type and format of the data contained in the attribute value field, such as integer or floating numbers, dates, string characters, and sequences. The VM specifies the cardinality or number of values that are codified in the value field. The length contains the attribute's value size in bytes. Finally, the value field stores the attribute data, according to its respective presence type.

For each IOD there are defined operation sets and named services that are executed on the information objects. When an entity needs to perform an operation over an IOD, it must request the proper service to another entity, which behaves as a server. Each object-defined service establishes a service-object pair (SOP), and the whole service set that is applicable to a particular IOD is a named SOP class.

The *media storage service class* and the *queryrRetrieve service class* are examples of SOP classes. The first one comprises the M-READ, M-WRITE, and M-DELETE services set, applied to the reading, writing, and deleting of files with image IODs coming from acquisition modalities. The second class groups the C-FIND, C-MOVE, and C-GET services that can be requested for querying and transferring information from IODs associated with different entities.

An SOP instance is the occurrence of an IOD that is operated within a communication context, through a specific service set. For example, DICOMDIR directory files are SOP products from requesting the information writing service from a directory IOD to a physical storage media.

The simplified model from our research considers only four entities and their corresponding IODs. The patient entity contains data of the patient for whom radiological studies were performed as described by the study entity. The series entity models information resulting from radiological studies, such as images or signals, and keeps some kind of spatial or temporal relationship among them. The image entity represents the images coming from some of the existing acquisition modalities, for example, computed tomography (CT), magnetic image resonance (MRI), or ultra sound (US).

DICOM File Format

The DICOM file format, described in part 10 of the standard (NEMA, 2004e), defines the way data representing a SOP instance is stored in a physical storage media. The data is encapsulated as a stream of bytes, preceded by a header with metainformation required for identifying the SOP instance and class.

The header has an organized sequence of components, named file ID, which organizes files hierarchically. An ID has up to eight components, where each one is a string of one to eight characters separated by backslashes. The file ID generally corresponds to a directory path and to a filename, for example, SUBDIR1\SUBDIR2\SUBDIR3\ABCDEFGH.

Located after the header is the data set associated with the information model entity that is stored in the file. Depending upon the entity nature, this stream of bytes could represent some of the following objects: images, curves, signals, overlay annotations, lookup tables for transforming images pixel values according to acquisition modalities or values of interest, presentation images descriptions, structure reports, or raw data.

Within the context and scope of our research, we considered only DICOM files containing images associated with studies performed on patients. Therefore, the stored data describes the image plane and the pixels features, as well as values for mapping the image to color or gray scales, overlay planes, and other specific features.

DICOM files are gathered in collections sharing a common name space, such as storage volumes or directory trees, and having unique file identifiers within it. The file collection is an abstraction of a container where files can be created or read. Each collection must be accompanied by an index file with a DICOMDIR identifier, corresponding to a

DICOM directory object instance. Part 12 of the standard (NEMA, 2004g) describes the way the DICOM file information is encoded inside the physical storage media. It depends upon the file system used by the computer system for the files creation and interchange, as well as the physical media used for it.

DICOMDIR File Format

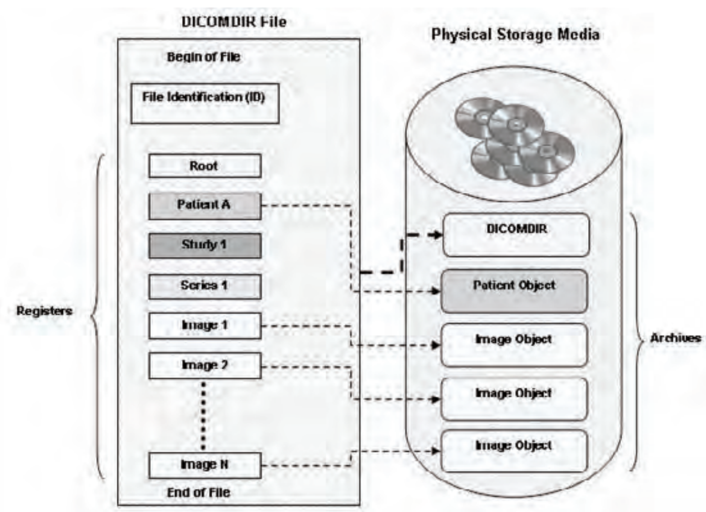
The DICOM standard defines a special object class as a named basic directory object, whose purpose is to serve as an organizing index for DICOM files stored in a physical media. The instance of a DICOM directory class object is a file with a unique filename and ID named DICOMDIR. The formal definition of the DICOMDIR object and its content are in part 3 annex F of the DICOM standard (NEMA, 2004b), whereas its structure complies with the DICOM files format specified in section 7 of part 10 (NEMA, 2004e).

There are registers in the DICOMDIR file with the information associated to objects stored in a DICOM files set, and it does not make reference to files that do not belong to the set. Each register contains a field identifying the represented information type, such as patient, study, series, and image, besides a group of specific fields with attributes extracted from the stored SOP instance. The registers are hierarchically sorted, and they are linked among themselves in the same hierarchy level and to the next lower level in the hierarchy (Figure 3).

The preceding information is generally present in the file, such as indicated in the standard, but some data fields can be optional, such as those related to the complementary information of the patient (birth date, sex, and age) and the studies (referring physician, institution, protocols, and diagnoses), as well as image descriptive information, for example, dimensions, samples by pixel, photometric interpretation, and gray-level window.

There should be a unique DICOMDIR file for each DICOM files set contained in a storage media. The DICOMDIR file location is related to the storage media directory organization, and it is commonly found at the root directory. DICOMDIR files help to make fast queries and searches throughout media contained images, without the need for reading whole file sets. Otherwise, searching and browsing images and information

Figure 3. Structure of a DICOMDIR file and its representation in a physical storage media



within file sets becomes an intensive, tedious, and difficult task.

DICOMDIR FILE READER IMPLEMENTATION

Description

The implemented tool enables DICOMDIR files to be opened in order to obtain the most relevant information from the DICOM-file collection they index. This information is organized in a hierarchical data structure, which can be easily consulted, and when coupled to a suitable graphic interface, permits the interactive browsing of the information related to the collection. The implementation was made using C++ language, based upon an object-oriented design that allows new DICOM data fields to be added to the reader. Some DCMTK library classes and methods (DCMTK, 2005) were used to facilitate the searching and decoding of the data contained within DICOMDIR files, thus avoiding the inherent complexity in the handling of the information stored under DICOM standard specifications.

Data Structures

The references to SOP instances contained in DICOMDIR files are linked according to the entities' hierarchy of PATIENT-STUDY-SERIES-IMAGE, which is implicitly established in the DICOM information model, thus establishing a natural correspondence between the hierarchy and a tree data structure. This tree has heterogeneous content nodes that correspond to the DICOM simplified application model (Figure 2), so there are five node types: root, patient, study, series, and image. The entities' relationship cardinality sets the offspring multiplicity, that is, a patient could be object from several studies, whereas each one of them could have several image series.

During the reading of a directory file, all of the file's hierarchical links are traveled, and the related collection file information is gathered for filling the nodes' specific fields, thus building the data tree structure. At the end of the reading process, the tree has the relevant collection information, having a structure similar to that shown in Figure 4. This data structure could be browsed in order to consult the information without the need of accessing the whole file set again.

An approach was considered for the tree structure implementation where the nodes behave

Figure 4. Structure of the data tree structure built by the reading tool

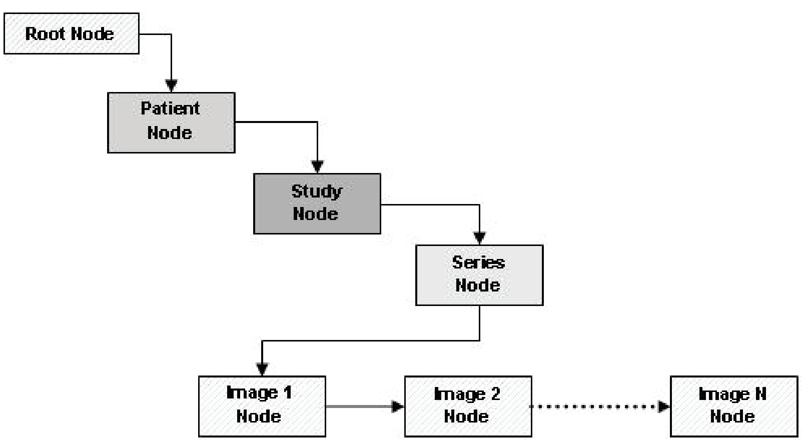
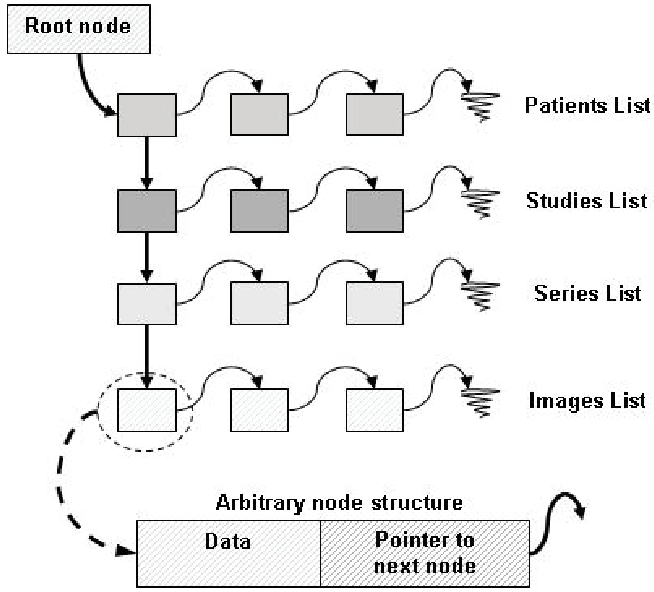


Figure 5. Data tree implementation using linked node lists as data containers



simultaneously as structural elements and data containers. In this way, the nodes establish the tree hierarchical structure and each one of them also stores the information associated to the corresponding entity. Hierarchy levels are made up of linked node lists, and each one of these lists have a children's list, corresponding to elements from the next lower hierarchy level (Figure 5). This approach facilitates the travelling of the tree and its coupling to browsing interfaces.

A class hierarchy was defined for enabling the nodes polymorphic handling through virtual methods. In this hierarchy, the base class, DICOMDIRNode, defines the basic structure for the other nodes, as well as the virtual methods for accessing and travelling the tree. All subclasses deriving from it, that is, DICOMDIRRootNode, DICOMDIRPatientNode, DICOMDIRStudyNode, DICOMDIRSeriesNode, and DICOMDIRImageNode, redefine their contents and behavior according to the corresponding DICOM application model entity to which they belong.

Collected DICOM Data Fields

DICOM files contain rather large data-element groups that are cataloged in the DICOM data dictionary (NEMA, 2004d). However, not all of these elements are useful in the scope of software applications that handle medical images, and for that reason, a selection of relevant elements was made in order to limit the number of data fields gathered during the DICOMDIR files reading process. Table 1 shows the collected data fields with their tags, value representations, multiplicities, presence attributes, and corresponding DCMTK classes and C++ standard types.

One selection criteria for the data fields is based upon the information provided by the image processing. Among the selected fields are the image dimensions, the samples per pixel number, and the gray-level window. Another criterion is the general information that guides the user during the search and selection of images, for example, patient's data, description, and modality of the study. Complementary fields that do not provide significant information for the image processing or to the user applications were not used for the gathering process, for example, address, occupation, and medical antecedents of the patient, and the technical information from the image acquisition devices.

Reading Process

The reader uses Algorithm 1 for the DICOMDIR information retrieving process.

It can be observed that in this algorithm starting from the parent nodes, each tree node and its children's nodes are recursively travelled, level by level, until reaching the deepest tree level. Only nodes from PATIENT, STUDY, SERIES, and IMAGE types are taken into account, though it is possible to extend the consideration to other types of nodes. For each visited node, the selected fields are compiled according to the node type (see Table 1). However, as explained in the DICOMDIR File Format section, not all the fields have an obligatory presence inside the DICOMDIR file, therefore the missing information must be completed from the corresponding DICOM image file, once its file identifier is known.

Some methods from the DCMTK library are used to travel through the DICOMDIR structure and retrieve its data. These methods can handle

Table 1. DICOM attributes compiled during DICOMDIR file reading process. Attributes are grouped by node type, showing main DICOM data fields and corresponding DCMTK/ C++ data types for each one of them

Node type	Tag	Field name	VR	VM	Presence	DCMTK data type	C++ data type
PATIENT	(0010,0010)	Patient name	PN	1	Obligatory	DcmPersonName	char[ ]
	(0010,0020)	Patient ID	LO	1	Obligatory	DcmLongString	char[ ]
	(0010,0030)	Birth date	DA	1	Optional	DcmDate	char[ ]
	(0010,0040)	Sex	CS	1	Optional	DcmCodeString	char[ ]
	(0010,1010)	Age	AS	1	Optional	DcmAgeString	char[ ]
STUDY	(0020,0010)	Study ID	SH	1	Obligatory	DcmShortString	char[ ]
	(0008,0020)	Study date	DA	1	Obligatory	DcmDate	char[ ]
	(0008,0030)	Study time	TM	1	Obligatory	DcmTime	char[ ]
	(0008,1030)	Description	LO	1	Obligatory	DcmLongString	char[ ]
	(0018,1030)	Protocol name	LO	1	Optional	DcmLongString	char[ ]
	(0008,0090)	Referring physician	PN	1	Optional	DcmPersonName	char[ ]
	(0008,0080)	Institution	LO	1	Optional	DcmLongString	char[ ]
	(0008,1080)	Diagnoses	LO	1..N	Optional	DcmLongString	char[ ]
SERIE	(0020,0011)	Serie number	IS	1	Obligatory	DcmIntegerString	signed long
	(0008,0060)	Modality	CS	1	Obligatory	DcmCodeString	char[ ]
	(0018,0015)	Body part	CS	1	Optional	DcmCodeString	char[ ]
	(0008,0021)	Serie date	DA	1	Optional	DcmDate	char[ ]
	(0008,0031)	Serie time	TM	1	Optional	DcmTime	char[ ]
	(0008,103E)	Description	LO	1	Optional	DcmLongString	char[ ]
	(0008,1050)	Performing physician	PN	1..N	Optional	DcmPersonName	char[ ]

continued on following page

Table 1. continued

Node type	Tag	Field name	VR	VM	Presence	DCMTK data type	C++ data type
IMAGE	(0020,0013)	Image number	IS	1	Obligatory	DcmIntegerString	signed long
	(0008,0008)	Image type	CS	1..N	Optional	DcmCodeString	char[ ]
	(0008,0023)	Image date	DA	1	Optional	DcmDate	char[ ]
	(0008,0033)	Image time	TM	1	Optional	DcmTime	char[ ]
	(0028,0010)	Rows number	US	1	Optional	Uint16	unsigned short
	(0028,0011)	Columns number	US	1	Optional	Uint16	unsigned short
	(0028,0100)	Bits allocated	US	1	Optional	Uint16	unsigned short
	(0028,0101)	Bits stored	US	1	Optional	Uint16	unsigned short
	(0028,0102)	High bit	US	1	Optional	Uint16	unsigned short
	(0028,0002)	Samples per píxel	US	1	Optional	Uint16	unsigned short
	(0028,0103)	Pixel representation	US	1	Optional	Uint16	unsigned short
	(0028,0004)	Photometric interpretation	CS	1	Optional	DcmCodeString	char[ ]
	(0018,0050)	Slice thickness	DS	1	Optional	DcmDecimalString	float
	(0028,0030)	Pixel spacing	DS	2	Optional	DcmDecimalString	float
	(0028,1050)	Window center	DS	1..N	Optional	DcmDecimalString	signed long
	(0028,1051)	Window width	DS	1..N	Optional	DcmDecimalString	unsigned long
	(0028,1053)	Rescale slope	DS	1	Optional	DcmDecimalString	float
	(0028,1052)	Rescale Intersection	DS	1	Optional	DcmDecimalString	float
	(0004,1500)	Referenced file ID	CS	1..8	Optional	DcmCodeString	char[ ]

data elements encoded either in *big Endian* or *little Endian* byte ordering, or with any kind of representation values and cardinalities. The retrieving methods are invoked by the data element tag each time a node is visited, whereas the travelling methods are used during recursive calls to the tool's main reading method.

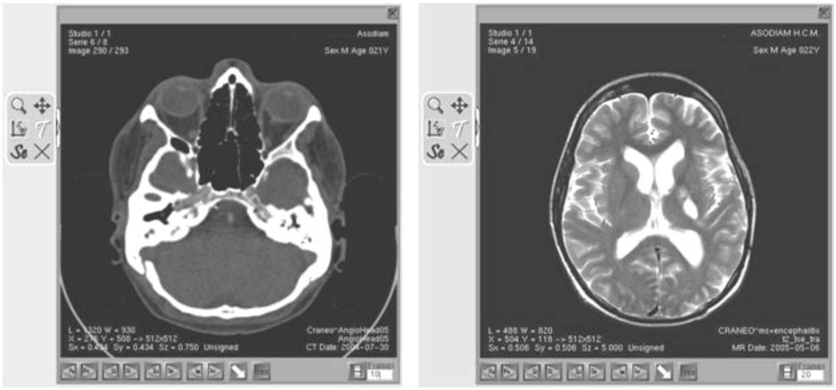
RESULTS AND DISSCUSSION

The DICOMDIR reader was integrated for its test and validation into software for neurosurgery and brachytherapy planning, developed with the VVM library (Montilla, Bosnjak, & Villegas, 2003). A graphical interface was coupled to the reader in order to enable the navigation of the information in a simple way. By just opening the associated DICOMDIR file, users will be able to have the interface display the patients' study images and

Algorithm 1.

```
Open DICOMDIR file for reading
Assign DICOMDIR file root register to NODE variable
With NODE, do recursively /* Begin of recursive block */
{
  Assign NODE first child to Next_Register variable
  While (Next_Register <> NULL)
  {
        Select according to Next_Register type
        {
        Case PATIENT:
                  Get patient node related information
                  Create a new DICOMDIR_PatientNode node object
                  Fill node members with gathered information
                  Insert node in the data tree
                  Assign Next_Register to NODE
                  Call recursive block with new NODE value
Complete missing patient information extracted from related image file
        Case STUDY:
                  Get study node related information
                  Create a new DICOMDIR_StudyNode node object
                  Fill node members with gathered information
                  Insert node in the data tree
                  Assign Next_Register to NODE
                  Call recursive block with new NODE value
Complete missing study information extracted from related image file
        Case SERIES:
                  Get series node related information
                  Create a new DICOMDIR_SerieNode node object
                  Fill node members with gathered information
                  Insert node in the data tree
                  Assign Next_Register to NODE
                  Call recursive block with new NODE value
Complete missing series information extracted from related image file
        Case IMAGE:
                  Get image node related information
                  Create a new DICOMDIR_ImageNode node object
                  Fill node members with gathered information
                  Insert node in the data tree
                  Assign Next_Register to NODE
                  Call recursive block with new NODE value
        }
        Assign NODE next child to Next_Register
  }
}                 /* End of recursive block */
```

Figure 6. Results achieved from the integration of the DICOMDIR reading tool into a medical application for neurosurgery planning. The tool is coupled with a graphical interface for browsing throughout studies, series and images.



their related medical information. Control widgets provided by the interface were used for interactively browsing and selecting the images.

The reading process was verified with DICOMDIR files associated to several image sets, coming from studies of CT and MRI performed on different patients. It could be proved that the reader recollected the information specified by the class definitions for each type of node, enabling display of its structure through the interface. Two examples from the tests are shown in Figure 6. In both examples, the studies were made on the patient's head but with two different modalities. The CT study has an eight-image series, whereas the MRI study contains 14 images.

It was observed that access to the information collected from the studies can be made in an organized and fast way, making transparent the navigation process of data structures and improving the efficiency in the use of the software. The tool is able to be integrated with other applications under development that require the handling of DICOM images and DICOMDIR files.As is usual in software projects implying code reusability, benefits from using our tool are going to be directly reflected in ease and speed of development.

CONCLUSION

The design used for the DICOMDIR file reader allows the medical application programmer to effortlessly incorporate the feature for the handling of this kind of file in software under development. Also, it avoids exhaustively understanding the DICOM standard and DCMTK classes and methods, which can be really hard and bothersome, enabling the programmer to focus on the integration of the tool into the application, as well as on the development of a navigation interface, according to the application's needs. Because the reading tool has open-source code, it can be reused and modified at will by the programmers. In addition, the design used for the tool development enables inclusion of new data fields to the presently used entities, as well as to add other information model entities to the data tree structure.

By integrating the tool into a medical application that handles DICOM image sets, it was proved it facilitates the browsing and searching of the information contained in the image sets, accelerating the fulfillment of these tasks and improving the efficiency and performance of the application. Open-source programming tools of this type also facilitate the development of medical applications and help to reduce software costs, thus making it more accessible to health institutions, physicians, and patients, particularly in regions where investment in health care solutions is either limited or not a priority issue.

FUTURE WORKS

We are considering the future implementation of tools for converting images from other formats to the DICOM format and for the creation of DICOMDIR files from nonindexed studies files. A tool for anonymization of DICOM fields also could be useful in protecting patient confidentiality during the sharing of clinical data among research teams. We hope that the implementation and subsequent use of this tool set will represent an incremental increase in the efficiency, quality, and speed of development of medical informatics software, producing applications that will be more complete and flexible at same time.

ACKNOWLEDGMENT

The authors want to express their gratitude to Dr. Luis Iván Jara, a neurosurgeon appointed to the Hospital Metropolitano del Norte (Valencia-Venezuela), for providing us the DICOM image sets used for the functional verification and evaluation of the reading tool. Financial support for this research was granted by Scientific and Humanistic Development Council of the University of Carabobo, Venezuela, under project CDCH-UC No. 0461-06.

REFERENCES

Digital imaging and communications in medicine took kit (DCMTK). (2005). *DICOM toolkit software documentation*. Oldenburger Forschungs und Entwicklungsinstitut für Informatik-Werkzeuge und Systeme (OFFIS). Retrieved on September 29, 2005, from http://dicom.offis.de/dcmtk.php.en

DicomLib. (2005). *DICOM software documentation*. Sunnybrook and Women's College Health Sciences Center Imaging Research Group. Retrieved on September 29, 2005, from http://dicomlib.swri.ca/dicomlib/html/index.htm

GDCM. (2005). *Grass roots DICOM software documentation*. Centre de Recherche et d'Applications en Traitement de l'Image et du Signal (CREATIS). Retrieved on September 29, 2005, from http://www.creatis.insa-lyon.fr/Public/Gdcm

Montilla, G., Bosnjak, A., Jara, L. I., & Villegas, H. (2005). Computer assisted planning using dependent texture mapping and multiple rendering projections in medical applications. *Proceedings from 3rd European Medical & Biological Engineering Conference and Ifmbe European Conference on Biomedical Engineering. Ifmbe* (pp. 4420-4425). Praga, Czech Republic: .

Montilla, G., Bosnjak, A., & Villegas, H. (2003). Visualización de mundos virtuales en la medicina. bioingeniería en Iberoamérica: Avances y desarrollos. In C. Müller-Karger & M. Cerrolaza (Eds.), *Centro internacional de métodos numéricos en ingeniería CIMNE* (pp. 519-545). Barcelona: .

National Electrical Manufacturers Association (NEMA). (2004a). *Digital imaging and communications in medicine. Part 1: Introduction and overview* (NEMA Standards Pub., PS3.1).

National Electrical Manufacturers Association (NEMA). (2004b). *Digital imaging and communications in medicine. Part 3: Information object definitions* (NEMA Standards Pub., PS3.3).

National Electrical Manufacturers Association (NEMA). (2004c). *Digital imaging and communications in medicine. Part 5: Data structures and encoding* (NEMA Standards Pub., PS3.5).

National Electrical Manufacturers Association (NEMA). (2004d). *Digital imaging and communications in medicine. Part 6: Data dictionary* (NEMA Standards Pub., PS3.6).

National Electrical Manufacturers Association (NEMA). (2004e). *Digital imaging and communications in medicine. Part 10: Media storage and file format for media interchange* (NEMA Standards Pub., PS3.10).

National Electrical Manufacturers Association (NEMA). (2004f). *Digital imaging and communications in medicine. Part 11: Media storage application profiles* (NEMA Standards Pub., PS3.11).

National Electrical Manufacturers Association (NEMA). (2004g). *Digital imaging and communications in medicine. Part 12: Media formats and physical media for media interchange* (NEMA Standards Pub., PS3.12).

Revet, B. (1997). *DICOM cookbook for implementations in modalities* (Tech. Rep.). Philips Medical Systems.

APPENDIX

Glossary of Terms

American College of Radiology (ACR)/ National Electrical Manufacturers Association (NEMA)(ACR): This joint committee is responsible for the development and maintenance of the DICOM standard.

Acquisition modalities: These imaging techniques and devices provide radiological images from patients' anatomy. Most commonly used modalities are CT (computed tomography), MRI (magnetic resonance imaging) and US (ultra sound).

Application programming interfaces (APIs): A set of specialized functions, class libraries, and tools used by software programmers to facilitate the application development process.

Attribute: A property of an information object represented as a data element.

Big Endian: An encoding scheme where multiple byte values are encoded with the most significant byte first, followed by the remaining bytes in decreasing order of significance.

Browsing interfaces: These mechanisms and widgets provide a way to navigate information within software applications.

Data element: This is a single atomic information unit that is related to a real-world object attribute and is defined by an entry in the DICOM data dictionary.

Digital imaging and communications in medicine (DICOM): This standard establishes norms for handling, storing, and interchanging medical images and associated digital information.

DICOM application model: An entity/relationship diagram used to model the relationships existing between real-world objects within the DICOM standard's scope.

DICOM data dictionary: This is a catalog of DICOM data elements that describes the semantics and contents of each one of them.

DICOM information model: An entity/relationship diagram used to model the relationships between the information-object definitions representing classes of real-world objects defined by the DICOM application model.

DICOMDIR file: This is a unique and mandatory file that accompanies a file set and indexes these files.

Entity/relationship diagram: This is a graphical representation of a set of objects and the relationships existing among them.

Image-based diagnosis: Techniques used by health care specialists to diagnose patients' diseases through analysis of radiological imaging studies.

I**nformation object description (IOD):** This is an abstract definition for real-world objects that share the same properties.

Little Endian: An encoding scheme where multiple byte values are encoded with the least significant byte first, followed by the remaining bytes in increasing order of significance.

Open source: Programming paradigm based upon software engineering principles that pursues software code reutilization to facilitate the development of applications.

Proprietary code: Programming paradigm that establishes that the software applications code is protected and is not available either for modification or reutilization.

Software development kits (SDKs): See APIs.

Service: This is an operation that can be requested for acting over an information object.

Service-object pair (SOP): This is a relationship that is established between an information object and an operation or the service applicable over it.

SOP class: This is the whole set of services applicable over an information object.

Surgical planning: This describes the preoperative procedure where the surgeon determines surgical protocols and approachg trajectories to be followed during the intraoperative stage of the surgery.

Value multiplicity (VM): This is the DICOM data field that specifies the cardinality or number of values existing for a data element.

Value Representation (VR): This is the DICOM data field that specifies the data type and format of values existing for a data element.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 2, Issue 1, edited by J. Tan, pp. 54-70, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 3.23
Technology in Physician Education

Michelle LaBrunda
Cabrini Medical Center, USA

Jose A. Cortes
Cabrini Medical Center, USA

ABSTRACT

How physicians are trained has been heavily influenced by the advent of the technology era. Technology has progressed faster than society has been able to integrate it. The same is true within schools of medicine and residency training programs. Many technological advances are available to medical educators, and the goal is to make educators aware of the possible educational tools. Traditionally, medicine has been a learn-by-doing discipline. This is becoming less and less acceptable in modern society, and new training methods are being sought, developed and implemented. Some of the modalities available to medical educators include intranet, hand-helds, virtual reality, computerized charting, computerized access to information and electronic monitoring student education. Technological advances in medical education have their uses, but there are also many drawbacks, including hardware limitations, computer failure, security issues, patient confidentiality issues, property rights, maintenance and poor attitude of those required to implement new learning systems.

INTRODUCTION

In the United States (U.S.), more than $810.7 billion per year is spent on medical expenses (Agency for Health Care Research and Quality, 2005). This is about 8% of our entire gross national income. Technological improvements have allowed for significant advances in all aspects of medicine, which aid in saving many lives each year. However, this same technology is associated with the rising costs of medical care. How this $810.7 billion is utilized within the health care system is largely determined by physicians. A physician in training is taught to use a specific, methodical thought process as part of his or her training. First, all possibilities, even the improbable ones, are considered. The most likely causes of a problem are determined and the best diagnostic tests to verify or disprove clinical suspicions are done. Only then is treatment determined and a course of action implemented. The tremendous infusion of technology into the medical field has complicated physician training by making available a plethora of diagnostic and testing options. Medical educators are faced with the challenge of appropriately integrating technological considerations into a fledgling physician's clinical thought process. This challenge is further exacerbated by the need to educate physicians to utilize appropriate technology in training and in practice among patients from a multitude of educational, cultural and linguistic backgrounds. In a nation where hospitals and medical centers must serve the needs of an enormously diverse population, it is important to be aware of the interaction of medical technology and culture.

This chapter describes the structure of medical education and the way technology is being used in the education of physicians. It discusses some advantages and disadvantages of technology in medical education and how these technological advances may impact a culturally diverse population.

BACKGROUND

It has been difficult for the medical education community to keep pace with the ever more sophisticated technologies developed on a seemingly weekly basis. There are more and more tests that can be ordered, increasingly innovative procedures that can be performed, and the emergence of new treatments that can be offered. Keeping in mind that all medical interventions have side effects, some of which are predictable and others not, where is the line to be drawn between what is medically warranted and what is too invasive, too unlikely or too expensive? As a physician, these issues must constantly be addressed, and often, a concrete demarcation does not exist. Promoting skill in the use of decision-making technology while continuing the intense depth and scope of traditional physician training is a major challenge facing medical educators today. Technology is seen in all facets of medical education, and it affects medical students on at least two levels. First, in the continually evolving world of medicine, students must not only be aware of new advances but must also learn when to apply them. Second, technology may be used to assess the quality of training or to evaluate the learning and performance of a medical student. An intranet, personal digital assistants (PDAs), virtual reality and computerized charting, as well as other available technologies, have the potential to revolutionize not only the way medicine is practiced, but the way it is taught, as well.

Medical technology, however, is not without some controversy. Unfamiliarity of some patients with the imposing appearance of even common technologies does little to promote a healthful and healing experience. While technology is generally touted as a marvelous educational and diagnostic tool, it can also have the effect of interfering with the learning process of a physician in training. Physicians must be able to function effectively while utilizing the latest technological innovations; custom and language cannot be barriers. It is equally important, however, that physicians be able to function effectively without an abundance of technological gadgetry. Verbal communication is still the cornerstone of the patient-physician relationship, and all too frequently, this relationship is sacrificed on the altar of technological advances.

It is important to remember when training physicians that emergency situations often arise at places where the latest technology is not readily available. Although there are many benefits to using technology over more invasive, traditional methods of diagnostic assessment, treatment and diagnosis of patients is trending toward an over-reliance on technology. This over-reliance is being perpetuated in the way medical students are being trained. The medical education community must determine how to most effectively optimize the use of technology without forgetting to look at the face of the patient on the other side of the device; in other words, how to practice the science of medicine without losing the art of medicine. The art of medicine includes empathy, communication skills and bedside manner. This is particularly essential in cases where patients do not have a clear command of English and may not have a full understanding of what is occurring. Technology should assist in the utilization of both clinical acumen (the science) and productive patient-physician relationships (the art), not handicap the problem-solving skills of medical students. With time, a proper balance will be found between the integration of technology and traditional medical skills, but the technological explosion has been so rapid that medical educators have been struggling to keep up.

Structure of Medical Education

Becoming a physician in the U.S. is a lengthy process, requiring years of formal education and practical training. First, the student has to complete at least 3 years of undergraduate work, although most students typically complete a bachelors of science degree in a biological science. Next, a satisfactory score on the arduous Medical College Admission Test (MCAT) exam must be achieved. The MCAT is a 1-day exam consisting of multiple-choice blocks of questions on the physical, biological and chemical sciences as well as an essay writing section that demonstrates proficiency in written communication. Once the required transcripts documenting sufficient grade point average have been compiled with MCAT test scores and letters of recommendation, a student wishing admittance to medical school may then fill out the lengthy application form and, if accepted, enter medical school.

Most medical schools are designed so that the first 2 years consist of the basic sciences and the second 2 years comprise the clinical sciences. The basic sciences are the building blocks of medicine, and include subjects such as biochemistry, histology (the study of tissues), pathology (the study of diseased tissue), physiology (the study of normal functioning) and statistics. The clinical sciences can be considered application of the basic sciences. These courses tend to focus on how to examine a patient, the recognition of disorders and diagnosis of disease. They teach skills that bridge the basic sciences and their application at patient bedsides. Courses in legal medicine and ethics are usually included throughout the curriculum and, increasingly, sensitivity to and awareness of cultural diversity is entering the course of study. After graduation from medical school, the medical doctor (MD) or doctor of osteopathic medicine (DO) is granted and the student is considered a physician, but for many, this may only be the halfway point of their education. Most physicians choose to specialize by entering into a residency program. A residency program consists of physicians with minimal clinical experience practicing medicine in a specific specialty under close supervision of more experienced physicians. Depending on the area of specialization, residency may last anywhere from 3 to 5 years. In addition, after residency many physicians choose to enroll in a fellowship program. A fellowship allows a physician to sub-specialize after residency. It is like residency in that a more experienced physician supervises a less experienced one. A fellowship usually lasts 1 to 3 years, but could be longer, and multiple fellowships may be completed by a diligent physician.

To become a physician in the U.S., mastery of academic medical content as well as effective application of knowledge and skill must be demonstrated by satisfactory scores on a rigorous series of medical examinations. Passing scores on four exams are required to obtain a professional license to practice medicine: the U.S. Medical

Licensing Exams (USMLE) Step 1; USMLE Step 2 Clinical Knowledge (CK); USMLE Step 2 Clinical Skills (CS); and USMLE Step 3. Three of the four exams are multiple choice in format and computer based. USMLE Step 2 CS involves actors feigning illnesses and an evaluation of the future physician's ability to interact with a patient, perform an adequate physical exam and determine appropriate diagnostic steps. USMLE Step 3, in addition to the multiple-choice blocks, has a section of interactive case studies that require insightful application of academic knowledge and competent judgment in the assessment of patient treatments. Step 1 evaluates the basic sciences, Step 2 (CK and CS) the clinical knowledge base and skills, and Step 3 the application of the clinical sciences. Additional exams are administered after the completion of residency and fellowship. These additional examinations are not required to practice medicine within the area of specialization, but are required if the physician wishes to be board certified.

The requirements for licensure vary from state to state. All states currently require that the USMLE exams or equivalent be passed. Some require additional years of supervised training, especially if the physician graduated from a foreign medical school.

On July 1, 2003, the Accreditation Council for Graduate Medical Education imposed a new set of guidelines limiting the workweek of physicians in residency to 80 hours. Prior to this date, residents were often required to work many more than 80 hours per week. This exposed them to the largest number of clinical scenarios possible during their years of residency training and maximized the hands-on learning experience. With the new restrictions on resident work hours, new, more efficient methods must be employed to ensure sufficient, practical physician training.

Technology in Medical Education

Traditionally, medicine was a learn-by-doing discipline. This has become less and less acceptable in modern society due to risks associated with inexperience. To cover any possible gaps in the training of resident physicians, new training methods are being sought, developed and implemented. There are a variety of technology modalities through which medical education may be enhanced. PDAs, virtual reality and computerized charting are technologies gaining a strong foothold in the diagnosis of patient ailments, the organization and availability of pertinent patient records and the training of new physicians.

Intranets

Many medical schools and hospitals have installed campus-wide intranet networks. The intranet systems are similar to the Internet except they have more limited access. These local systems provide a structured environment for residents and allow great access to information. This information pertains specifically to issues of importance to medical students. It could be information on schedules, required reading, lecture notes, photographs, or audio and video clips. Since many medical students are off campus for the last year of medical school, these local networks give the possibility of a virtual campus and provide continuity of education (Ward, Gordon, Field, & Lehmann, 2001). Medical tasks now require more processing, communication and transfer of greater amounts of information. Intra-institutional communication systems, such as intranets, are useful in organizing this information. Intranets can allow the most current information on a patient to be immediately available to many people at various sites. This allows laboratory values, radiology results and consultations from specialists to be available to both students and medical workers at the same time. Patient care is increasingly shifting from inpatient to ambulatory care, and training models have not kept up with the change. Hospital settings create a "captive" population from which medical students can learn. As medicine shifts toward a more outpatient-oriented environment, physicians in training have less access to patients undergoing treatment. Intranets make possible a simplified media through which many people may have access to an interesting case while respecting the patient's right to privacy. An intranet has much more utility for medical students than just permitting access to patient information. It can serve as a repository for evaluation forms, instruction, schedules, directory, and as a bulletin board for transmitting important information (Zucker, White, Fabri, & Khonsari, 1998). Intranets increase the accuracy of information, because the data is only entered one time rather than being transferred and transcribed numerous times as seen in traditional paper-based systems.

PDAs

PDAs include the Palm system and Pocket PCs. Both are being used more and more commonly in the field of medicine. PDAs have many utilities in the field of medical education. They allow for quick access to medication information, including drug-drug interactions, diagnostic manuals and medical calculators.

Currently, there are programs integrating physicians' PDAs into the hospital computer system. This permits rapid and accurate access to patient information as well as bedside data entry. Wireless communication is rapidly facilitating communication between health care providers of different departments, even though they are not able to see the patient at the same time. It is common during physician education that a student is only exposed to one aspect of a patient's care. The surgery residents are minimally exposed to a patient's other medical problems, and the medical residents are peripherally aware of the surgical issues. PDA systems allow students and residents to have access to all pertinent information simultaneously. Legibility problems are corrected by the typed format and subsequent errors are minimized. PDAs have other useful purposes in the education of physicians. They can serve as a procedural log. As a log, it has been shown that PDAs have high reliability, dependable data integrity, low data-entry workload and rapid feedback for residents and program directors for residents' procedural experience (Nicolaou & Davis, 2001). They are relatively low in cost, are very easy to use and require minimal or no training. It is clear that soon physicians will use PDAs for prescribing, ordering studies, checking results, capturing charges and dictating notes. Information can be integrated from multiple sites and gives students and residents access to pertinent information, scientific calculators and drug interaction information, and alerts physicians and pharmacists of potential conflicts with medications. Basic clinical guides and medication prescribing information may be stored on PDAs and quickly accessed. This is important, since medical residents are usually responsible for initiating therapy in emergency situations at hospitals with residency programs. PDAs are increasingly being integrated into the hospital environment, and it is important that medical students become confident in employing them as effective tools that provide access to critical information, even when they are not integrated into hospital databases.

Virtual Reality

As technology has progressed, so have the available treatment and diagnostic modalities. The traditional learning-by-doing methodology has become less acceptable as more invasive and delicate procedures are required for optimal patient care. Students are expected to master complex skills with no errors and minimal experience. Medical educators are forced to create options other than practicing on live patients. These alternative options include Web-based education, virtual reality and high-fidelity human patient simulator. An additional benefit of this training is the ability to practice patient comfort in the presence of technological devices. Care for the patient includes more than just the application of modern technologies. A patient's background, experience and command of English can have a significant influence on his sense of well-being during treatment and can affect the success of an important procedure. A trusting patient-physician relationship is especially important when working with patients with different cultural and linguistic backgrounds.

One of the most promising types of virtual reality is called augmented reality. Augmented reality consists of superimposing a virtual world on the real one so both are experienced simultaneously (Vozenilek, Huff, Reznek, & Gordon, 2004). For example, a virtual image of a complicated fracture may be superimposed on a patient's leg, allowing a surgeon to "see" the broken bone before beginning a surgery. Fluoroscopy is an X-ray that may be taken during surgery. It has many applications, most of which are related to placement of external devices into the body. Fluoroscopy allows the position of these devices to be observed and adjusted while a surgical procedure is still in process. The disadvantages include radiation exposure of medical personnel and patients from the fluoroscopy machine and poor quality of fluoroscopic images. Continuing with the broken leg example from above, virtual reality allows a surgeon to take an initial set of fluoroscopic images and load them into a virtual reality program, then observe an image of the broken bone projected by the virtual reality simulator. These virtual images can be used during surgery to obtain proper alignment of the bones while minimizing radiation exposure. These true life images can be recorded and used by medical faculty for educational purposes while avoiding exposure of students and residents to radiation (Gruntzner & Suhm, 2004).

Most virtual reality has been designed with particular surgical specialties in mind, although some are also useful for emergency medicine. Many of these virtual-reality systems have been designed with the goal of producing high-fidelity patient simulators that can give realistic tactile, auditory and visual stimuli. Full-body simulation allows physicians to practice procedural skills with no patient risk. Immersive simulators are emerging that may also be useful in emergency medicine. They provide a complete computer-generated environment, not just a specific portion of the environment, which allows for organized team simulations (Vozenilek et al., 2004; Medical Readiness Trainer Team, 2000).

Medical schools in the field of psychiatry are also beginning to integrate virtual reality into their programs. One of the treatments for phobias is to gently expose a patient to that which he or she fears. Traditionally, this was done through photographs and visualization techniques. Virtual reality is being used by many psychiatrists as part of the treatment regime. Immersing a patient with a phobia into a virtual simulation requires a delicate approach, and medical students clearly must be adequately supervised and trained (Krijin, Emmelkamp, Olafsson, & Biemond, 2004; Klinger et al., 2005). Rehabilitation medicine is also adapting virtual-reality training to assist recuperating patients. Virtual reality enables patients to try motor skills without risk of injury. Although there may have been some initial doubt as to the effectiveness of this treatment, it has been shown that patients can learn motor skills from virtual reality (Holden, 2005). Medical students need to be aware that these types of rehabilitation options exist, and rehabilitation residents should have some exposure to virtual rehabilitation as part of their training.

Simulated patients made out of electrical conductors, computer chips and circuits all wrapped in a thin layer of plastic emulating human skin have been designed for use in training medical personnel in various diagnostic simulations. There are a number of types of these patient simulators and each is designed for educational accommodation in a specific medical area. They may be used to learn how to place IVs, for auscultation of heart sounds and for practicing management of a patient whose heart stops, among other uses. Depending on the model, different degrees of reality are mimicked and students can practice procedures related to both common medical situations and emergencies. Interactive patient simulators allow students to learn to anticipate common reactions, avoid common mistakes and become acquainted with high-risk, adrenaline-filled situations in a safe environment.

Integration of simulators into medical programs is becoming more and more common, but is still a slow and expensive process. Currently, robotic surgeries are being done and robotic surgery simulators are being implemented in a number of programs. Both in the simulator and during the real surgery, robotic arms are controlled at a console. The advantage of the simulator in this case is that the simulator can be set on record mode and can give feedback as to whether or not the surgical movements were safe. Also, alternate options and approaches can be explored by the physician. What would have happened if X or Y had been performed are questions that cannot be answered using patients as practice subjects, but simulators can answer these sorts of questions. Currently, each surgical robot costs about $1.2 million (Conn, 2004), but as technology advances the price will drop, making these robots available to more training programs.

Pulmonologists who train on bronchoscopy simulators are just as fast and competent as someone who had done several bronchoscopies on human subjects (Conn, 2004). It can be inferred from the very limited amount of data available that high-quality patient simulators would decrease the number of procedural complications by allowing physicians with less experience to master difficult skills in a non-life-threatening environment. It is possible that as technology progresses, simulator training will become so integrated into the medical system that it will become part of the certification process for medical specialties that require procedural skills. In fact, expertise with video games has been shown to enhance surgeons' skills. Surgeons who spend more time on video games work 27% faster and make 37% fewer mistakes than those who do not (Conn, 2004).

Computerized Charting

Currently, many hospitals are converting from a paper-based charting system to computerized patient tracking. Electronic medical records give medical staff easy access to all patient information at one site. The information can be accessed from computers throughout the hospital or from physicians' PDAs. In hospitals, there is often difficulty assembling all pieces of information relevant to a patient because there are a number of different departments and sources of information involved. Specialists are consulted, radiographs taken and blood work drawn, and decisions are made by a number of different medical services. A large portion of a student physician's time may be spent trying to get all relevant information together in one place at one time for any specific patient. To provide the optimum care, access to information from a patient's previous admissions should be readily available and consulted. Computerized charting gives students and residents access to all relevant patient information, allowing for a more comprehensive understanding of each case. From an economic standpoint, computerized charting saves money. It decreases the rate at which diagnostic tests are inadvertently repeated on the same patient, helps prevent tests from being performed on the wrong patient or at the wrong site, and generally decreases the total length of hospital stay. Medical education must include training for students in how to deal with financial issues within the hospital as well as in the outpatient clinics if the future physician is to achieve economic stability. Part of this training should include how to utilize computerized charting to manage patient care in an expedient, practical and economical form.

Another technological advance is the picture-archiving and communication system. This system allows x-rays and other radiological studies to be viewed on computer rather than via printed films. A picture-archiving communication system makes it easy for students to compare films from various patients with similar diseases or from one patient over a period of time, both of which are important in clinical diagnoses as well as in learning to read radiological studies. This system is slowly being integrated by many institutions and is preferred by physicians over regular films (Nissen, Abdulla, Khandheria, Kienzle, & Zaher, 2004).

ACCESS TO INFORMATION

The current model in medical education is the mentored student. In this model, the one with least experience is responsible for undertaking all the responsibilities of patient care. When a challenge is met that the student/resident is unsure how to handle, he or she contacts the next higher-up doctor in the chain of command, and so on. On rounds, the physician with the most experience takes time to go over the individual cases and redirect medical care if needed.

This system presents a number of difficulties. Recently, there has been a move towards treating patients in a more outpatient environment. There are fewer hospitalizations, so physicians in training have less access to patients. Conversely, these patients have significantly less opportunity for treatment from a senior or highly experienced physician. The tolerance of mistakes in medicine is zero. In an effort to reduce medical mistakes so the number of errors is as close to zero as possible, physicians with more experience may be reluctant to allow residents to manage their patients. Physicians at various times are expected to care for their patients in the hospital, see patients in their office, supervise residents, perform and publish leading-edge research, and write grant proposals that bring funding to the hospital. Often, one or more of these services is performed with no compensation. The demands on a physician's time are enormous and seem to be increasing all the time. Physicians have less and less time to devote to supervision and teaching. Because of the new direction medicine is taking, all learning opportunities must be fully utilized and new learning opportunities made available. One way to maximize the learning process is through improved access to information. The paperless charting systems, intranet and PDAs already mentioned are examples of improved access to information, but other information modalities are available. For example, sound and video clips may be integrated into presentations. Some computer systems even allow for interactive videos (Nissen et al., 2004). A newer approach to teaching medicine is problem-based learning (Ward et al., 2001). Problem-based learning is different from the traditional approach to medicine in that a specific problem is simultaneously addressed from multiple vantage points. Environmental, behavioral, nutritional, social, economic, cultural and religious factors can be taken into consideration, as well as any obvious medical problems. Technology allows a single problem to be addressed from many perspectives. Chat rooms and video conferences allow experts from various fields in geographically distant locations to interact with each other, provide consultation advice and give feedback to medical students (Whitcomb, 2003).

A shortcoming of many medical education programs is the failure to treat medical students as adults (Barnete, 1995). A student's unavoidable and inevitable lack of experience is sometimes confused for ignorance. Partially because of this attitude, and partially because of the sheer amount of material that must be mastered, students are increasingly finding the lecture-based method of education unsatisfactory. Strømø, Grøttum, and Lycke (2004) showed that the introduction of computer-supported, problem-based learning systems had a positive influence on student learning. The students tended to increase their use of Web-based resources, expand their interaction with experts and lessen the use of textbooks, which are often outdated by the time they are in print (Strømø et al., 2004). Lectures can be informative, but applying skills and knowledge to case studies and interactive simulations can be much more meaningful and effective. Health care professionals must continually learn new skills, procedures and diagnostic techniques. Education never actually ends. Physicians must frequently update their knowledge base and train in new techniques. Modern communication and computerized learning tools help achieve these goals (Zak, 2004). In addition to requiring re-certification exams, many states also require practicing physicians to participate in a program of continuing medical education credits. These credits are becoming increasingly available through Web-based teaching environments. This makes both delivering and receiving the required medical education credits easier and more flexible. These more accessible educational credits may help improve compliance of physicians with the credit requirements. This requirement also offers opportunities for residents and physicians to enhance their knowledge of how cultural and religious traditions among diverse patients can impact medical treatment, and how best to respectfully and appropriately treat an individual rather than a disease or disorder.

MONITORING STUDENT EDUCATION

Computer technology can be used to directly monitor students' progress through their academic studies. There are numerous computer applications available, such as access to journal articles and abstracts, downloadable textbooks and drug information databases that can be consulted to improve educational outcomes and enrich the educational experience. Also, problems that result from poor penmanship are avoided. Presently, when one medical shift changes to another, a process called sign-out occurs. During sign-out all relevant patient information and a to-do list is passed from one shift to the next. It is done by medical residents, attending physicians, nursing staff and anyone else responsible for patient care. Medical mistakes may result if important information is not transmitted correctly or is forgotten by either shift. Technology, especially PDAs, allows for electronic sign out. Specific data on each patient may be transferred electronically. A resident, especially during off shifts, may be responsible for the well-being of 50 or more patients, often on different wards. Electronic access to basic patient information, such as a list of medical problems, current medication list and any active issues (such as a blood transfusion, for example) minimizes the risk of medical errors made due to lack of information on a specific case. In addition, electronic sign-out allows alarms to be set, reminding on-call residents of tasks that need to be done or patients that need to be checked.

Databases can be accrued for each student that monitor clinical experience and ensure that students receive comprehensive training. Students can enter patient contacts and residents can log procedures into PDAs for easy evaluation, allowing any omissions in education or experience to be identified (Fischer, Stewart, Mehta, Wax, & Lapinsky, 2003). Additionally, computers and PDAs can be used in data collection for research. Every year, more than 2 million medical journal articles are published (Hancock, 1996). This is in addition to the existing base of medical knowledge. Medical educators are already maximizing the amount of information that a student can learn in the amount of time allotted. It is impractical to extend the length of time of medical training, so methods of being more efficient must be evaluated. Technology plays a primary role in this trend toward greater efficiency.

DISADVANTAGES

People today are living longer lives than ever before, and the productive period of those lives has been greatly extended. Survival rates of the ill or injured are at an all-time high. A major contributor to this happy state of affairs is technology. This includes not only technology for diagnostic and treatment purposes, but also innovations used for development of medications and training of highly capable physicians. Technological advances have led the way to many improvements in modern medicine. However, there are also drawbacks to a dependence upon technology. These include hardware and software limitations, computer failure, security issues, patient confidentiality issues, property rights, maintenance and poor attitude among those required to implement new learning systems (Ward, Gordon, Field, & Lehmann, 2001). Currently, one of the primary problems in medical technology is lack of bandwidth. Adequate bandwidth is required for proper data transfer (Fischer et al., 2003). Also, the resources required to develop a specific technology can be prohibitively expensive (Whitcomb, 2004).

Technology is not without patient risk. Risks and costs are often underemphasized in medicine for financial and social motives (Schroeder, 1981) or because the health care provider underestimates or is unaware of the risks associated with a given technology. For example, a full-body computed tomography (CT) scan has a 0.08% chance of causing cancer (Brenner & Elliston, 2004). Students must be well aware of the potential side effects of all diagnostic procedures and treatments, and must learn to convey this information in an understandable manner to the patient.

As the health care industry evolves, the demands on physicians continue to grow, and time constraints in the physicians' day are surfacing as a common and worsening problem. In order to adapt to new technologies directed at enhancing medical education, instructors must spend additional time learning the technology themselves. Extensive training may be required of physician educators to gain adequate mastery before introducing procedures to students. Many health care professionals simply do not have time to devote to this requirement. Medical education includes extensive role modeling and learning through observation. It is through this role modeling that students learn bedside manner and communication skills with both patients and other physicians (Zucker et al., 1998). The introduction of many of the technologies previously described decreases the amount of role-model learning and peer interaction. Strømø et al. (2004) showed that the introduction of computer-supported, problem-based learning made students' attitudes towards cooperation more negative, making them less likely to work together (Strømø et al., 2004). Technology that decreases role modeling could hinder the acquisition of patient-doctor interaction skills. This may be particularly significant in treating those whose linguistic abilities limit their ease of understanding physicians and procedures in the first place.

The medical field tends to attract science- and technology-directed thinkers, but these individuals may not be the best people for treating a patient in a holistic manner. In medicine, there is a need to treat the whole patient, and treatment often includes behavioral modifications. The vast majority of disease is preventable through proper diet, regular exercise, not smoking, consuming little alcohol and safe sexual practices. It seems most people are not taught good habits as children. Changing behavior to promote healthy lifestyles does prevent disease. Behavior modification requires educating patients so they are convinced they need to change bad habits. As technology becomes more advanced, the tendency is toward a deterioration in close patient-doctor relationships, which can result in lessening ability to persuade patients toward long-lasting behavioral changes. In addition, patient diversity is greater than ever, and medical students might never learn proper communication skills that allow them to influence their patients' choices. Preventive medicine is the discipline where the greatest impact can be made in improving overall health. The science-geared minds frequently attracted to medicine tend to be less interested in preventive care than in the challenges of curing disease (Clawson, 1990). To adequately apply preventive medicine, social, cultural, ethnic and behavioral factors affecting a patient's life must be weighed equally with scientific knowledge, or the desired change in patient attitude or behavior is less likely to occur. That is to say, someone with a solid grounding in patient care and communication skills but possessing minimal clinical knowledge can have a more profound impact on long-term health of a patient than a doctor with outstanding clinical knowledge who lacks interpersonal skills. As technology increasingly enters into the educational process there, is the possibility that health care providers may continue to become more science-centered and less person-centered.

Medicine is taught as both an art and a science. The science in medicine is obvious, but the importance of non-scientific factors within the field of medicine is underestimated by both medical personal and the lay person. Traditionally, doctors accept a sick person as a patient and then try to treat the medical condition. The best doctors also identify risk factors for disease and try to prevent disease. The preventive branch of medicine includes vaccination, smoking cessation programs, diabetes screening and education, weight loss promotion, nutrition counseling, regular Pap smears and exercise consultations. Communication, not technology, is required for disease prevention. Medical technology becomes important primarily when prevention has failed. The art of medicine requires physicians to learn how to conduct an interview with an endless variety of patients in a non-judgmental format. This puts the patient at ease with him or herself and the physician. The interview must follow a well-organized format, so relevant medical information is extracted and a trusting doctor-patient relationship is forged. Technology may misguide inexperienced physicians and medical students into forgetting that it is the subjective symptoms of the patient that first and foremost lead to the correct diagnosis. Only with this basic piece of information clearly in mind is the physical exam and laboratory data of any value. The principle skill used in clinical work is problem solving. Medicine is like detective work. First, information must be obtained from the patient and his or her family. Next, that information, which may or may not be completely accurate, has to be combined with the physical exam, laboratory and radiology studies. A physician must be able to combine all the information, look for the source of the problem and determine possible treatments, then select the one most suitable for the individual patient. Technology is a master at providing useful information, but the physician must solve the problem (Clawson, 1990). With the pervasiveness of advanced technologies, there is a risk of over-reliance on computer-generated information and an under-emphasis on transmission of thinking and problem-solving skills.

An additional problem with technology is equitable application; economically underprivileged patients often have a medical disadvantage. However, there is a tendency when new technology is found to be advantageous to utilize that technology at every promising opportunity. Over time, the expense of the technology is spread out over many patients, and the price gradually decreases, making it available to more people. During the price-dropping interim, the poorer population may have less access to cutting-edge technology. The medical residents and students who train in facilities serving a lower socioeconomic population will also have less opportunity to work with high technology, resulting in a potentially poorer education.

Other problems exist associated with the integration of technology into medical education. One of these is that students are sometimes more experienced with common technologies than their teachers (Salas & Anderson, 1997). Or, sometimes educators lack enthusiasm for new technological concepts and this is inevitably conveyed to the students (Moberg & Whitcomb, 1999). Occasionally, technology projects are promoted by an individual instructor who leaves that teaching position and the technology component falls by the wayside, which indicates a lack of long-term support and organizational interest (Barnete, 1995). In addition, it is very expensive to fund medical education, and most of this expense is shouldered by the student physician. It is a problem if the cost of medical education sees a significant increase due to more technological requirements (Jones & Korn, 1997). Educating future doctors is a difficult task. Technology provides an almost limitless number of directions educator can pursue. The key is to be aware of available options and the advantages and disadvantages of each, and choose the options most likely to provide long-term educational benefit to the educator, future physicians and the great diversity of patients whose lives this physician will impact. It is inevitable that technology will play a vital role in medical education, and educators must be ready for the challenge.

FUTURE TRENDS

There is much promise on the horizon in medical technology. Technological improvements include better organization of data and improved access to information; collegial exploration of ethical, cultural and social issues; improvements in communication and consultation capabilities; an increase in online educational opportunities; and the evolution of medical technology itself.

One of the major problems in the utilization of technology in physician education is organizing the amount of available information into a usable format. The organizational systems available are lagging behind the technological advances. An expanding number of organizational systems to enhance a physician's ability to effectively utilize technology will be emerging in the near future. One method the medical community has employed to help assimilate and utilize the body of available information is through an approach called evidence-based medicine. Evidence-based medicine is relatively new and operates by collecting all the available research studies on a topic, rating them according to their quality and combining them into one easily accessible location so the work can be summarized and meaningful conclusions can be drawn from the combined findings of all the studies. This is a large change from the mentor-based system, where medical traditions were passed on, often without question, from one generation of physicians to the next. There is considerable enthusiasm for evidence-based medicine. It helps to provide optimum care for patients, and in the future is likely to become even more well known.

Future applications of medical technology contain a number of ethical issues, and the study of ethical implications is already on the rise. Ethics is generally an understudied aspect of technology, because there is little to be gained financially by research sponsors. Ethical issues will play a growing role in medical education as technology and the procedures that technology makes possible continue to evolve. Stem cell research, deoxyribonucleic acid (DNA) testing, preserving the lives of comatose patients indefinitely and replacement organs for treatment of the elderly or infirm are but a few of the issues already at hand, and the list is growing daily.

Newer, faster, smarter, better, smaller, more powerful and so forth are descriptions anticipated in the future of medical technology. In general, physicians and medical students even in the near future must be prepared to work with computers in ways we can hardly imagine today. The trend is a move toward exclusively electronic data storage and retrieval in all aspects of patient care. Basic keyboarding, once a class only for secretaries, is now one of the most important preparatory classes a student can take. Even now, it is often being taught in elementary school, where traditionally it was a high school course. Innovative technologies emerge more rapidly than an educator can master, so training must focus on basic technological skills with an accent on thinking logically and adapting creatively to new circumstances.

It is becoming apparent that the federal government may have a greater regulatory hand in technology and its application in the future. At a national level, the U.S. government is proposing standards for health information technology (HIT) and backing it up with legislative proposals. The National Health Information Incentive Act of 2005 aims to facilitate the development and adoption of national standards in health technology, provide initial financial support and ongoing reimbursement incentives for physicians in smaller practices to adopt the HIT policies, and support quality improvement activities (ACP Press release Feb 11, 2005). The National Nanotechnology Initiative (NNI) (www.nano.gov) is a federal research and development program established to coordinate the multi-agency efforts in nanoscale science, engineering and technology. The goals of the NNI are to maintain a world-class research and development program intended to realize the full potential of nanotechnology, and to facilitate the transition of new technologies into products for economic growth, jobs and other public benefit. These goals include development of educational resources, training of a skilled workforce, creation of a supporting infrastructure, development of tools to advance nanotechnology and finding ways to support responsible development of nanotechnology. Awards have been given to medical schools and research institutes to establish new research initiatives focusing on creating advanced nanotechnologies to analyze and detect arterial plaque formation on the molecular level in its early stages. Another group has been awarded grants to detect, monitor, treat and eliminate vulnerable plaques. Nanotechnology clearly has a large role in the future of medicine, and is already used for noninvasive imaging and sensing, targeted therapies and drug delivery systems, as well as for treatment of heart, lung, blood and sleep disorders.

At this point, there is limited application for nanotechnology in the training of medical students. The field is in its infancy. It is important though, that students are aware that the technology exists and to understand the tremendous potential it holds in the field of medicine.

CONCLUSION

Looking back to the earliest days of modern medical science, we can see that the progress of medicine, as it evolved from conjecture and superstition into a science of diagnosis and treatment, has been fueled by daring minds that had the courage to question, challenge, explore, postulate new ideas, try new approaches and prove to others the need for change. We have moved from an informal, poorly structured and chaotic system to one that attempts to tie the clinical practice of medicine to the scientific evidence available. Technology has stoked the current momentum and provided a fertile environment that fosters the exploration of new frontiers in health care. During the renaissance period, we ventured into the “inner sanctum” of the human body to learn its anatomy; we printed materials to distribute to others; we expanded knowledge and communicated theory; all of which obligated us to ask more questions. We have moved at the pace our technology has allowed and grappled with serious ethical dilemmas along the way. As a society, we have had to make room in our ideals to evaluate and assimilate current technologies and develop plans for future advances we can only imagine. Those who work as medical educators and health care professionals are obligated to create an ethical balance and provide safeguards for society while preserving life through available technology.

One can only wonder at the expanse of the unknown and where current technology will take us. Medical students are becoming familiar with a virtual world that makes information available immediately. Medical educators are earnestly responding by integration of evidence-based educational tools into the medical school curriculum and teaching early on the need to discern the true quality of the evidence being presented by a publication. This is a life-long skill needed by every physician. Physicians as a group have embraced the concept and practice of intranet, handheld devices, virtual teaching, microsurgery and technology-dependent modalities. Technology will continue to push the boundaries of diagnosis and treatment of diseases and, in the hands of a caring and skilled physician, can foster the treatment of people as diverse, unique individuals rather than emphasizing treatment of disease.

REFERENCES

Agency for Health Care Research and Quality. (2005). *Health expenditure statistics*. Retrieved May 26, 2005, from www.meps.ahrq.gov/papers/st61/stat61.pdf

Barnete, G. O. (1995). Information technology and medical education. *Journal of the American Medical Informatics Association*, *2*(5), 285-291.

Brenner, D. J., & Elliston, C. D. (2004). Estimated radiation risks potentially associated with full-body CT screening. *Radiology, 232*(3), 735-738.

Clawson, K. (1990). The education of the physician. *Academic Medicine*, *65*(2), 84-88.

Conn, J. (2004). The games doctors play. *Modern Healthcare, 34*(30), 32-33.

Fischer, S., Stewart, T. E., Mehta, S., Wax, R., & Lapinsky, S.E. (2003). Handheld computing in medicine. *Journal of the American Medical Informatics Association, 10*(2), 139-149.

Grutzner, P.A., & Suhm, N. (2004). Computer aided long bone fracture treatment. *Injury, 35*(1), S-A57-64.

Hancock, L. (1996). *Physicians' guide to the Internet.* Philadelphia: Lippincott-Raven.

Holden, M. K. (2005). Virtual environments for motor rehabilitation: Review. *Cyberpsychological Behavior, 8*(3), 187-211, 212-219.

Jones, R. F., & Korn, D. (1997). On the cost of educating a medical student. *Academic Medicine. 72*(3), 200-210.

Klinger, E., Bouchard, S., Legeron, P., Roy, S., Lauer, F., Chemin, I., et al. (2005). Virtual reality therapy versus cognitive behavior therapy for social phobia. *Cyberpsychological Behavior, 8*(1), 76-88.

Krijin, M., Emmelkam, P. M., Olafsson, R. P., & Biemond, R. (2004). Virtual reality exposure therapy of anxiety disorders. *Clinical Psychology Review*, *4*(3), 259-281.

Medical Readiness Trainer Team. (2000). Immersive virtual reality platform for medical training: A killer-application. *Studies in Health Technology and Informatics*, *70*, 207-213.

Moberg, T. F., & Whitcomb, M. E. (1999). Educational technology to facilitate medical students' learning: Background paper 2 of the medical school objectives project. *Academic Medicine, 74*, 1145-1150.

Nicolaou, D. D., & Davis, G. L. (2001). A distributed asynchronous resident procedure log for hand held devices. *Academic Emergency Medicine, 8*(8), 1200-1203.

Nissen, S., Abdulla, A., Khandheria, A., Kienzle, M., & Zaher, C. (2004). Working group 6: The role of technology to enhance clinical and educational efficiency. *Journal of the American College of Cardiology, 44*(2), 256-260.

Salas, A. A., & Anderson, M. B. (1997). Introducing information technologies into medical education: Activities of the AAMC. *Academic Medicine, 72*(3), 191-193.

Schroeder, S. A. (1981). Medical technology and academic medicine: The doctor producers' dilemma. *Journal of Medical Education, 56*(8), 634-639.

Strømø, H. I., Grøttum, P., & Lycke, K. H. (2004). Changes in student approaches to learning with the introduction of computer supported problem based learning. *Medical Education*, *38*(4), 390-398.

Vozenilek, J., Huff, J. S., Reznek, M., & Gordon, J. A. (2004). See one, do one, teach one: Advanced technology in medical education. *Academic Emergency Medicine*, *11*(11), 1149-1154.

Ward, J. P., Gordon, J., Field, M. J., & Lehmann, H. P. (2001). Communication and information technology in medical education. *Lancet, 357*(9258), 792-796.

Whitcomb, M. (2003). The information technology age is dawning for medical education. *Academic Medicine, 78*(3), 247-248.

Whitcomb, M. (2004). More on competency based education. *Academic Medicine, 79*(6), 493-494.

Zak, R. (2004). Continuing medical education. *Healthcare Informatics, 21*(7), 40.

Zucker, S., White, J. A., Fabri, P. J., & Khonsari, L. S. (1998). Instructional intranets in graduate medical education. *Academic Medicine*, *73*(10), 1072-1075.

This work was previously published in Technology and Diversity in Higher Education: New Challenges, edited by Y. Inoue, pp. 104-121, copyright 2007 by Information Science Publishing (an imprint of IGI Global).

Chapter 3.24
Technology in Primary and Secondary Medical Education

Sarah R. Edmonson
Baylor College of Medicine, USA

ABSTRACT

There is an increasing need for cost control and improved outcomes in both primary and secondary medical education. This chapter reviews the characteristics of medical learners and summarizes shortcomings of traditional medical training that may be amenable to technological intervention. Technologies useful for educating medical students and practicing physicians will be described along with examples pertaining to each technology. The chapter concludes with a summary of potential barriers to the adoption of medical education technology.

INTRODUCTION

Doctors spend the first years of their career completing an intensive primary medical education. Sadly, this primary medical training can become obsolete in a remarkably short time; new products and research constantly change the definition of "best medical practice." Consequently, secondary training—usually referred to as continuing medical education or CME—is also an important component of a complete medical education.

Medical education presents some unique challenges for curriculum design. Further, ample evidence suggests that the traditional medical curricula could be considerably improved upon. In developing new approaches to clinical training, medical educators draw from a rich field of theory and evidence about professional education in other arenas. These generalizations tend to be quite successful when the source material was designed for learners with similar qualities; they can be stunningly unsuccessful when the two learner populations differ. The first section in this chapter describes qualities of medical learners that may be relevant to instructional design, and the following section outlines barriers to successful learning in a traditional medical education setting.

The next section describes technologies that hold particular promise for solving specific medical education challenges. Examples of successful implementations provide concrete examples of these technologies in action. In recent years, a broad variety of technical tools has been employed in an effort to provide cheaper, more accessible, and more effective medical training. However, the application of technological tools in medical education is not without controversy, and the chapter closes with a discussion of some barriers to medical education technology implementation.

THE MEDICAL LEARNER

As noted, medical education begins with an intensive primary training and ongoing learning that continues throughout the physician's career. Medical learners at any stage share some common characteristics, but important differences between the two groups may influence which educational approach is most successful.

DEFINITIONS

Primary medical education describes the period in which the future physician builds a knowledge base *de novo* and learns to apply that knowledge effectively (D'Eon & Crawford, 2005). Learning tasks during medical school are both didactic and experiential; the student absorbs a fund of knowledge and also masters a repertoire of procedural skills. Generally, the student begins with a heavily didactic courseload, which transitions to a primarily experiential curriculum by the end of his training.

Secondary medical education refers to the ongoing efforts of a practicing physician to renew and update his or her knowledge and skills. Because medical practices change over time, most licensing organizations require doctors to complete ongoing medical education—although the content and format of this education is often not specified. Thus, the main educational tasks a practicing physician faces are the identification of knowledge deficiencies, and the selection of appropriate training to correct those deficiencies (Curran & Fleet, 2005). In response to that need, the physician will find a selection of continuing medical education programs offering a variety of educational content and structure.

The formal medical education process begins with university and/or medical school, then proceeds through a series of additional training steps, which may include internship, residency, public service, and fellowship training. At some point, a transition from "medical student" to "doctor" occurs. Studies of the medical education process suggest that the transition occurs as a dramatic step, rather than a gradual evolution (Prince, Van De Wiel, Scherpbier, Can Der Vleuten, & Boshuizen, 2000). Unfortunately, the exact timing of this transition is unclear and may differ from student to student. Thus, one of the major challenges in medical education is determining whether an educational tool should be optimized for the primary or secondary learner.

Common Characteristics

For both primary and secondary medical training, the designer can count on at least some commonality among medical learners. All are adult learners—no nation trains children to be physicians. For most, a record of academic achievement encourages robust self-efficacy for medical learning tasks. The direct applicability of the material provides an inherent motivation to learn. And, the learning content is all drawn from a finite, if large, pool of knowledge.

Considerations for Primary Medical Education

Superficially, curriculum design at the medical school level is simplified by a reasonably uniform body of learners and by a fairly universal curriculum. The core material to be mastered does not vary significantly, and standardized testing is available for measuring how effectively this material has been taught. Learners share similar educational backgrounds and a common career goal; in addition, medical students are typically people who have previously demonstrated high levels of motivation and academic skill. The students also share similar levels of ignorance at the outset. Unfortunately, a truly generalizable primary medical curriculum must account for many types of diversity, which may influence the effectiveness of a learning tool both within an institution and between institutions. In the United States, medical school is a post-secondary degree. Globally, the requirement for a separate university-level degree is often waived in lieu of a longer medical school curriculum. Thus, medical learners from different nations differ, in maturity and academic experience. The impact of these differences on their response to educational technology has not been well studied. In addition, a student's understanding of the material may be driven by cultural background. Language differences may bias how the material is received. Learning style also varies widely between individuals, an effect that is likely to impact educational effectiveness (Cook, 2005).

Considerations for Continuing Medical Education

How does continuing medical education differ from primary medical education? In many ways, the established physician's learning task is easier than that of the medical student. New knowledge can be incorporated into an existing framework and assessed based on the doctor's own experience.

The diversity among medical students increases enormously among secondary medical learners, and the task of preparing educational material for this group is correspondingly harder. Physician learning requirements will be driven not only by personality and cultural background, but also by specialty, experience levels, and the unique requirements of individual professional duties (Holm, 1998).

Motivation for learning also changes as the physician's career progresses. Primary medical education provides a feedback structure of grades, teacher assessments, and direct supervision; these mechanisms offer immediate motivation for educational performance. After graduation, the availability of this sort of feedback drops. Practicing physicians cite many reasons for seeking continuing education, including professional responsibility to their patients, curiosity, entertainment, fear of litigation, and maintenance of licensing and accreditation requirements (Dolcourt, Zuckerman, & Warner, 2006).

BARRIERS TO EFFECTIVE MEDICAL EDUCATION

In an atmosphere of increasing concerns about the quality of healthcare, medical education reform takes a prominent role (Bradley, 2006; Harden, 2005). Unfortunately, the most commonly used methods for medical education—including printed materials, lectures, clinical practice guidelines, and opinion leaders—have not been very effective for improving clinical care or patient health (Bloom, 2005; Davis, Thomson, Oxman, & Haynes, 1995). Educational reform will need to address the factors underlying the failures observed with traditional educational approaches: the body of information to be mastered is overwhelming, skill acquisition requires ethically questionable hands-on practice, the learners cannot always be assembled in a central location, and the cost of education is increasingly high.

Information Overload

The body of knowledge required to practice medicine effectively is overwhelmingly large and highly inter-related (D'Eon et al., 2005). In traditional lecture and text-based learning, material is presented in a linear fashion and segregated into discrete subject categories such as anatomy or pharmacology. This approach makes it more difficult for the medical learner to see the relationships between topics. Educational tools that facilitate mastery of large bodies of facts, or that highlight the relationships between topics, would be a welcome improvement. Additionally, if students could master the necessary clinical knowledge more efficiently, more time would be available for instruction in clinical and communication skills (Bradley, 2006).

Hands-On Experience

Medical students must also learn a variety of procedural skills. Like all such things, medical procedures are most effectively mastered with practice—exhaustive repetitions of the same process until it becomes automatic. Two factors make this element of the doctor's training problematic. The first is that the procedures will ultimately be applied to human patients and there are obvious logistical and ethical difficulties with practicing an uncertain skill on real patients. The second problem is the mastery of less-commonly needed skills. The student is in training for a limited number of years, and may not see enough patients who actually need the procedure in question.

Geographic Barriers

Physicians must practice where their patients live, and in some cases patients live a long way from any medical education facility. Often, practicing physicians in remote locations have only limited resources available for their continuing education (Curran et al., 2005). The geographic separation of medical practices and educational facilities has been another strong argument for new approaches to medical training.

While issues of distance and travel have been particularly prominent for doctors seeking continuing education, even primary teaching settings have limited access to training resources in some parts of the world. The "brain-drain" phenomenon—emigration of healthcare professionals from impoverished to wealthier health-care systems—is highly related to educational opportunity since the majority of emigrants initially relocate in order to participate in educational opportunities (Dodani & LaPorte, 2005). Educational tools that can be disseminated worldwide may reduce the rate of emigration, as well as improving training for those who choose not to leave low-resource nations. Educational technology, then, may play a vital role in achieving a globally uniform standard of medical practice.

Financial Considerations

Many innovations in medical education arise from a desire to decrease the cost of clinical training. The primary educational cost is substantial—in 2005, the average cost of 4-year medical school in the United States was estimated to be around \$140,000 for public medical schools and \$225,000 for those private schools whose curriculum is not supplemented by state and national funding (Morrison, 2005). These bills have escalated far in excess of inflation, and 80% of U.S. physicians graduate with debt between \$104,000 and \$140,000 (Morrison, 2005). Continuing medical education costs seem trivial in comparison to those of the primary training, but the expenses associated with arranging facilities, materials, faculty, and time away from a medical practice can add up—particularly for practitioners without easy access to an academic medical center. Logically, the cost of medical care reflects, in part, the cost of training the providers of medical care, so the high cost of medical education translates to higher prices for healthcare consumers as well. Consequently, cost reduction is a major goal of educational innovation at all levels.

TECHNOLOGY MODALITIES PARTICULARLY HELPFUL FOR MEDICAL EDUCATION

Following are descriptions and examples of medical education technologies. Some of these represent improvements on techniques that are already commonly used by medical learners. Others demonstrate adoption of useful educational tools from other industries. All, however, address one or more problems inherent to traditional medical education.

Automated Drilling

Medical students memorize a stunning quantity of facts and figures. Anatomy classes alone introduce hundreds of body structures, each with its own name, function, and location. Clinical terminology is often arbitrary, nonsystematic, and yet essential to the practice of medicine.

Automated tools can be enormously helpful for memorizing blocks of material. Automated flashcard programs are an excellent example of how computerized tools can improve the flexibility and efficiency of a classic learning tool. Electronic cards can be edited and reproduced very easily, offering an immediate advantage over traditional handmade or commercial paper cards.

Flashcards have some inherent flaws and an electronic flashcard tool overcomes many of these flaws. The most important flaw is that in repetitively viewing a set of cards, the student may "learn the card" —generating the answer on the card because of cues about the card itself, such as color, shape, handwriting, or order in the stack. A flashcard computer program can randomly vary card features as well as continuously reshuffling the card order so cues about the shape or color of card do not confound the mastery of the information. The program may also customize instruction to individual learning needs, and adaptively change the current card stack in response to student performance. Accomplishing such tasks with traditional flashcards might be possible, but would be time-consuming.

Games can also be helpful adjuncts to memorization. Software programs that generate crossword puzzles, word-find tasks, and other games can be applied to medical material. Both the creation and the solving of the puzzle serve as learning mechanisms (Bailey, Hsu, & DiCarlo, 1999).

Example: Flashcard Exchange

Flashcard Exchange (http://flashcardexchange.com) is an example of a drilling tool that can easily be adapted to medical learning. The site currently contains hundreds of collections of flashcards, entered by users, covering medical topics from acupuncture to neuroscience. For card review, users can randomize the cards, display them with either question or answer first, sort them according to prior performance, and alter visual features such as font, size, and text alignment. Session performance is reported in table and pie chart format. Cards can also be reviewed by playing a memory game.

In summary, computerized tools can offer additional functionality to improve student drilling on material to be memorized. These programs overcome several educational barriers: they enable mastery of larger volumes of information, are generally inexpensive, can be shared electronically with anyone in the world, and can be edited easily to reflect new understanding. Further, automating the memorization task frees up instructor time for more sophisticated material.

Computerized Tutorials

Brute memorization may help the student pass an upcoming exam, but may be an ineffective approach to long-term learning (D'Eon et al., 2005). In some cases, the task of memorization can be ameliorated by better understanding of the material. So, for example, anatomy terms may be better retained if they are presented in the context of human physiology and medical problem solving (Miller, Perrotti, Silverthorn, Dalley, & Rarey, 2002). Relationships between different components of clinical knowledge may be seen as branching or networked, rather than linear. As a result, the linear format of traditional lecture and text-based instruction cannot entirely represent the material. Further, the constructivist school of adult learning suggests these learners may master the material most efficiently when allowed to navigate the material freely, according to their own curiosity (Harden, 2005).

Fortunately, we can employ technical solutions to allow better representation of the interrelationships between medical concepts. One of the simplest approaches is to indicate, within a block of learning material, where other material relates. Hyperlinked text—text on the screen that, when clicked, connects the learner to a different page with the related information—was quickly recognized as a powerful tool for adult constructivist learning. Within a hyperlinked tutorial student may move freely between discussions or multimedia demonstrations of the anatomic, physiologic, and clinical aspects of the subject matter. New terms in the hyperlinked tutorial are presented in context, but link directly to a glossary. The insights from one domain improve the understanding and retention for the others. Hyperlinked learning materials are relatively easy to design, but are interactive and empowering for the student. As an added bonus, the computer environment easily incorporates tools to track the progress of the student through the material (Hilty et al., 2006).

Example: IVIMEDS

The goal of the International Virtual Medical School (IVIMEDS) is to provide a complete, four-year undergraduate medical education using a combination of e-learning and face-to-face learning with medical educators and institutions all over the world. Founded by a worldwide partnership of institutions, IVIMEDS proposes a 4-year medical school curriculum whose core elements include problem-based and system-based lessons based on a bank of virtual patients and an extensive database of reusable digital learning objects. The learning objects comprise the heart of the IVIMEDS philosophy. Each object consists of a small, discrete chunk of information—perhaps a short block of explanatory text, or a video clip, or an audio file—which can be combined with other objects to create a course of study. When medical knowledge changes, only those objects relating to that knowledge need to be updated. Thus, the program design includes a plan for low-impact maintenance. (Harden, 2005)

In addition, self-directed hyperlinked tutorials are quite simple to deploy to an Internet environment. The availability of such materials on the World Wide Web means that good quality and effective training can be made available globally, alleviating the struggle to provide high-quality education in remote or low-resource parts of the world.

Virtual Reality and Simulation

In addition to absorbing the knowledge base required for effective medical practice, medical students eventually apply that knowledge to real-world patients. Mastery of the facts and figures may not be adequate to ensure that the student will be able to apply the knowledge in the complex, rapidly changing environment found in the hospital or clinic (Hilty et al., 2006). Most students master this synthesis with practice; unfortunately, in order to practice, the lives of real patients are put at risk.

In addition to the practice required for intellectual synthesis, students need to master a set of manual skills. These involve physical examination skills such as using a stethoscope, and procedural skills, such as drawing blood or suturing a wound. Again, skills are only perfected through extensive practice—and how many patients wish to be the subject of a student's practice?

Not only is it ethically questionable to ask patients to submit to the inexperienced hands of an untried student, but the patient supply may also be inadequate for the student's education. For example, a student is fortunate to have a single opportunity to recognize and treat a rare syndrome such as Kawasaki's syndrome in pediatric patients.

Historically, to gain experience that couldn't be ethically carried out on humans, students turned to alternatives—dissecting cadavers to perfect surgery skills, treating animals to understand physiological responses to drugs, and interviewing actors imitating real patients to practice diagnosis and bedside manner. While helpful, all of these approaches have limitations. Actors may allow a student to listen to the heart, but not to administer electric shocks to restart it. Actors also require training, expect to be paid, and can meet with only one student at a time. Subjecting animals to procedures to which no human would consent is ethically troublesome, and furthermore animals differ from humans in physiologically important ways. And, cadavers are surprisingly difficulty to come by, can be dissected only once, and cannot demonstrate all the possible anatomical variants the student needs to learn.

Simulation tools, if integrated into the broader medical curriculum, allow the student to gain those experiences that are difficult to acquire in the real world (Hilty et al., 2006). Simulation may involve simple programs or devices, designed to demonstrate a simple concept. Or, a simulation model may be highly complex, designed to represent the interactions of as many variables as possible.

A well-known example of simulation technology is the life-sized dummies used for cardiopulmonary resuscitation (CPR) (Bradley, 2006). These simple tools allow the user to learn through trial and error how to position the body correctly to administer mouth-to-mouth breathing and chest compressions. Integrated circuitry provides warning lights or tones when the student is not performing the procedures correctly. And, the dummy can be used over and over again by one or many students—unlike actual hospital resuscitations! Higher resolution resuscitation simulations incorporate animated behaviors such as breathing, pulse, or blood pressure, responsive behaviors such as eye blinking, and feedback in the form of changes in the above behaviors as well as visual or auditory alarm systems (Bradley, 2006).

Example: "SimMan" Full-Scale Patient Simulation

The SimMan full-scale patient simulator assists students in practicing anesthesiology procedures and scenarios. Used worldwide, this adult-size simulator not only offers physical representations of the mouth and airway, but also produces breath sounds and chest wall movements, and can recognize and respond realistically to more than eighty types of injected drugs (Gabriel, 2001). Experienced anesthesiologists judged the SimMan tool to be "acceptably realistic" or "highly realistic" in fifteen of sixteen parameters (Hesselfeldt, Kristensen, & Rasmussen, 2005).

Simulations can take the form of interactive case studies, often using the hyperlinked multimedia format described above; physical models with integrated feedback like the resuscitation dummy; or high-fidelity "virtual reality" equipment which may incorporate haptic simulators, eye movement tracking, and highly complex algorithms to simulate real-world events (Bradley, 2006).

Intuitively, learning through simulation should lead to better mastery of the material and the higher the simulation fidelity the better the learning should be. A recent review of the literature on simulation suggested that, in fact, the most successful simulations were those who adopted sophisticated approaches such as feedback, variable difficulty levels, a range of scenarios, and adaptability (Issenberg, McGaghie, Petrusa, Lee Gordon, & Scalese, 2005). Research evidence supporting this intuition remains limited, however, perhaps because the technology to produce high-resolution simulation has only recently been available at a reasonable cost.

Traditional Content Delivered Remotely

Remote delivery technologies offer a chance to vastly improve the educational offerings available to physicians in remote practices. Geographic barriers are particularly formidable for practicing physicians. Medical students live near academic facilities and follow an established curriculum, but medical practices position themselves in proximity to the patients. Traditionally, community physicians receive continuing medical education in two ways: they read printed materials such a journals and textbooks, and they attend educational conferences. Unfortunately, with both of these formats, the content of the education is determined externally; the doctor receives what the editor thought was important, which may not reflect what she actually needed to learn (Curran et al., 2005; Harden, 2005).

Certainly a physician can seek out an educational conference addressing the topics of most interest. However, when the practicing doctor needs continuing education, he must locate appropriate training, schedule time to participate, and, often, travel to the training location--if, in fact, he or she finds an available program about the topic of interest at all. The productivity loss, from time spent traveling to and from the training, increases the overall cost of continuing education dramatically. Thus, remote learning—accessing medical education resources from remote locations—offers a distinct benefit to practitioners.

The simplest form of remote learning is simply broadcast of a traditional lecture session. Providers can observe the session by way of television, radio, telephone, or Internet access points, depending on the technology chosen. For those individuals who are unable to be present at the time of the actual conference, audio or video tapes of educational conference sessions can be distributed, allowing a broader audience to benefit. This mechanism is particularly helpful when the topic under discussion is new or poorly disseminated, so that very few people are qualified to teach it.

However, as previously noted, traditional lecture formats are often a less effective approach for presenting clinical knowledge; adult learners benefit from interactive settings. Interactivity can be incorporated into real-time lecture broadcasts when the technology allows two-way communication. Telephone conference calls, for example, allow the lecturer to invite comment or respond to questions from the audience. As satellite-based telephone and Internet communication become more comprehensive, more and more locations can benefit from this type of offering.

Remote access to traditional lecture and classroom discussion resources does expand the practicing physician's educational options, but does not address any of the inherent flaws in traditional medical teaching. Fortunately, continuing medical education learners may find it easier to follow and retain lecture material, because they often already know part of the material. Their broader knowledge base also makes it easier for them to recognize how new material relates to the things they already know.

Web-Based Learning

Web-based education is increasingly available and popular for continuing medical education (Harden, 2005). Hundreds of sites offer free or low-price "e-learning" about a broad variety of topics. Additional sites list and catalog online continuing medical education resources, so that clinicians can easily find a program that fits their needs.

At minimum, the Internet serves as a useful medium for distributing print materials or recorded lectures to a wide audience. At this level, Internet-based learning already offers obvious advantages. Large numbers of people, in a variety of places, can partake at trivial additional cost. And, learners can access the material when it is convenient for them, allowing them to pursue instruction without disrupting clinical care (Curran et al., 2005).

The best of the Web-based education offerings are pedagogically impressive, offering far more value than mere availability and time efficiency. They present content via multiple media, including text, pictures, audio, and movie clips. They offer interactivity—mechanisms such as self-test questions increase the learner's motivation to pay close attention to the material, and have been associated with improved efficacy in educational technology (Harden, 2005). They can monitor learner progress, and direct the learner to material that hasn't yet been viewed.

Just-in-Time Learning

Theories on adult learning all agree that motivation is highly correlated with successful learning. It makes sense that students would make the strongest effort to learn when they know the material will be immediately valuable to them.

For clinicians, the point at which knowledge is most desirable may be in the clinic at the moment when they realize they don't know something relevant to the patient in front of them. Estimates vary on how often doctors have unanswered questions while caring for patients, but all agree it happens daily and probably multiple times every day (Curran et al., 2005). Thus, integration of learning resources into the clinical workflow—"just-in-time" learning—appears to be an excellent opportunity.

Just-in-time learning can take two forms. The first is a passive availability of knowledge resources in the clinical setting so that the physician can access information as needed. Often passive just-in-time resources are computer-based; the electronic format allows storage of many resources in a confined space, and these resources can be indexed and cross-referenced to speed up query response time. The effectiveness of passive just-in-time resources, however, seem to be limited by the willingness of clinicians to stop and check the resources in the midst of clinical care (Butzlaff et al., 2004).

The second type of just-in-time learning attempts to more actively offer information that appears to be needed during the course of clinical care. So, for example, if an abnormal lab value is typically associated with a diagnosis, the system might offer information about that diagnosis when the result is viewed. The implementation of active just-in-time learning requires an extensive infrastructure, including a structured database of patient information, a decision logic about when to offer information, and a forum for offering instruction that the physician will see at the appropriate time (Hilty et al., 2006). Such systems are usually associated with existing clinical systems including online lab reporting, computerized order entry, or fully implemented computerized medical records.

The same programming used to offer just-in-time learning can also be employed for monitoring physician performance (Holm, 1998). If a review of clinical behavior reveals that a physician is performing incorrectly in some area, an offer for specific education coupled with a status report to the physician may increase learning motivation.

BARRIERS TO THE USE OF TECHNOLOGY-BASED MEDICAL EDUCATION

The technological innovations previously described have the potential to address existing shortcomings in both primary and secondary medical education. Why, then, haven't all medical schools replaced their existing teaching methods with high-tech tools? Barriers to the use of medical education technology include concerns about their efficacy for long-term learning; struggles with development costs; doubts about equipment reliability; and lack of academic recognition for faculty who develop educational technology.

Efficacy

The role that technological innovation may play in medical education is intellectually compelling. However, the effort is not worthwhile unless the innovations are shown to improve educational outcomes (Hilty et al., 2006). Unfortunately, when new educational tools are reported on in the literature, efficacy is not well tested and variable results are seen. In part, this is because it is difficult to design good studies in this arena (D'Eon et al., 2005). Even within medicine, educational experiences are highly individual with respect to medium, setting, learner profiles, instructional approach, and presentation of the material. Consequently, comparative studies are overwhelmed by potentially confounding factors and variable interactions (Cook, 2005).

Kirkpatrick (1994) outlined four potential levels of evaluation for training programs: user satisfaction, learning outcomes, performance improvement, and the result of the training on the process outcome. Literature supporting user satisfaction with technological innovations in medical education is plentiful. The data supporting learning outcomes is less impressive; while many studies show that the students learn from these modalities (Curran et al., 2005), few of them report findings in comparison to traditional education. In at least some studies, short- and long-term behavior changes are seen with the use of educational technology; again, these studies do not compare the effect to that seen with more traditional approaches.

Published studies demonstrating improved clinical outcomes as a result of training are difficult to find for any area of medical education. However, only educational developers from academic settings typically report their findings in the peer-reviewed professional literature. Commercial developers may be satisfied with internal technical reports to be shared with users and clients. Outcomes data that is reported to accrediting agencies also often does not appear in the published literature. Thus, substantial unrecognized outcomes data may be found in the "grey literature" (Curran et al., 2005).

Automated education systems have potential shortcomings that may undermine their effectiveness, particularly in at the performance improvement and process outcome levels. For example, one of the finer points of medical reasoning is recognizing when a case does not fit the common condition. There is a danger that students exposed to pre-programmed educational tools will come to believe that there is always one right answer to every scenario. It is also difficult to program automated education to teach and test at the conceptual level rather than simply presenting a set of facts and figures. Particularly when both teaching and testing are automated, it may be difficult to identify students who have mastered the facts without understanding the big picture, or to identify students with subtle misunderstandings. Finally, the role of personal factors in the effectiveness of technology-based education remains unclear. Students differ in their prior exposure to technology, their preferred learning styles, and their cultural viewpoints. Any of these factors could cause a student to find diminished value in a computer-based educational intervention.

Some relevant outcomes associated with new educational interventions are not assessed by Kirkpatrick's schema. For example, many studies document the usage of a resource (Curran et al., 2005). Intuitively, if a new form of education matches the efficacy of traditional forms, but is used more frequently, better learning should result.

Cost

The cost of developing high-quality computer-based educational interventions can be substantial (Hilty et al., 2006). Even a simple automated flash card program requires an initial investment to create the utility and enter the learning material. Each modification to improve learning efficacy raises the complexity of the program, with corresponding impact on set-up cost. The specialized equipment needed to deploy educational technology can also be very costly. Simulations and virtual reality programs are particularly expensive to produce. Even the classic resuscitation dummy is a fairly expensive specialty item.

The cost of developing educational technology does not stop when the program is deployed, either. Medicine is constantly changing; new knowledge is introduced and old knowledge becomes obsolete. Traditional lecturers update their notes; textbook publishers produce new editions; likewise, developers of educational technology must update their materials to reflect the state of the art.

Further, many physicians are not comfortable using computers. A lack of familiarity with the equipment is a commonly cited reason for why physicians do not participate in Web-based continuing medical education, for example (Curran et al., 2005). Poor technical skills may also deter medical faculty from choosing high-tech forums for training; often the students are more familiar with technology than are their teachers (Hilty et al., 2006). Training programs, perhaps associated with traditional CME programs, may ameliorate this barrier but add another cost to the adoption of technology-based educational tools.

Reliability

As noted, specialized equipment is necessary for both the creation and the deployment of technology tools. Virtually every application demands that the student have access to a personal computer. Often, good quality recording equipment is needed to capture audio or visual media. Web-based materials must be designed, programmed, and maintained on an active server. Even when educational technology is affordable and demonstrated to be effective, potential users may be deterred by concerns about the reliability of the medium.

Because these tools depend on their specialized equipment to run, technology failures such as a power outage or computer virus may completely stall the educational effort.

Professionalism and Ethics

The tools described in this chapter all, to some extent, replace teaching by human faculty. It is important to remember that the practice of medicine involves more than application of facts and mastery of procedures. In a traditional medical education setting, the ongoing contact between learners and faculty leads to lessons in ethical conduct, bedside manner, and professional values (Mel-B, 2002) —all things, even well written computer programs, are unlikely to supply.

Faculty Incentives

The adoption of educational technology may also be stalled by a lack of well-developed material and a lack of incentive to develop such material. Unfortunately, development of educational technology doesn't fit well into the academic productivity paradigm. Faculty who invest time in writing a book chapter or presenting a lecture are rewarded with academic recognition; a teacher who devotes energy to the development of pedagogically strong automated learning tools often receives no recognition for the effort.

CONCLUSION

Traditional medical education has many shortcomings, both at the level of the primary medical learner in medical school, and at the level of the practicing physician seeking continuing medical education. In particular, medical educators struggle with the high cost of medical education, the difficulty of delivering quality education to remote locations, the challenge of a large and continuously changing body of medical knowledge, and the need to provide extensive hands-on experience to medical learners.

Technological innovations that have the potential to address those shortcomings include programs for automated drilling like computerized flashcards, self-directed tutorials utilizing hyperlinks to establish the relationships between ideas, virtual reality and simulation tools, and remote learning mechanisms such as teleconferencing, Web-based learning, and just-in-time learning technologies.

While these educational technologies show promise for improving the overall quality of medical education, their efficacy is not yet firmly established. In addition, the cost of developing and deploying such tools has proved a barrier to adoption, as has concern about their reliability. When these barriers are overcome, medical learners will have access to a much broader set of tools for both primary and continuing medical education.

REFERENCES

Bailey, C. M., Hsu, C. T., & DiCarlo, S. E. (1999). Educational puzzles for understanding gastrointestinal physiology. *American Journal Physiology, 276*(6 Pt 2), S1-18.

Bloom, B. S. (2005). Effects of continuing medical education on improving physician clinical care and patient health: A review of systematic reviews. *International Journal Technology Assessment in Health Care, 21*(3), 380-385.

Bradley, P. (2006). The history of simulation in medical education and possible future directions. *Medical Education, 40*(3), 254-262.

Butzlaff, M., Vollmar, H. C., Floer, B., Koneczny, N., Isfort, J., & Lange, S. (2004). Learning with computerized guidelines in general practice? A randomized controlled trial. *Family Practice, 21*(2), 183-188.

Cook, D. A. (2005). The research we still are not doing: An agenda for the study of computer-based learning. *Academic Medicine*, *80*(6), 531-538.

Curran, V. R., & Fleet, L. (2005). A review of evaluation outcomes of Web-based continuing medical education. *Medical Education, 39*(6), 561-567.

D'Eon, M., & Crawford, R. (2005). The elusive content of the medical-school curriculum: A method to the madness. *Medical Teacher, 27*(8), 699-703.

Davis, D. A., Thomson, M. A., Oxman, A. D., & Haynes, R. B. (1995). Changing physician performance. A systematic review of the effect of continuing medical education strategies. *Journal of the American Medical Association, 274*(9), 700-705.

Dodani, S., & LaPorte, R. E. (2005). Brain drain from developing countries: how can brain drain be converted into wisdom gain? *Journal of the Royal Society of Medicine, 98*(11), 487-491.

Dolcourt, J. L., Zuckerman, G., & Warner, K. (2006). Learners' decisions for attending Pediatric Grand Rounds: A qualitative and quantitative study. *BioMed Central Medical Education*, *6*, 26.

Gabriel, B. A. (2001). Replicating the "real thing": How virtual reality is transforming medical education. *American Association of Medical Colleges Reporter Online*, *10*(12).

Harden, R. M. (2005). A new vision for distance learning and continuing medical education. *Journal of Continuing Education in the Health Professions, 25*(1), 43-51.

Hesselfeldt, R., Kristensen, M. S., & Rasmussen, L. S. (2005). Evaluation of the airway of the SimMan full-scale patient simulator. *Acta Anaesthesiologica Scandinavica, 49*(9), 1339-1345.

Hilty, D. M., Hales, D. J., Briscoe, G., Benjamin, S., Boland, R. J., Luo, J. S., Chan, C. H., Kennedy, R. S., Karlinsky, H., Gordon, D. B., Yager, J., & Yellowlees, P. M. (2006). APA Summit on Medical Student Education Task Force on Informatics and Technology: learning about computers and applying computer technology to education and practice. *Acad Psychiatry, 30*(1), 29-35.

Holm, H. A. (1998). Quality issues in continuing medical education. *Bmj, 316*(7131), 621-624.

Issenberg, S. B., McGaghie, W. C., Petrusa, E. R., Lee Gordon, D., & Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: A BEME systematic review. *Med Teach, 27*(1), 10-28.

Kirkpatrick, D. (1994). *Evaluating training programs: The four levels.* San Francisco: Berrett-Koehler.

Mel-B, K. A. (2002). What is happening to bedside clinical teaching? *Med Educ, 36*(12), 1185-1188.

Miller, S. A., Perrotti, W., Silverthorn, D. U., Dalley, A. F., & Rarey, K. E. (2002). From college to clinic: Reasoning over memorization is key for understanding anatomy. *Anat Rec, 269*(2), 69-80.

Morrison, G. (2005). Mortgaging our future--the cost of medical education. *N Engl J Med, 352*(2), 117-119.

Prince, K. J., Van De Wiel, M., Scherpbier, A. J., Can Der Vleuten, C. P., & Boshuizen, H. P. (2000). A qualitative analysis of the transition from theory to practice in undergraduate training in a PBL-Medical School. *Adv Health Sci Educ Theory Pract, 5*(2), 105-116.

KEY TERMS

Primary medical education: Describes the period in which the future physician builds a knowledge base de novo and learns to apply that knowledge effectively.

Secondary medical education: Refers to the ongoing efforts of a practicing physician to renew and update her knowledge and skills.

Information overload: Refers to the phenomenon in which the body of information to be mastered is too large to be assimilated by the student.

Procedural Skills: Are those physician skills that require hands-on manipulation of the patient, such as physical examination, obstetrical delivery, or surgery.

The "Brain-Drain" Phenomenon: Refers to emigration of healthcare professionals from impoverished to wealthier health-care systems.

Automated Flashcard Programs: Are computer programs that allow the user to quiz himself about terms and their definitions, onscreen. They are an excellent example of how computerized tools can improve the flexibility and efficiency of a classic learning tool.

The Constructivist School of Adult Learning: Suggests these learners may master the material most efficiently when allowed to navigate the material freely, according to their own curiosity.

Hyperlinked Text: Text on the screen that, when clicked, connects the learner to a different page with the related information.

Simulation Tools: Learning tools in which a real-life situation is simulated using models or interactive computer programs.

Just-In-Time Learning: Learning tools that attempt to offer knowledge at the point when the physician would most need it, often in the course of clinical care.

Kirkpatrick's Levels of Evaluation: Four potential levels of evaluation for training programs: user satisfaction, learning outcomes, performance improvement, and the result of the training on the process outcome.

This work was previously published in Handbook of Research on Instructional Systems and Technology, edited by T. T. Kidd and H. Song, pp. 312-325, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.25
Care2x in Medical Informatics Education

Andreas Holzinger
Medical University Graz (MUG), Austria

Harald Burgsteiner
Graz University of Applied Sciences, Austria

Helfrid Maresch
Graz University of Applied Sciences, Austria

ABSTRACT

In this chapter the authors report about their experiences in education of both students of healthcare engineering at Graz University of Applied Sciences, and students of medicine at the Medical University Graz, gained during the winter term 2004. Care2x is an open source Web-based integrated healthcare environment (IHE). It allows the integration of data, information, functions, and workflows in one environment. The system is currently consisting of four major components, which can also function independently: hospital information system (HIS), practice management (PM), a central data server (CDS) and a health exchange protocol (HXP). Although the components are under heavy development, the HIS has reached a degree of stability, where one can use it at least for educational purposes. Various groups also report the usage of enhanced versions of Care2x in real life settings. Our experiences in both—very different—student groups have been very promising. In both groups the acceptance was high and Care2x provided good insights into the principles of a hospital information system. The medical students learned the principal handling of a HIS, whereas the engineering students had the possibility to go deeper into technical details.

INTRODUCTION

In this chapter, the authors report about their experiences in the education of students of healthcare engineering (HCE) at Graz University of Applied Sciences, and students of medicine at the Medical University Graz, gained during the winter term of 2004. Care2x is an open-source Web-based integrated healthcare environment (IHE). It allows the integration of data, information, functions, and work flows in one environment. The system currently consists of four major components, which can also function independently: the hospital information system (HIS), practice management (PM), a central data server (CDS), and a health exchange protocol (HXP). Although the components are under heavy development, the HIS has reached a degree of stability so that one can use it at least for educational purposes. Various groups also report the usage of enhanced versions of Care2x in real-life settings. Our experiences with both—very different—student groups have been very promising. In both groups, the acceptance was high and Care2x provided good insights into the principles of a hospital information system. The medical students learned the principal handling of an HIS, whereas the engineering students had the possibility to go deeper into technical details.

How to prepare both medical and engineering students in the best possible way for their later work with modern HISs is a common question. Whereas students of engineering are rather enthusiastic about IT, students of medicine are skeptical in general about using it. However, HISs are not widely accepted by healthcare professionals; that is, barriers to the use of HIS are primarily sociological, cultural, and organizational rather than technological (Moore, 1996).

It seems plausible to not only give students theoretical background about the structure, functions, and common tasks of an HIS, but to also let them work with a fully functional HIS during lectures. This is essential, particularly if students are required to be able to work with possibly any HIS in practice after only a short period of vocational adjustment. However, it depends on many different factors regarding which HIS to choose. One of the most important is whether it is necessary to teach (with) a particular HIS of a certain vendor, for example, if this system is deployed in a network of local hospitals. Another key factor, especially for noncommercial educational institutions, is the economic impact of the introduction of a commercial HIS at the university. Third, for the education of students of medical informatics, it might also be reasonable to teach the process of developing (parts of) a bigger software engineering project. Hence, the need for an open-source system arises if one does not want to start the development of his or her own HIS. Although there are many more factors to consider in general, we chose Care2x as our primary educational HIS for the following reasons.

CARE2X

Care2x is a generic multilanguage, open-source project that implements a modern hospital information system (the Web page of Care2x is located at http://www.care2x.org/). The project was started in May 2002 with the release of the first beta version of Care2x by a nurse who was dissatisfied with the HIS in the hospital where he was working. As of today, the development team has grown to over 100 members from over 20 countries. Care2x is a Web-based HIS that is built upon other open-source projects: the Apache Web server from the Apache Foundation (http://www.apache.org/), the script language PHP (http://www.php.org/), and the relational database-management system (RDMS) mySQL (http://www.mysql.com/). There exist several source-code branches that try to integrate the option to choose from other RDBMSs like Oracle and postgreSQL. The latter one is already supported in the current version at the time of this writing (Deployment 2.1). For our investigations, we chose the most feature-rich version that was available from the Care2x Web page in early fall of 2004. This release had the version number 2.0.2. Some minor deficiencies that we report later may already be fixed in the current version, Deployment 2.1.

Care2x is a very feature-rich HIS that is fully configurable for any clinical structure. It is built upon different modules, which include, for example, in and outpatient administration, admission, pharmacy, radiology (including DICOM [Digital Imaging and Communications in Medicine] image uploads), laboratories, ambulatories, nursing, medics, DRGs (diagnosis-related groups), and so forth. Online help is available for some clinical paths. See Figure 1 for an example.

REVERSE ENGINEERING

The reverse engineering of existing complex software packages starting at the source-code level has a higher value for practical education than a new development. Bothe (2001) argues that groups of students will rarely be able to develop a project further than to a prototype stage during a single lecture. Access to the source code is not available for most commercial HISs, which is another advantage of using Care2x as an educational system. In our first lecture, the students of HCE were asked to test all functions and paths of Care2x. They had to set up a small virtual clinic and employ doctors, nurses, and technical stuff. Finally, patients had to be admitted, attended to, and dismissed at all stations. In a second lecture in the upcoming semester, our students have the assignment to analyze a fully functional HIS at the source-code level. Since Care2x is built upon a modular structure, small teams of programmers have tasks like finding and fixing bugs in the current version, adding simple modules for special functions not included in the official version, or implementing interfaces to other existing information systems or medical equipment. In the spirit of open-source projects, reasonable additions and modifications can and should be published to the Care2x community Web page.

LESSONS LEARNED

Approximately 100 students from medicine and 25 from HCE participated (Figure 2). The whole lecture was built in the following way:

Figure 1. Help page describing the clinical path for starting a new surgery-operation document

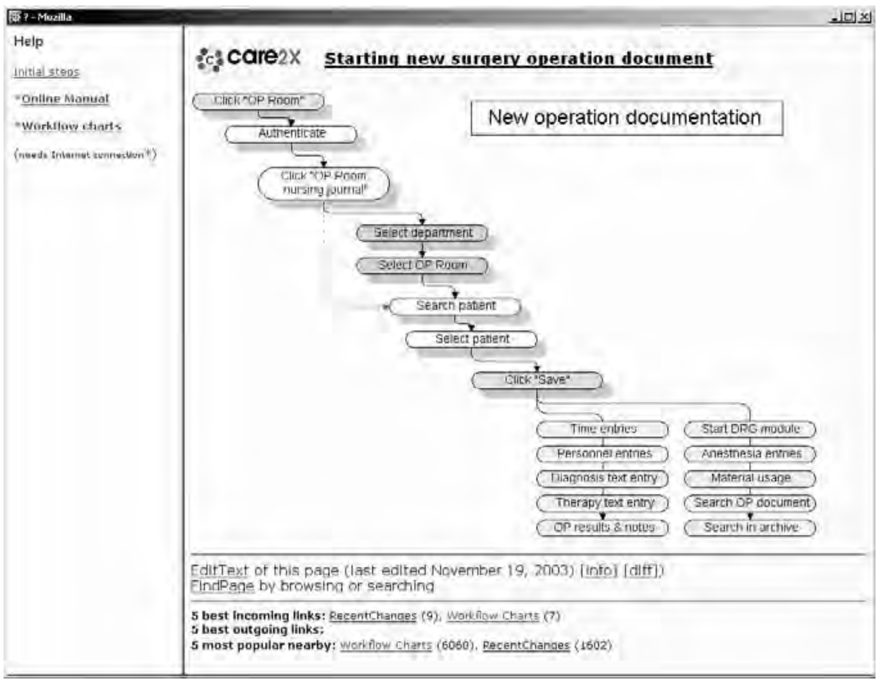


Figure 2. Students at work with Care2x (We assigned groups of 2-3 students with different tasks related to the administration of a virtual hospital)

1. Theoretical foundations of HIS in traditional lectures
2. Principles of Care2x explained (HCE group's lecture was more technology orientated)
3. Familiarization with Care2x in practical sessions
4. Practical work, specific work flows
5. Applying reverse engineering (HCE group only in the second part of the lecture)
6. Examination (both theoretical and practical)

During the education, the students were faced with the following strengths and weaknesses of Care2x.

Strengths:

- Everyone can make his or her own tools
- Work does not have to be done in a strict order
- Very flexible
- Easy to handle
- Continuing design and development
- Open source
- Lots of different languages
- Bg community that takes care of Care2x
- Easy to select the different departments and stations

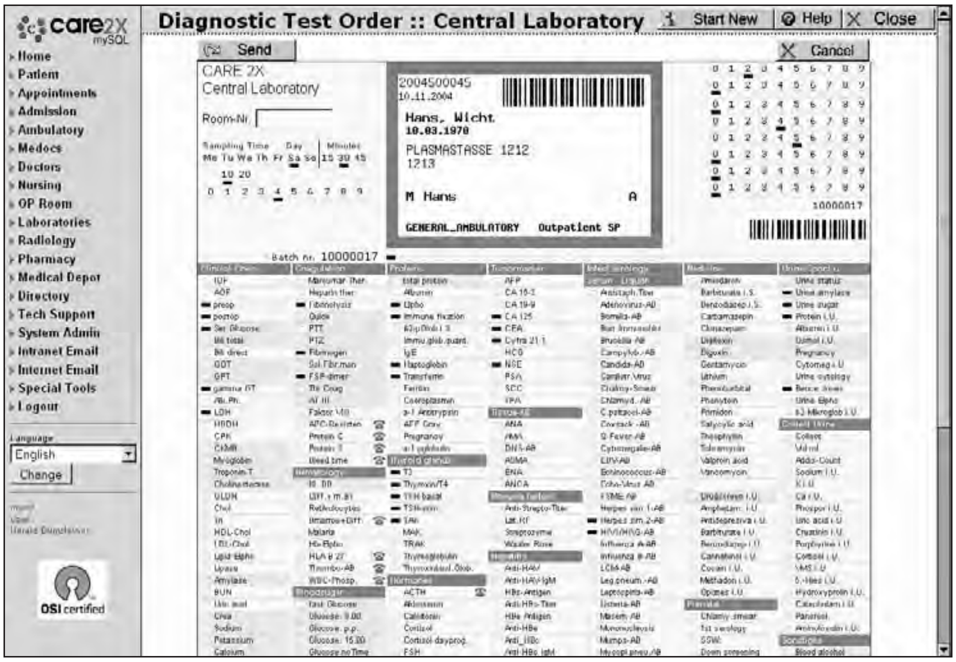
Weaknesses:

- No real standard between the modules
- Documentation is only rudimentary
- A few tools are not really easy to interpret
- Lack of security measures
- Not a state-of-the-art user interface
- There is no global list of patients from which to select one

OBSTACLES IDENTIFIED

During our lectures and trainings, there emerged several problems while using Care2x. There are

Figure 3. Example of a graphically embedded complex form, the diagnostic test order



a lot of small bugs that caused troubles. The biggest problem was that sometimes the browser responded with an "inactivity error" and the session would time out. Most of the time this error message was shown, the last click had not been made but one single minute ago. The next problem with the handling was that sometimes the back button on the Web browser would lead to nowhere because Care2x does not manage this. Much later, we found out that the back button of the browser is unnecessary because the program has included this function. That did not solve the problem completely: Every now and then the integrated back button of Care2x led to nowhere, too. In addition, some pages did not include the Care2x back button (inconsistency), resulting in a blank page. This required the user to restart at the very beginning and click through all the menus once again, which was boring for the students.

The general software problems that did not concern the running process were not severe. However, there is a serious problem when it is possible to admit a patient to more than one station, or when it is possible to alter a patient's record after his or her death.

A severe problem that has to be solved is that patients have to be discharged and then hospitalized again when we just want them to be transferred from the ambulatory to a station.

There are some translation errors and missing notes. For example, if a new patient record is being applied, there are red stars above some properties. Although this is an obvious sign for experienced users, it is not noted anywhere why these stars appear. The students found out that these stars show the minimum amount of data that is required to create a patient record, but how would the students of medicine with little experience in IT know this fact?

It is also sometimes annoying that bits of information are hidden behind a link. For example, if you want to hospitalize a patient, you have to remember the social insurance number because it is not shown in the place it is needed. This is due to the fact that Care2x works with only one window. Sometimes there might just be too little space to provide all the information needed, and then the user has to write this information down or remember it; this cannot be the aim of an HIS.

CONCLUSION

Care2x is flexible open-source software. Although there are some bugs, it has the potential to become functional software to support work flows within a (real) hospital. We think the biggest problems are the documentation and the deduction of treatments. Working with Care2x as a beginner is not very comfortable, and the software is not very intuitive. However, if one trains with Care2x, the work flows become clearer and more logical. The online help of Care2x should be better and more comprehensive. Working with the software was very fun because you really can play with a virtual hospital. Care2x is a very good possibility for training with work flows in a hospital. Further improvement of Care2x will open new areas to work with this software.

REFERENCES

Alpay, L., & Murray, P. (1998). Challenges for delivering healthcare education through telematics. *International Journal of Medical Informatics, 50*(1-3), 267-271.

Bemmel, J. H. v., & Musen, M. A. (2000). *Handbook of medical informatics*. Berlin, Germany: Springer.

Bothe, K. (2001). Reverse engineering: The challenge of large-scale real-world educational projects. *14th Conference on Software Engineering Education and Training* (p. 115).

Fieschi, M. (2002). Information technology is changing the way society sees health care delivery. *International Journal of Medical Informatics, 66*(1-3), 85-93.

Haux, R. (1998). Health and medical informatics education: Perspectives for the next decade. *International Journal of Medical Informatics, 50*(1-3), 7-19.

Haux, R. (2002). Health care in the information society: What should be the role of medical informatics? *Methods of Information in Medicine, 41*(1), 31-35.

Haux, R., Hasman, A., Leven, F. J., Protti, D. J., & Musen, M. A. (1997). Education and training in medical informatics. In J. Bemmel, J. V. Bemmel, & M. A. Musen (Eds.), *Handbook of medical informatics* (p. 537ff). Heidelberg, Germany: Springer.

Haux, R., & Knaup, P. (2000). Recommendations of the International Medical Informatics Association (IMIA) on education in health and medical informatics. *Methods of Information in Medicine, 39*(3), 267-277.

Hovenga, E. J. S. (2004). Globalisation of health and medical informatics education: What are the issues? *International Journal of Medical Informatics, 73*(2), 101-109.

Moore, M. B. (1996). Acceptance of information technology by health care professionals. *Symposium on Computers and the Quality of Life* (pp. 57-60).

KEY TERMS

Care2x: An open-source HIS available from http://www.care2x.org/. Care2x is a quite mature and stable product that can be used at least for educational purposes for both students of medicine and students of medical informatics. Some groups report the deployment of enhanced and adopted versions in real hospitals.

Diagnosis-Related Group (DRG): The DRG system is an inpatient classification system based on several factors: the principal diagnosis, secondary diagnosis, surgical factors, age, sex, and discharge status. Under the Medicare prospective payment system, hospitals are paid a set fee for treating patients in a single DRG category, regardless of the actual cost of care for the individual.

Digital Imaging and Communications in Medicine (DICOM): The DICOM image format is commonly used for the transfer and storage of medical images. Visit Chris Rorden's DICOM page for information about the format and free software to view and manipulate it.

Hospital Information System (HIS): It is the central medical information system in most hospitals in which most healthcare-related data (e.g., personnel, stations, patients and their medical history, etc.) are stored.

Medical Informatics: The rapidly developing scientific field that deals with biomedical information, data, and knowledge: their storage, retrieval, and optimal use for problem solving and decision making. The emergence of this new discipline has been attributed to advances in computing and communications technology, to an increasing awareness that the knowledge base of medicine is essentially unmanageable by traditional paper-based methods, and to a growing conviction that the process of informed decision making is as important to modern biomedicine as is the collection of facts on which clinical decisions or research plans are made (Shortliffe, 1995).

Open Source: The idea of sharing the source code of applications or tools for free. Other people are invited to elaborate on future extensions and improvements. Most open-source projects are committed to one of the Gnu public licenses (see http://www.gnu.org/licenses/licenses.html).

RDBMS (Relational Database Management System): A software package that manages a relational database, optimized for the rapid and flexible retrieval of data. It is also called a database engine.

Reverse Engineering: Taking apart an existing system to analyze smaller or single parts. The reduced complexity simplifies the process of enhancing or understanding its functions.

This work was previously published in the Handbook of Research on Informatics in Healthcare and Biomedicine, edited by A. A. Lazakidou, pp. 81-88, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.26
ePortfolios in Graduate Medical Education

Jorge G. Ruiz
University of Miami, VA GRECC, and Stein Gerontological Institute, USA

Maria H. van Zuilen
University of Miami, USA

Alan Katz
University of Miami, USA

Marcos Milanez
University of Miami, USA

Richard G. Tiberius
University of Miami, USA

ABSTRACT

Residency education is the period of clinical education that follows graduation from medical school, and prepares physicians for the independent practice of medicine. The Accreditation Council for Graduate Medical Education (ACGME) is an organization responsible for accrediting residency education programs. The ACGME is increasingly emphasizing educational outcomes in the accreditation process. The authors will discuss the experience of GME programs using ePortfolios for both formative and summative evaluation of residents and the integration of ePortfolios as part of institutions' learning management systems. ePortfolios can be especially useful for evaluating and documenting mastery of educational outcomes such as practice-based improvement, use of scientific evidence in patient care, and professional and ethical behaviors that are difficult to evaluate using traditional assessment instruments. The authors also review the literature describing the use of ePortfolios as a tool that is both powerful and reflective, for the assessment of program outcomes by administrators and faculty.

BACKGROUND

Residency education or graduate medical education (GME) is the period of clinical education that follows graduation from medical school, and prepares physicians for the independent practice of medicine. Depending on the specialty, resident physicians require between three to seven years of full-time experience in a training program to graduate as qualified specialists ready to practice. Resident physicians care for patients under the direct supervision of teaching physicians. The clinical experiences occur in a range of venues from community settings and outpatient practices to institutional environments such as hospitals and long-term care facilities. These clinical experiences are integrated into a comprehensive educational program that includes didactic activities and research.

Keeping track of residents' progress and assuring that all residents acquire the necessary knowledge, skills, and attitudes to become competent physicians in their areas of specialty can be a challenge given this variety of training experiences. Increasingly, accrediting agencies are holding training programs accountable for documenting outcomes. Traditionally, the Accreditation Council for Graduate Medical Education (ACGME) has focused on evaluating the adequacy of the process or structure of the residency programs to educate residents. In 1999, the ACGME introduced a new paradigm, the Outcomes Project, which places greater emphasis on a program's actual accomplishments through an assessment of program outcomes (ACGME, 2004).

In order to accomplish this goal, the ACGME has outlined six competencies: patient care, medical knowledge, professionalism, interpersonal and communication skills, practice-based learning and improvement, and systems-based practice. The rationale for this emphasis on outcomes is the need to ensure that physicians become and remain competent to meet the health care needs of their communities. At the end of their training, physicians must develop competence in lifelong learning strategies, reflective clinical practice, skills, and appropriate attitudes. Achieving these outcomes and documenting the achievement presents challenges for planners of postgraduate teaching, learning, and assessment. Medical educators and trainees must meet these new challenges in the face of dramatic changes in the U.S. health care system. Mounting clinical and academic activities due to changes in health care delivery and advances in medicine have increased demands on academic faculty, resulting in less time for teaching and mentoring (Ozuah, 2002).

To be able to assess this expanded range of competencies, training programs must redefine their current assessment approaches. Graduate medical education programs need to move from an almost exclusive reliance on traditional evaluations such as global subjective ratings of performance and written examinations, towards a competency-based model that requires multidimensional evaluations such as objective structured clinical examinations (OSCEs), standardized patient exams, chart reviews, and peer and patient evaluations. Even these additional assessment methodologies will not enable us to effectively evaluate all of the competencies. The assessment methodologies appear to be more effective in the evaluation of patient care, knowledge, and communication than they are in the evaluation of competencies such as practice-based learning and improvement (Lynch, Swing, Horowitz, Holt, & Messer, 2004), systems-based practice (Ziegelstein & Fiebach, 2004), and professionalism. There is a need for new tools with which to conduct valid and accurate assessments of these competencies. Moreover, these new tools should be compatible with the learner-centered model that emphasizes self-reflection and self-directed learning, critical skills that put learners in control of their own learning.

PORTFOLIOS IN MEDICAL EDUCATION

Portfolios document the evidence that demonstrates a doctor's education and practice achievements (Wilkinson et al., 2002). Medical educators have used portfolios at the undergraduate, graduate, and continuing medical education levels across various medical specialties such as psychiatry (Jarvis, O'Sullivan, McClain, & Clardy, 2004), obstetrics and gynecology (Lonka et al., 2001), internal medicine (Hayden, Dufel, & Shih, 2002), pediatrics (Carraccio & Englander, 2004), and emergency medicine (O'Sullivan & Greene, 2002). Medical education experts advocate the use of portfolios at each level of training to aid these specialties in the evaluation of competencies set forth by accrediting agencies such as the ACGME. Some portfolios focus primarily on assessment and reflect a learner's achievement of competency in specific areas of medicine. These types of portfolios are often used for summative evaluation purposes. Others are more developmental in nature and present evidence of improvements in learning over time. Such portfolios can be used by program directors and by the learners themselves for formative purposes to track progress. Yet other portfolios might showcase a person's best work or achievements in specific areas such as clinical work, research, or education. Most portfolios contain a combination of elements. The table lists some components that have been included in medical portfolios (Carraccio & Englander, 2004; Hays, 2004; Ozuah, 2002).

Portfolios can be tailored to the medical trainee, and the content can be determined by the trainee's learning needs. They can accommodate a diversity of practice and academic evidence. They foster self-directed learning, lifelong learning, critical thinking, and self-reflection (Carraccio & Englander, 2004; Challis, 2001; Hayden et al., 2002; Lynch et al., 2004; Stanton & Grant, 1999).

There are some disadvantages to the use of portfolios. Portfolios are time consuming to complete and assess, and immediate documentation of an experience is not always practical, given competing clinical obligations. Portfolios sometimes lack structure and standardization, making their use in evaluation difficult. Finally, portfolios do not really tell us much about the actual level of performance in many clinical activities (Challis, 2001; Snadden, 1999; Stanton & Grant, 1999).

RATIONALE FOR THE USE OF ELECTRONIC PORTFOLIOS

Electronic portfolios (ePortfolios) can circumvent some of the difficulties with hardcopy portfolios. Portfolios in their traditional written form are difficult to update, store, search, access, and distribute. These problems are compounded by the rotation of residents through different clinical venues. ePortfolios enable program administrators to gather and keep track of all the portfolio components that need to be submitted by residents, supervising faculty, and other contributors, especially when these persons are geographically dispersed. ePortfolios give residents access to their portfolio at all times, so they can document their clinical training in a timely fashion and allow their physician mentors and colleagues to review their progress.

Most ePortfolios are based on Internet technologies that offer multiple useful features such as accessibility, easy updating, learner control, distribution, standardization, tracking, and monitoring. Accessibility refers to the trainee's ability to find in the ePortfolio what is needed, when it is needed. Improved access to ePortfolio content is crucial, as learning is often an unplanned experience. Updating electronic content is easier than updating printed material (Chu & Chan, 1998); ePortfolios allow trainees to revise their content simply and quickly. Learners have control over the content, learning sequence, pace of learning, time, and often media, thereby allowing them to tailor the ePortfolio to meet their own personal learning objectives (Chodorow, 1996). Internet technologies allow the widespread distribution of ePortfolio content to many users simultaneously, anytime and anywhere (Chu & Chan, 1998). An additional strength of ePortfolios is that they may standardize content and delivery. Automated tracking and reporting of a trainee's activity lessens the faculty administrative burden. Moreover, ePortfolios can be designed to include learner assessments to determine whether learning has occurred. Since documentation of outcomes is the new ACGME mandate for residency programs, ePortfolios will likely grow in popularity.

There are some downsides to the use of ePortfolios in medical education, including issues related to computer server space, technical support and maintenance, use of portfolio software, and security and confidentiality. Medical educators considering the use of ePortfolios need to work with their local information technology staff to identify whether the level of expertise is available to develop the electronic platform and have access to dedicated support. There are a variety of commercial products available, but these are generally limited in their ability to meet the specific needs of medical education programs and may not fulfill the needs of learners, faculty, and program administration.

ePORTFOLIOS IN GRADUATE MEDICAL EDUCATION

ePortfolios can be especially helpful in documenting the ACGME competencies for graduate medical education. The ACGME has suggested methods for the evaluation of each competency (ACGME, 2004). Data from these evaluation methods can be effectively captured in the ePortfolio. For example, discipline-specific medical knowledge can be documented in the ePortfolio by including oral, written, or MCQ test results. Most of these tests can easily be administered electronically, which greatly facilitates the inclusion of results in the ePortfolio. The patient care competency can be documented through a variety of information sources, including record reviews, chart-stimulated recall exercises, patient surveys, and results from performance-based competency assessments such as standardized patient exams and OSCEs. The interpersonal and communication skills competency can also be effectively assessed through patient surveys and the performance-based assessments. Some of this data can be gathered electronically. For the other elements it is relatively easy to enter the results in an online template.

ePortfolios offer particular benefits for the remaining competencies, which are more difficult to assess. Aside from the evaluation methods proposed by the ACGME, medical educators can experiment with a wide array of innovative evaluation methods to assess these competencies (Carraccio & Englander, 2004). The practice-based learning and improvement competency requires residents to be able to investigate and evaluate their own patient care practices, appraise and assimilate evidence-based scientific evidence, and consequently improve their patient care. ePortfolios provide an accessible platform where residents can document evidence of their patient care practices, allowing them to reflect upon these experiences, analyze their thought processes, identify strengths and weaknesses, and under the guidance of supervising faculty with access to the same ePortfolio, set up a plan for improvement of overall clinical performance. These reflective activities can also provide evidence for the professionalism competency, as it fosters residents' commitment to ethically perform their professional responsibilities to an increasingly diverse patient population. Finally, the documentation of medical errors, the critical incidents during patient care activities, and the performance of quality improvement projects serve to fulfill the systems-based practice competency where residents must demonstrate an awareness of and responsiveness to the larger health care system, and an ability to provide optimal patient care by effectively mobilizing the health care system resources (Carraccio & Englander, 2004).

There are several examples in the literature of how ePortfolios have been used in graduate medical education to address the above ACGME competencies. Fung et al. (2000) described a learner-driven ePortfolio that was successfully implemented in an obstetrics and gynecology training program. They demonstrated improved self-directed learning and lifelong learning attitudes. Since then, their ePortfolio, called KOALA, has been used successfully with other trainees. Carraccio and Englander (2004) developed a Web-based ePortfolio to evaluate pediatric residents' performance in ACGME competencies. Their ePortfolio consists of self-assessment, a learning plan, tracking features, asynchronous discussion boards, and a system to address critical incidents during pediatrics training. Chisholm and Croskerry (2004) present an ePortfolio entry that can assess the Practice-Based Learning and Improvement competency. An example is of a senior faculty member who describes a medical error that occurred, say "delay in diagnosis." The learner is encouraged to reflect on the system errors and personal errors that might have contributed to the delay in diagnosis. The faculty member can also identify a number of personal learning points. This case study can then be distributed to residents and faculty in an emergency medicine residency. This type of portfolio entry promotes introspection and self-reflection. The authors note that senior-level faculty may need to model the portfolio for novice learners such as students and residents that have little experience with this process.

ePORTFOLIOS IN A GRADUATE GERIATRICS EDUCATION PROGRAM

The Geriatrics Fellowship at the University of Miami Miller School of Medicine is a one- to two-year training program following residency. Currently, nine residents (called fellows at this level) are enrolled in training. The fellowship includes several affiliated institutions with diverse training venues through which the fellows rotate including the nursing home, acute and sub-acute care settings, hospice, geriatric evaluation and management, home care, and outpatient settings. Although the training program is based in Miami, the fellows rotate at locations geographically distant from the home institution.

At the beginning of the 2004-2005 academic year, the fellowship committee decided to incorporate the ACGME recommendations, anticipating the full implementation of the outcome project in 2006. Our technical group, in conjunction with fellowship faculty and fellows, initiated a pilot ePortfolio program to capture the ACGME outcomes. To meet the challenge of reliably measuring all of the six competencies, the committee expanded the assessment of the fellows from global subjective evaluations to a more comprehensive assessment methodology based on multiple assessment tools and multiple raters. One example of a multiple assessment tool is the 360-degree evaluation in which fellows are evaluated by different members of the interdisciplinary teams they work with. A 360-degree evaluation can address aspects of several ACGME competencies, including interpersonal and communications skills, professionalism, patient care, and systems-based practice (ACGME, 2004). These evaluations are integrated with other training materials and self-reflection components as part of an electronic portfolio. Most of these evaluations were previously in written format, which complicated the ongoing evaluation of fellows. The use of an ePortfolio offers the opportunity to enhance documentation and reduce the administrative burden of dealing with nine fellows dispersed across different locations.

The Geriatrics Fellowship relies on a learning management system (LMS) called GeriU: the Online Geriatrics University (http://www.geriu.

org). GeriU is based on the Angel LMS platform, a SCORM (Sharable Content Object Reference Model) conformant LMS which is a database-driven eLearning platform. This LMS offers several collaboration features such as e-mail, chat, and discussion groups. It also serves as a repository for digital materials ranging from digitized text documents to highly interactive multimedia eLearning tutorials. The fellowship homepage was built as an ePortfolio to allow fellows to document the clinical, educational, and research activities they engage in during their fellowship.

DESCRIPTION OF THE GERIATRICS FELLOWS' ePORTFOLIO

The ePortfolio our fellows use is organized into four major sections: toolbox, clinical, education, and research, reflecting the major components of the fellowship training program. A description of each section follows:

Toolbox: This section contains guidelines for completion of the ePortfolio, pertinent articles about ePortfolios, the ACGME competencies, an overview of the assessments used to evaluate each competency, and a frequently asked questions section.

Clinical: This section contains information related to the supervised clinical activities of Geriatrics fellows. For each clinical rotation, the portfolio includes an evaluation of the fellow by the attending physician, an evaluation of the attending completed by the fellow, and a learning objective checklist that allows fellows to check whether they have completed a core set of activities for that rotation. After finishing each rotation, fellows complete a reflective exercise in which they describe some of their key learning experiences. The clinical section of the portfolio also contains a patient log, patient surveys, 360-degree evaluations, chart reviews conducted by the fellows, and an analysis of critical incidents.

Education: The education section comprises fellows' educational activities, products, and evaluations. Fellows conduct teaching and assessment sessions for medical students, residents, their peers, and sometimes other audiences. A list of educational activities conducted by fellows is included in the ePortfolio, as well as any teaching materials the fellows developed such as PowerPoint presentations, journal club reviews, and morbidity and mortality case conference summaries. The education section also contains evaluations of some of these teaching sessions from the learners.

Another aspect included in this section is a list of educational activities the fellows have participated in, as well as results from competency assessment activities. This section also contains a reflective component asking fellows to describe what they learned from each of these experiences.

Research: Each Geriatrics fellow completes a research project under the mentorship of Geriatrics faculty. This section includes the description of the research project with its goals, objectives, research plan, result, discussion, and conclusions. Any abstracts, manuscripts, or presentations resulting from this and other research in which the fellows are engaged are included in this section. Fellows' reflections on their experience with the research project are also included.

Assessment Process: The program director and associate program director are in charge of conducting the ePortfolio evaluation. These two faculty members will review each fellow's portfolio independently following an agreed-upon rubric. After completion of their summative evaluations, they will meet to discuss their evaluations and try to reach agreement. If no agreement is reached, the ePortfolio will be reviewed by the fellowship committee, which will then arrive at a final decision.

Preliminary Indications: The Geriatrics fellows who are currently using the ePortfolio have given some preliminary evaluations. They find the ePortfolio to be a very useful component of their training and describe it as easy to use and navigate. They are very satisfied with the opportunity to analyze their performances and see the milestones during their training clearly documented. The fellows gave mixed evaluations about the reflective exercises for the different sections. A clearer explanation of the rationale for self-reflection may be needed. Fellows indicated they would like to continue having access to their ePortfolio at the end of their training.

CONCLUSION

Medical educators have used portfolios for both formative and summative evaluation of trainees. Portfolios promote self-directed learning, lifelong learning, critical thinking, and self-reflection—important characteristics for competent physicians. The introduction of electronic portfolios offers medical educators advantages over hardcopy portfolios by making it easier to access, update, store, search, and distribute portfolio content. This greatly facilitates the documentation of the training experience. The ACGME is increasingly emphasizing educational outcomes in the accreditation process of graduate medical education programs. ePortfolios can be used to evaluate all six ACGME competencies, but they are especially useful in evaluating practice-based improvement, professional and ethical behaviors, and systems-based practice—competencies that are difficult to evaluate using traditional assessment instruments. The authors described their experience implementing an ePortfolio as part of a graduate medical education program in Geriatrics and provided examples of how their portfolio entries address the ACGME requirements. The ePortfolio can easily accommodate a broad variety of materials and reflect the spectrum of academic activities that medical trainees participate in. Setting up an initial ePortfolio structure that is capable of capturing this broad array of information is time consuming and costly. The authors will refine their existing ePortfolio structure based on feedback from the different stakeholders, with the goal to provide a template that can readily be adapted by other graduate medical education programs to document their outcomes.

REFERENCES

ACGME (Accreditation Council for Graduate Medical Education). (2005). *ACGME outcome project: Enhancing resident education through outcomes assessment*. Retrieved May 5, 2005, from http://www.acgme.org/Outcome/

Carraccio, C., & Englander, R. (2004). Evaluating competence using a portfolio: A literature review and Web-based application to the ACGME competencies. *Teaching and Learning in Medicine, 16*(4), 381-387.

Challis, M. (2001). Portfolios and assessment: Meeting the challenge. *Medical Teacher, 23*(5), 437-440.

Chisholm, C. D., & Croskerry, P. (2004). A case study in medical error: The use of the portfolio entry. *Academic Emergency Medicine*, *11*(4), 388-392.

Chodorow, S. (1996). Educators must take the electronic revolution seriously. *Academic Medicine, 71*(3), 221-226.

Chu, L. F., & Chan, B. K. (1998). Evolution of Web site design: Implications for medical education on the Internet. *Computers in Biology and Medicine*, *28*(5), 459-472.

Fung, M. F., Walker, M., Fung, K. F., Temple, L., Lajoie, F., Bellemare, G., et al. (2000). An Internet-based learning portfolio in resident education: The KOALA Multicenter Program. *Medical Education, 34*(6), 474-479.

Hayden, S. R., Dufel, S., & Shih, R. (2002). Definitions and competencies for practice-based learning and improvement. *Academic Emergency Medicine, 9*(11), 1242-1248.

Hays, R. B. (2004). Reflecting on learning portfolios. *Medical Education, 38*(8), 801-803.

Jarvis, R. M., O'Sullivan, P. S., McClain, T., & Clardy, J. A. (2004). Can one portfolio measure the six ACGME general competencies? *Academic Psychiatry, 28*(3), 190-196.

Lonka, K., Slotte, V., Halttunen, M., Kurki, T., Tiitinen, A., Vaara, L., et al. (2001). Portfolios as a learning tool in obstetrics and gynecology undergraduate training. *Medical Education, 35*(12), 1125-1130.

Lynch, D. C., Swing, S. R., Horowitz, S. D., Holt, K., & Messer, J. V. (2004). Assessing practice-based learning and improvement. *Teaching and Learning in Medicine, 16*(1), 85-92.

O'Sullivan, P., & Greene, C. (2002). Portfolios: Possibilities for addressing emergency medicine resident competencies. *Academic Emergency Medicine, 9*(11), 1305-1309.

Ozuah, P. O. (2002). Undergraduate medical education: Thoughts on future challenges. *BMC Medical Education, 2*(1), 8.

Snadden, D. (1999). Portfolios—attempting to measure the unmeasurable? *Medical Education, 33*(7), 478-479.

Stanton, F., & Grant, J. (1999). Approaches to experiential learning, course delivery and validation in medicine. A background document. *Medical Education, 33*(4), 282-297.

Wilkinson, T. J., Challis, M., Hobma, S. O., Newble, D. I., Parboosingh, J. T., Sibbald, R. G., et al. (2002). The use of portfolios for assessment of the competence and performance of doctors in practice. *Medical Education, 36*(10), 918-924.

Ziegelstein, R. C., & Fiebach, N. H. (2004). "The mirror" and "the village": A new method for teaching practice-based learning and improvement and systems-based practice. *Academic Medicine, 79*(1), 83-88.

KEY TERMS

ACGME: The Accreditation Council for Graduate Medical Education (ACGME) is responsible for the accreditation of post-MD medical training programs within the United States. Accreditation is accomplished through a peer-review process and is based upon established standards and guidelines.

Competencies: Complex set of behaviors built on the components of knowledge, skills, attitudes, and "competence" as personal ability.

Geriatric Fellows: Geriatric fellows are physicians who are first trained in family practice or internal medicine and then complete additional years of fellowship training in geriatrics.

Geriatrics (or Geriatric Medicine): The branch of medicine dealing with the medical management and care of older people.

Graduate Medical Education: Residency education or graduate medical education (GME) is the period of clinical education that follows graduation from medical school, and prepares physicians for the independent practice of medicine.

Resident: A medical resident is a physician who has received a postgraduate medical degree (M.D. or D.O.) and is enrolled in a clinical training program in the United States of America.

This work was previously published in Handbook of Research on ePortfolios, edited by A. Jafari, pp. 283-291, copyright 200g by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 3.27
Multimedia Computing Environment for Telemedical Applications

V.K. Murthy
University of New South Wales at ADFA, Australia

E.V. Krishnamurthy
Australian National University, Australia

ABSTRACT

This chapter describes the system design for a multimedia telediagnostic computing environment (MMTE) for telemedical applications. Such an environment requires the design of: (i) a wired-in or wireless computing facility based on currently available technology with a high bandwidth for fast, reliable, and efficient communication of data, voice, and image; (ii) a database query system to access data, voice, and medical images from a fixed server to the mobile or fixed hosts; and (iii) suitable audiovisual software communication tools among the cooperating fixed and mobile hosts to help visualize pointer movements remotely (telepointers) and for teleconferencing. Appropriate software and hardware tools for the design of the cooperative environment are described. We also provide an up-to-date bibliography.

INTRODUCTION

Telemedicine (in short, e-medicine) is a means of delivering medical services to any place, no matter how remote, thereby removing the limitations of space and time that exist in today's health-care settings. Computers are indispensable in telemedicine since they provide for efficient, relevant data gathering for large-scale applications. Besides providing immediate feedback of results to patients and doctors, they also can compare past patient records and evaluate relative improvement or deterioration. Further, they are readily available at any time, are fatigue-free, and can be more objective.

Also computers provide for multimedia imaging—ultrasound, digital X-rays, 3D spiral CAT scanning, magnetic resonance imaging, PET scanning, etc.—and can fuse them into a single multipurpose image using fusion software. Adding mobility to computers enhances their role in telemedical applications considerably, especially at times of emergency, since the patients, doctors, and data collecting and retrieval machines, as well as their communication links, can always be on the move.

For instance, very simple, inexpensive mobile communication and computing devices can be of great help in telemedicine, as illustrated below:

- **Low-cost radio:** Even the simplest of mobile devices—such as a low-power radio that can transmit messages to a home computer, from which medical data can be sent through the telephone line and the Internet—can be of great value in saving lives (Wilson et al., 2000).
- **PDA (personal digital assistant):** The simplest of the computers, such as palmtops and PDAs, can assist the doctors for instant nomadic information sharing and looking for diagnoses of different diseases and treatments. PDAs can help the doctors to figure out drug interactions, storing summaries of sick patients and their drug lists. Further, PDAs can provide for downloading suitable programs from the Web and can be programmed for alert, sending and receiving e-mail, jotting down pertinent points, and storing immediately needed clinical results to carry out ward rounds.
- **Internet:** The Internet is an important tool for medical professionals and will completely change the manner in which medical consultations are provided (Coiera, 1997). For minor ailments, Internet-based consultations to doctors can provide prescriptions for medical/pathological examinations by laboratories. The results are then posted on the Internet for subsequent reading of the results by the concerned doctors, who can prescribe medicines that can be posted on the Internet. This prescription can then be handled by a pharmacy to dispense the medicines to the concerned individual. Kim and Hwang (2001) have proposed a password-controlled Internet-based medical system that brings in a variety of services to doctors, patients, pharmacists, and healthcare professionals. It allows people to receive medical examinations and medical advice on the Internet enables examinations that are not possible in the Internet to be treated to have a direct contact with the doctor.

TELEMEDICAL INFORMATICS

The first step in telemedicine is the telemedical diagnosis (or telediagnosis) based on information obtainable from medical images, blood, urine, and other pathological test reports. Usually, for diagnostic purposes, the doctor sends a patient for such examinations. The laboratory assistant takes the required X-ray or ultrasound images or carries out pathological tests and passes these images (or readings) on to a radiologist/pathologist, who then makes an analysis and sends a report to a doctor. These manual actions are totally sequential and slow. This whole procedure can be made cooperative and faster if the images and data are stored in a database and these can be simultaneously retrieved by doctors, radiologists, and specialists in their offices or homes, using personal computers to make a cooperative diagnosis. This is the basis for **telemedical (e-medical) informatics** (Alfano, 1997; Coiera, 1997; Ganapathy, 2001; Gomez et al., 1997; Jameson et al., 1996; Kleinholz et al., 1994; Lauterbach et al., 1997; Mathew et al., 1999; Pham & Yearwood, 2000; Yearwood & Pham, 2000).

Principal Aims

The principal aims of e-medical informatics are to:

i. Provide online services of patient records (medical and pathological databases) to medical practitioners and radiologists.
ii. Provide primary specialist diagnosis, offer a second opinion, and provide pre- and post-treatment advice through e-mail.
iii. Reduce the cost of imaging equipment, reduce delays, and increase the speed and volume of diagnosis.
iv. Aid cooperative diagnosis and provide assistance for remote surgery.
v. Provide student/resident education.
vi. Reduce professional isolation and increase collaboration.
vii. Provide home care.

Advantages

E-medicine offers the following advantages:

i. Provides health care to underserved and isolated areas so that we can make a better allocation and utilisation of health resources.
ii. Since communication cost is much cheaper than the transportation cost, patients in remote areas can outreach physicians quickly.
iii. Increases the speed of diagnosis and treatment, especially when used for teleradiology, cardiology, and psychiatry.
iv. Allows access to speciality care using time-oriented clinical data.
v. Real-time monitoring of public health databases to prepare and respond during epidemics and biological and chemical terrorism.
vi. The Internet can provide the following support:
    a. Health information
    b. Administrative infrastructure
    c. Online health records
    d. Pharmaceutical information and sales outlets
    e. Online training for telemedical professionals

For an excellent treatment on the Internet and telemedicine, see Coiera (1997).

Thus the development of an integrated wireless (mobile) and wired-in networking environment for e-medical informatics can have a wide range of applications in health care for remotely situated areas. Such a computing environment has a particular relevance to vast countries (e.g., Australia, China, and India) with a widely distributed population, as well as to provide emergency medical assistance in areas of natural disasters and war-torn regions.

Prerequisites

The prerequisites for a successful implementation of a telemedical system are:

- **Infrastructure:** A suitable infrastructure of health-care providers, doctors, engineers, computing specialists, communication engineers, information technology professionals, and medical statisticians to analyse outcomes, and suitable outreach clinics with telemedical facilities.
- **Communication network:** A reliable, inexpensive, readily accessible communication network from outreach clinics to hospitals, doctors, patients, and pathological laboratories.
- **Low-cost computers:** Suitable low-cost hardware/software and a good communication bandwidth for transmission of medical data in different modes (radiological images, video images of signals, and text). While using wired-in or wireless mobile devices and monitors, the effect of electromagnetic interference (EMI) and radio frequency interference (RFI) on data collection and transmission, and the side effects on patients (both physiological and psychological

aspects) have to be taken care of so that improper diagnosis does not result.

- **Training facility:** Training of personnel for providing proper maintenance of equipment and safety standards to patients.
- **Security, reliability, efficiency:** Reliability, efficiency, security, privacy, and confidentiality in handling, storing, and communicating patient information.

Economic Necessity

In densely populated countries (e.g., India) if we want to keep the hospital bed to population ratio near the ideal ratio of one bed for every 500 persons, we will require more than 2 million beds. Assuming that we can build 1,000 hospitals with 200 beds every year, it will take 10 years to cater to the needs of the population in India (Ganapathy, 2001). Added to this is the annual rate of growth of population. The rate of growth in hospital beds to cope up with the increasing population is economically unsustainable and technically not viable since the number of medical specialists also cannot grow to meet this demand. It is intractable in reality. Thus in an ideal situation where every citizen needs to have an immediate medical attention unless we have some kind of telepresence of specialists and doctors. The use of telemedicine avoids unnecessary strain involved in travel and associated expenses, provides immediate attention and care, and can avoid hospitalisation and allow the patients to stay home, enjoying family support.

For example, in developing countries, congenital foetal abnormalities are major causes of prenatal mortality and morbidity. Suitable medical attention can be given at an early stage to help women, using the ultrasound imaging technology at remote places and transmitting the image either through telephone lines or the Internet. Chan et al. (2000) describe a real-time tertiary foetal ultrasound telemedical consultation system, using a standard integrated system digital network (ISDN), that operates in Queensland, Australia. This consultation system has gained acceptance from the clinicians and patients.

As another example, consider the home care for the elderly in developed countries.

Here telemedicine and telediagnosis play a different role. An aging population and rising health costs have created the need to care for more patients in their own homes. Hospitals without walls (e-hospitals, or virtual hospitals) provide for continuous monitoring of patients in certain diagnostic categories. The mobile communication and computing technology is vital to transmit information from a home computer through telephone lines and the Internet to appropriate medical professionals. Wilson et al. (2000) describe how to build such "hospitals without walls." The key technology used here is a miniature, wearable low-power radio that can transmit vital and activity information to a home computer, from which data is sent by telephone line and the Internet to the concerned doctors. This system has been experimented in Australia by the Commonwealth Scientific and Industrial Research Organization (CSIRO).

Thus telemedicine and telediagnosis are economic necessities for both the developing and the developed world, and the capital investment on a telediagnostic computing environment is of great benefit to society.

Challenges to Telemedical Service

Although telemedicine has many advantages, the following challenges arise:

- **Reimbursement:** How to charge patients (rationale), who will bear the cost of treatment, and how can the costs be reimbursed from health providers?
- **Costs:** Who will support the ongoing costs?
- **Human factors:** Will telemedicine gain wide acceptance from patients, doctors, governments, and politicians?

- **Liability, malpractice:** Who will be responsible for liability in case of wrong diagnosis and treatment, and how to avoid and eliminate malpractices?
- **Licensing:** How to evaluate telemedical professionals and provide medical licenses?
- **Telecommunication issues:** How to make an appropriate choice of the bandwidth requirement to make telediagnosis cost effective?
- **Interoperability**: Software interoperability is a key requirement in telemedicine since different modalities of patient records are used, running on different platforms. How to achieve this feature?
- **Funding/sustainability:** Setting up a hospital without walls (or a virtual hospital) requires capital expenditure and sustainable running expenses. Thus it is a national issue requiring cooperation of different agencies.
- **Evaluation/outcomes:** The evaluation of the different diagnostic categories suitable for telemedicine and the outcomes of the treatment procedures require monitoring.
- **Confidentiality and security requirements:** Security of patient data during storage and transmission are vital to safeguard confidentiality and privacy of patient information, as it can be abused or tampered to harm the concerned individuals. These criminal threats are more serious than the virus attack, overloading, and network disruption. Biometric authentication, (Nanavati et al., 2002) could be an answer in dealing with patient information. HIPAA (Health Insurance Portability and Accountability Act) has devised certain security measures in handling telemedical information (http://www.hipaadivsory.com / regs/securityoverview.htm).

We will not deal with these aspects as these are beyond the scope of the main theme of this chapter.

MOBILE MULTIMEDIA TELEDIAGNOSTIC ENVIRONMENT

A **mobile multimedia telediagnostic environment (MMTE)** permits users to work remotely on common shared resources and applications and simultaneously communicate both visually and through audio. Such an environment becomes more flexible and useful if it can work in a wireless and wired-in (integrated) environment so that the services can be provided for mobile hosts. As already mentioned, mobility of users becomes very important for many applications in telemedicine, where doctors, specialists, and health-record officials can cooperate and discuss particular medical treatment. In this application we are concerned with a large data set such as radiological images, video images of signals, and text that are transferred to the host from a server. Typically, in this case the communications are asymmetric in the sense that the server transfers a large amount of information, while the mobile host sends only a small amount of information in the opposite direction. Here the users have interfaces of PDA or laptops interacting remotely. In this application we need cooperation among the participants through special communication tools for conversation (Roth, 2001) and e-pointing (MacKenzie & Jusoh, 2001). The conversation can be of conference type, where two or more participants are involved. The e-pointers aid each participant to point out a particular segment of an image or a video image of a signal so that each of the other participants can visualise the movements of the remote pointer. This would provide for greater clarity and effectiveness of discussion about the particular image. Also in telesurgery (e-surgery) where two or more operating rooms are connected by a network, live video signals may require to be transferred from endo-camera (camera for photographing internal organs) and operating room camera to other remote locations for consultancy. This would allow surgeons not only to see but also visualise surgical instrument movements with 3-D models in real time during the surgery.

PREREQUISITES FOR THE DEPLOYMENT OF MMTE

Reliable Communication Network and Equipment

A reliable multimedia communication network that links the remote centres with hospitals is essential. The delivery of multimedia content in a timely fashion is very important. The key is to find a method of delivery that is universally acceptable, easy to use, and reliable. The answer comes from the Internet. The Internet offers a developer the ability to publish information very quickly. It provides a widely used and uniform communications medium to link users together to access or deliver multimedia information. However, while using mobile devices and the Internet as vehicles for multimedia delivery and interaction, the following factors are to be taken into account:

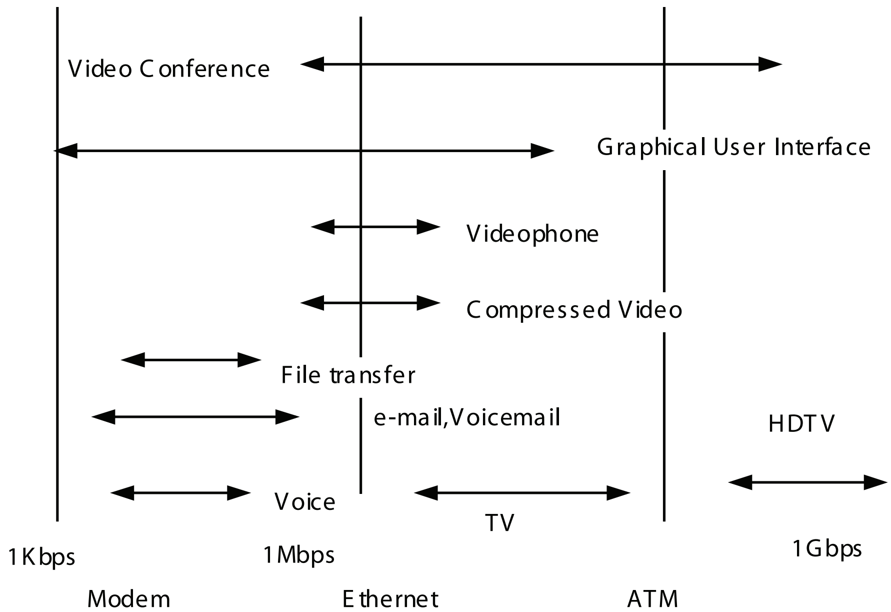
Type of Application

The application requirements can range from a very simple radio to very complex real-time image/signal processing system

Communication Bandwidth Requirements

Text requires the lowest bandwidth, while audio and video data and signals require a significant increase in bandwidth. Specification of bandwidth to be used and compression techniques used are to be laid down so that the data and images that are transmitted are of diagnostic quality. The following issues arise in using networked communication:

Figure 1. Bandwidth (log scale) requirements for applications



i. Choice of equipment and bandwidth management for fast, reliable, and secure transmission of patient data.
ii. Costs involved.

Bandwidth Management

Bandwidth determines the information capacity of a network per unit of time. Wireless networks deliver lower bandwidth than wired networks. Hence software techniques based on compression should be used. Also scheduling communication intelligently can save bandwidth. For use in telemedicine, the techniques should be extremely reliable. Current cutting-edge technologies are yet to develop. Bandwidth requirements along with applications are given in (approximate) logarithmic scale in Figure 1.

Three common technologies used are (bps = bits per second; K= Kilo; M= Mega; G = Giga):

Dial-up mode: Rate 28.8 Kbps
T1: 1.544 Mbps
DS3: 45 Mbps

As an example, standard X-ray transmission takes 30 minutes in dial-up mode, 30 seconds in T1 mode, and 1 second in DS3. It is obvious that the cost goes up as we want to increase the bandwidth to communicate voice, data, and pictures.

As mentioned earlier, even a low-cost dial-up mode can be helpful in telemedicine. Of course the more sophistication we introduce the more complex and costly the system gets, and the system has to be supported by governmental agencies or health insurance providers.

DS3 is a dedicated, private line service designed for point-to-point communication. This service uses fibre-optic cable. One can have the option of tiered DS3 service from lower to higher bandwidth from 5 Mbps to 45 Mbps, depending upon cost considerations and transmission requirements.

Choice of Multimedia Hardware

We need a proper choice of graphics and audio equipment for quality images and audio for diagnosis.

Low-Cost Computing Environment

Low-cost workstations are to be made available at the user's and health provider's end. Suitable software and hardware are to be developed so that the workstations are easy to use by doctors, nurses, and other paramedical staff. For this purpose, suitable training centres are to be made available.

Cost of Consultation

Low cost per consultation, faster diagnosis, faster reply to queries, and quicker medical aid when compared with the conventional method of a direct visit to the doctor.

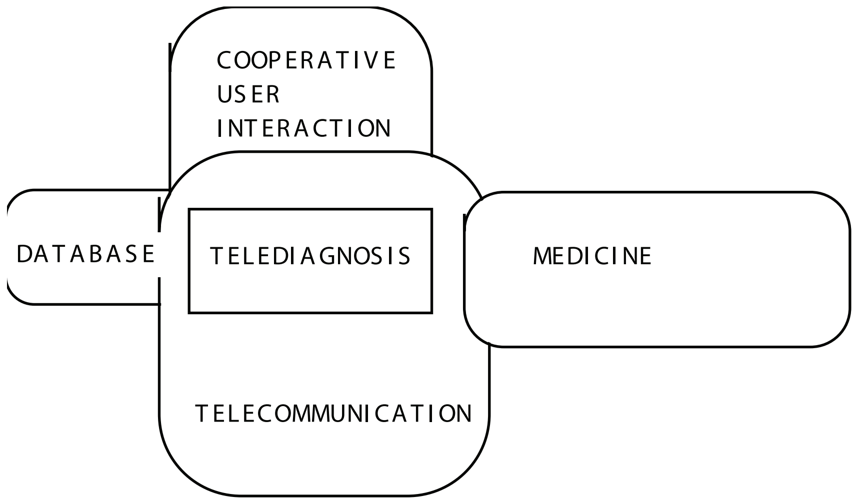
Case Evaluation and Emergency Attention

Quick evaluation of emergency cases and arrangement for transportation to hospitals supported with mobile telemedical personnel and equipment.

FEATURES OF A TELEDIAGNOSTIC ENVIRONMENT

A typical cooperative telediagnostic (e-diagnostic) environment is shown in Figure 2. Yearwood and Pham (2000) and Pham and Yearwood (2000) describe a cooperative telemedicine environment for interactive processing of visual data. This environment consists of databases containing medical/pathological images and textual information, a cooperative interaction facility, and telecommunication. The typical characteristics of such an environment are:

Figure 2. A cooperative diagnostic mobile environment



- **Remoteness of users:** Users work remotely on common shared resources and applications and simultaneously communicate both visually and through audio. This environment becomes more flexible and useful if it can work in an integrated (wireless and wired-in) environment so that the services can be provided for mobile hosts.
- **Cooperation:** Many e-medicine applications require doctors, specialists, and health-record officials to cooperate and discuss particular medical treatment. In this application we are concerned with a large data set such as radiological images, video images or signals, and text that are transferred to the host from a server. Cooperation among the participants requires special communication tools for conversation, e-pointing, and e-surgery.
- **E-pointing facility:** E-pointing is the act of moving an on-screen tracking symbol, such as a cursor, by manipulating the input device. Such a tracker can be placed over text, images, icons, and menu items and is basic to human interaction with graphic interfaces. Such a facility can aid conversation between two or more participants dealing with medical images or signals. The interactive e-pointers (MacKenzie & Jusoh, 2001) aid each participant to point out a particular segment of an image or a video image of a signal so that each of the other participants can visualise the movements of the remote pointer and discuss any abnormality for diagnostic purposes.
- **Facility for Internet link:** Also users should be able to remotely access worldwide knowledge bases/databases, download files from the Internet, and browse the World Wide Web using portable and mobile devices such as laptops, palmtops, and PDAs. Further technological developments are needed to provide the PDA users with facilities to handle text, graphics, and audio messages and access interactive multimedia services.

SYSTEM STRUCTURE OF MMTE

The integrated computing environment consists of fixed host computers and wired-in/wireless mobile client computers (Bates, 1994, Hatfield, 1996; Imielinski & Badrinath, 1994). Fixed hosts (FHs) are connected together via a fixed high-speed network (Mbps to Gbps). The mobile clients (MCs) are capable of connecting to the fixed network via a wireless link. The components in the fixed network are called fixed hosts. Fixed hosts (FHs) provide mobile application services and coordinate tasks to mobile hosts. MCs support query invoking and information filtering from FHs to provide personal information service.

The remote access to image and other related databases may appear like automatic teller machine transactions, but they differ in several respects:

i. They are long-duration transactions requiring the transfer of a large amount of data and require lengthy negotiation, cooperation, pointing, and decisions to reach a final diagnosis and therapy. In this cooperative environment, therefore, we need to sacrifice some of the properties of the traditional transaction model (Krishnamurthy & Murthy, 1992) that enforces ACID properties (atomicity, consistency, isolation, and durability). Note that, in particular, we cannot roll back any transaction in the context of e-surgery. Also isolation implies that the actions of one participant are invisible to another. Therefore, we need to introduce a more realistic model in which the isolation property is removed and intermediate results are made visible so that any catastrophic incidents can be avoided. Also precedence order in execution and context dependencies have to be taken care of. This means we must remove the atomicity constraint that requires "***all or none***" operation and rollback. This model is called "***a workflow model***." More detailed aspects of this model and relevant software design features are described briefly in a later section and in chapter 2 of this book.
ii. A service to permit two users to interact simultaneously with a medical imaging system for cooperative diagnosis. To help diagnosis we need to provide each user with a pointer of a desired identity that can be locally operated and at the same time can be visualised at the remote terminal. This is a concurrent action from two or more users on a read-only image database.
iii. Also an audio conference facility is needed for cooperative diagnosis.
iv. Collaborative multimedia environment allows medical specialists to cooperate in diagnosis. The environment has to support remote database access for medical images, the retrieval of relevant medical cases to support diagnosis, and communication among participants through telepointers and image annotation for freehand drawing.
v. A telepointing device with a very high reliability.

Thus the key elements in the design of a telediagnostic system are:

i. Design of an appropriate workflow model for telemedical applications.
ii. Supporting database access in a mobile cooperative environment for medical images, related text, and video signal images. Also additional support through handheld devices, such as PDAs, for supporting remote information sharing and processing.
iii. Providing a cooperative environment for communication among participants audiovisually through telepointers to visualize movements remotely.
iv. Appropriate hardware/software tools.
v. Implementation of the reliability, fault tolerance, and recovery features.

MOBILE WORKFLOW

Mobility is central to many applications involving telemedicine. This arises due to the fact that the patients, doctors, and data collecting instruments can always be on the move. Hence any transaction that occurs among the clients and servers is a mobile transaction. A mobile transaction is a distributed transaction that can be executed partly within that mobile client (MC) as internal transactions (intran) and partly in other fixed hosts (FHs) as external transactions (extran). Each FH has a coordinator FHC that receives external transaction operations from mobile hosts and monitors their execution in the database servers within the fixed network. Similarly each MC has a coordinator MCC.

In a mobile environment, some of the properties of the traditional transaction model (see Chapter 2 of this book; Krishnamurthy & Murthy, 1992) that enforces atomicity,consistency, durability, and isolation have to be replaced by a more realistic model in which the isolation property is removed and intermediate results are made visible. Also precedence order in execution and other dependencies have to be taken care of; this means we must remove the atomicity restrictions. Such a model turns out to be a workflow model between the MC and FH.

A **workflow** is a collection of tasks organized to accomplish some business activity. Here each task defines some unit of work to be carried out. A workflow ties together a group of tasks by specifying execution dependencies and the dataflow between tasks. Also there could be a constraint that a particular task cannot begin until some other task ends. Such constraints can be specified by event-action systems.

In a workflow model we need to ensure correctness and reliability of workflows. Correctness means that the concurrent execution of transactions and workflows is interleaved in a way that incorrect results such as lost update or inconsistent retrieval cannot occur. Recovery means both the tasks and the transactions are recoverable in the event of a failure arising due to disconnection.

Thus we need to satisfy the following requirements:

i. A programming paradigm to support a combination of transactions that satisfy atomicity, consistency, isolation, and durability (ACID) properties, and workflow tasks that do not have such ACID properties.
ii. A method of managing control flow of workflow and transactions.
iii. A suitable recovery model under failure of the workflow. This recovery is very complex compared to short transactions since it will require re-instantiation and following the control flow of actions strictly.
iv. Contextual information (Henricksen et al., 2002; Mattern & Naghshineh, 2002) that preserves the consistency of databases as well as the local state of the application.
v. Remembering the execution history and path and local states produced in the past.
vi. Externalization of preliminary results: workflow computations need to externalize their results before they are completely done. This implies that unilateral rollback is no longer possible; one needs to specify compensating actions as part of the control flow description.
vii. Concurrency and consistency control: Consistency can no longer be based on serializability only; we must now allow for application-oriented policies of synchronizing access to shared objects.
viii. Conflicts handling: In general, it is not feasible to let some surgical or diagnosis activity wait in case of a resource conflict until a long-duration activity has completed. Nor is it acceptable to roll it back to the beginning. Therefore the control flow description has to specify what should be done if a resource conflict occurs and how it can be resolved using priorities and context. For example,

in an emergency, it is necessary to suspend less critical tasks and postpone them for a later time.

HUMAN-COMPUTER INTERFACE AND SOFTWARE DESIGN

The software and human-computer interface system design consists of the following steps:

i. Design an appropriate transactional workflow model for telemedical applications.
ii. Support database access in a mobile environment for medical images and related video signal images.
iii. Provide a cooperative environment for communication among participants audio-visually through telepointers to visualize movements remotely.
iv. Adapt currently available tools to design the required software for the cooperative multimedia mobile environment.

Two key issues involved in the above steps are:

- Choice of an appropriate software tool to implement mobile workflow.
- Choice of suitable hardware for mobile access and cooperation.

We will now provide some practical aspects on the aforesaid design steps.

Software Tools

Presently, Java (Flanaghan, 1996; Hamilton, 1996; Yourdon, 1996) seems to be the best choice due to the following important features:

- **Portability:** Java is platform independent.
- **Efficiency, power, simplicity:** Java retains the efficiency, power, and simplicity of C++ but is less complex. There is no pointer arithmetic and no memory management. Memory management occurs automatically. Strict rules are enforced regarding variables; this permits detection of mistakes when the program is compiled.
- **Safety:** Java is safer than other low-level languages such as C++ since Java is free from pointer errors and has uniform reference semantics and automatic storage management. Also it provides security.
- **Openness:** It is a de facto standard for platform-independent computing and can run at any level: client, application, or database.
- **Transmission on network:** Java has been developed with the aspect of transmission across the network in mind; Java includes very developed network functionality in its core so that the application developer can concentrate more on users' needs rather than on details of network code.
- **J/SQL and Java Data Base Connectivity (JDBC):** JDBC is a simple call-level library interface. J/SQL provides a seamless integration of Java and SQL. JDBC provides for simultaneous connection to several databases, transaction management, simple queries, manipulation of precompiled statements with bind variables, calls to stored procedures, streaming access to long column data, access to the database dictionary, and description of cursors.

J/SQL is an integration of SQL statements in Java programs. It is more concise than JDBC and more amenable to static analysis and type checking. Use of J/SQL has several benefits:

i. Provides for compile-time type checking.
ii. Runs everywhere.
iii. Extends server: The Oracle database server will store a user's J/SQL program and execute it in its server.

iv. Bridges languages: The J/SQL preprocessor draws a clear distinction between the standard Java language and the supplementary J/SQL clauses, which add functionality to that language by exposing database objects (e.g., tables, cursors) and services (SQL, Persistence) as standard Java objects.

JDBC provides for both automatic commit as well as specific commits from the application before a transaction is committed to the database. JDBC does not support two-phase commit protocol. External support is needed to allow the Java application to have two-phase commits.

Persistent Java (PJava) allows for online transactions. Also the programmer can create new transactional styles. Such styles, as already mentioned, are very important for mobile applications. Transactions in PJava can be launched either synchronously (in the same thread) or asynchronously (in different threads) by invoking the Start () method of the transaction object.

Multithreaded processes are useful for clinical decision support in complicated medical diagnosis involving temporal reasoning for real-time medical data.

Nested transaction enables one to perform updates in a child transaction without affecting the parent transaction. A child transaction that completes successfully passes all its updates (the modified objects) to its parent transaction. If the child transaction aborts, none of its updates are ever reflected in the parent transaction. Also we can spawn parallel independent nested transactions. In this case each of the sibling transactions are isolated from all others and can commit or abort independently.

JetConnect allows real-world applications to be developed using Java by providing data access to any ODBC- or JDBC-enabled database, including DB2 and Oracle. JetConnect augments the Java programming language with easy-to-use, vendor-independent database connectivity that enables transaction-oriented access to corporate data stores from Internet/intranet applications. JetAssist is a GUI-based tool for rapid development of database applets that take advantage of JetConnect's rich set of features. JetAssist employs a wizard-style approach that steps the user through applet creation. Armed with only a casual acquaintance with database principles and a Java-enabled Web browser, an end user can construct and maintain a custom applet using an interface that requires absolutely no programming.

JetExpress provides high-speed connectivity between Java applets and applications and the corporate information resource. JetExpress provides transaction acceleration functionality when accessing DB2 databases, with specific optimization for other DBMSs scheduled in the near future.

To use JetConnect, the system must have these operating systems and software packages installed: either Windows 95/Windows NT or a supported version of UNIX; Java Developer's Kit (JDK); the 32-bit ODBC drivers for the database we want to access. Our experience reveals that Java, the design by contract tool in Java iContract (Kramer, 1998), Java Database Connectivity, JetConnect tools, UML (Gogalla & Kobryn, 2001), and OCL (Warmer & Kleppe, 1999) are eminently suited for our purpose. The type of relaxed transactions we need (subjunctive, or "what-if," programming) requires that we execute hypothetical or pseudo-transactions to test the intention of actions for trial-error design. Such a model uses "virtual copy" and arises in real-time transactions in patient-health monitoring, flight plan, and cooperating information systems.

Lee (2001) describes a Java-applet-based image-guided telemedicine system via the Internet for visual diagnosis of breast cancer. This system automatically indexes objects based on shape and groups them into a set of clusters. Any doctor with Internet access can use this system to analyse an image, query about its features, and obtain a satisfactory performance (http://dollar.biz.uiowa.edu/~kelly/telemedicine.html).

Software Agents

Software agents, which are personalized, continuously running, and semiautonomous objects, can play an important role in cooperative telemedicine. Such agents can be programmed for supporting medical diagnostic intelligence and keep a watchful eye to discover patterns and react to pattern changes that usually occur in epidemics and biological and chemical terrorism. Also agents can help in the appropriate choice of doctors and allocation of hospitals and beds for patients. Since different modalities used for inputting the patient diagnostic data require a heterogeneous computing environment, software interoperability is a major issue. Software agent technology helps to solve this issue; Ganguly and Ray (2000) describe a methodology for the development of a software-agent-based interoperable telemedicine system. This system has been illustrated using a tele-electrocardiography application. Unified Modelling Language, or UML (Gogolla & Kobryn, 2001; Warmer & Kleppe, 1999), has been used for this purpose. O' Hare and O'Grady (2002) describe the use of agents deployed from PDAs and mobile devices. This approach will be useful for online access of a large repository of medical information that is customised and personalised to match the interest profile of an individual. For medical education this will be a very useful tool.

FIPA (the Foundation for Intelligent Physical Agents; http://www.fipa.org) aims to improve agent interoperability by providing standards for protocols and languages. Also the Java 2 Micro Edition (J2ME) is targeted at PDAs. These developments will provide an agent execution environment in PDAs.

Hardware: PCs, Laptops, and PDAs

Presently, portable and mobile devices, such as laptops and personal digital assistants, are not very well suited to remotely access resources in an efficient and reliable manner for reliable telediagnosis (Roth, 2001). For instance, information sharing among doctors through handheld appliances is of limited applicability, except for transmitting small amounts of data, such as heart rate, blood pressure, and other simple monitoring devices. Typical data types for current handheld appliances are: text, data entries, numbers, and tables. None of the applications above can deal with multimedia data such as audio video, which requires considerable bandwidth, sufficient output devices, and a very powerful battery. Also, currently multimedia data are not yet suitable for handheld appliances. Very few applications support graphical data (Roth, 2001).

To provide robustness in the mobile environment, we must overcome different kinds of failures (e.g., radio link failures, mobile host and fixed host failures). Suitable protocols to recover under faults using the recovery mechanisms or rescue clauses are to be developed. WaveLAN radio cards (WaveLAN/AT cards for PCs and WaveLAN/PCMCIA cards for laptops or notebooks) are available (Fogle, 1995; Wong, 1995). WaveLAN is an indoor wireless LAN based on a cellular structure. A cell is the wireless coverage area in which WaveLAN stations can communicate with each other. Also, infrared-based wireless technology is available commercially. Using these devices, PCs are linked.

Telepointers

This is an interaction style for presentation systems, interactive TV, and other systems where the user is positioned at a remote site from the display. Pointing is the act of moving an on-screen tracking symbol, such as a cursor, by manipulating the input device from a remote site. Presently, two devices are commercially available: these are GyroPoint (a product of Gyration Inc., Saratoga, California) and RemotePoint (a product of Interlink Electronics, Camarillo, California).

GyroPoint works in two distinctly different modes of operation: as mouse or as a tracking pointer based on gyroscope principles so that spatial motion can be generated across a screen.

RemotePoint uses infrared for transmission and joystick for motion control. Also telepointers for remote pointing of images and conferencing are not fully developed yet. Currently available tools, such as GyroPoint and RemotePoint(MacKenzie & Jusoh, 2001), lack accuracy and speed. The telepointer technology has yet to mature to provide reliable service for e-medicine and e-surgery. Along with the use of virtual reality, the pointer technology can be useful in teaching surgery and for trial-error planning of operations.

CONCLUSION

Cooperative telemedical informatics will have a direct impact on the rapidly changing scene of information and communication technology (including advanced communication systems) and will provide for greater interaction among doctors, radiologists, and health-care professionals to integrate information quickly and efficiently and make effective decisions at all levels. Some important pilot studies have already taken place in India recently in Apollo hospitals for emergency consultations and remote health care in Indian villages using satellite links (Ganapathy, 2001). Virtual e-hospitals will be a reality in the near future, providing much needed health care that is of great benefit to the human society. Telemedical informatics offers exciting possibilities and is a vastly expanding area of research. For those interested in current developments in these areas in the United States (National Library of Medicine, http://www.nlm.nih.gov/reserch/telemedicinit.html; University of Iowa, Virtual Hospital, http://www.vh.org/index.html) and Australia (CSIRO), reference may be made to several journals in the telemedicine area and also the Web sites provided at the end of this chapter.

ACKNOWLEDGMENT

The authors thank the reviewers for providing valuable suggestions in revising this chapter.

REFERENCES

Alfano, M. (1997). User requirements and resource control for cooperative multimedia applications. In *Lecture notes in computer science: Vol. 1242. Multimedia applications, services and techniques* (pp. 537-553). New York: Springer-Verlag.

Bates, R. J. (1994). *Wireless networked communications.* New York: McGraw-Hill.

Chan, F. Y., et al. (2000). Clinical value of real-time tertiary foetal ultrasound consultation by telemedicine. *Telemedicine Journal*, *6*, 237-242.

Coiera, E. (1997). *Guide to medical informatics, the Internet and telemedicine*. London: Chapman & Hall.

Flanagan, D. (1996). *Java in a nutshell.* Sebastapol: O'Reilly.

Fogle, D. (1995, September). Room to roam. *LAN Magazine*, 123-127.

Ganapathy, K. (2001, November). Telemedicine in action—The Apollo experience. *Proceedings of the International Conferenceon Medical Informatics.*

Ganguly, P., & Ray, P. (2000). A methodology for the development of software agent based interoperable systems: A tele-electrocardiography perspective. *Telemedicine Journal*, *6*, 283-294.

Gogolla, M., & Kobryn, C. (2001). <<UML>> 2001. In *Lecture notes in computer science: Vol. 2185.*New York: Springer-Verlag.

Gomez, E. J., et al. (1997). The Bonaparte telemedicine ATM multimedia applications. In *Lecture notes in computer science: Vol. 1242.*

Multimedia applications, services and techniques (pp. 693-708). New York: Springer-Verlag.

Hamilton, M. A. (1996). Java and the shift to Net-centric computing. *IEEE Computer*, *29*(8), 40-46.

Hatfield, D. N. (1996). The technology basis for wireless communication. *Telecommunications*, *30*(11), 56-60.

Henricksen, K., et al. (2002). Modeling context information in pervasive computing systems. In *Lecture notes in computer science: Vol. 2414* (pp. 167-180). New York: Springer-Verlag.

Imielinski, T., & Badrinath, B. R. (1994). Mobile wireless computing. *Communications of the ACM*,*37*(10), 18-29.

Jameson, D. G., et al. (1996). Broad band telemedicine: Teaching on the information Superhighway. *Journal of Telemedicine and Telecare*, *1*, 111-116.

Kim, S. S., & Hwang, D.-J. (2001). An algorithm for formation and confirmation of password for paid members on the Internet-based telemedicine. In *Lecture notes in computer science*: Vol. 2105 (pp. 333-340). New York: Springer-Verlag.

Kleinholz, L., et al. (1994, Winter). Supporting cooperative medicine. *IEEE Multimedia*, 44-53.

Kramer, R. (1998). iContract—The Java design by contract tool. *26th Conference on Technology of Object-Oriented Systems (TOOLS USA'98).*

Krishnamurthy, E. V., & Murthy, V. K. (1992). *Transaction processing systems*. New York.

Lauterbach, Th., et al. (1997). Using DAB and GSM to provide interactive multimedia services to portable and mobile terminals. In *Lecture notes in computer science: Vol. 1242*. *Multimedia applications, services and techniques* (pp. 593-609). New York: Springer-Verlag.

Lee, K.-M. (2001). Image-guided telemedicine system via the Internet. In *Lecture notes in computer science: Vol. 2105* (pp. 323-333). New York: Springer-Verlag.

MacKenzie, S., & Jusoh, S. (2001). An evaluation of two input devices for remote sensing. In *Lecture notes in computer science: Vol. 2254. Engineering for human-computer interaction* (pp. 235-250). New York: Springer-Verlag.

Mathew, T. L., et al. (1999). *Modern technologies for biomedical information handling.* New Delhi, India: Institute of Nuclear Medicine and Allied Sciences.

Mattern, F., & Naghshineh, M. (2002). Pervasive 2002. In *Lecture notes in computer science:Vol. 2414*. New York: Springer-Verlag.

Nanavati, S., et al. (2002). *Biometrics*. New York: Wiley.

O'Hare, G., & O'Grady, M. (2002). Addressing mobile HCI needs through agents. In *Lecture notes in computer science: Vol. 2002* (pp. 311-314). New York: Springer-Verlag.

Pham, B., & Yearwood, J. (2000). Delivery and interactive processing of visual data for cooperative telemedicine environment. *Telemedicine Journal*, *6*, 261-268.

Roth, J. (2001). Information sharing with handheld appliance. In *Lecture notes in computer science: Vol. 2254. Engineering for human-computer interaction* (pp. 263-279). New York:Springer-Verlag.

Warmer, J., & Kleppe, A. (1999). *The Object Constraint Language*. Reading, MA: Addison-Wesley.

Wilson, L. S., et al. (2000). Building hospitals without walls: A CSIRO home telecare initiative. *Telemedicine Journal*, *6*, 275-281.

Wong, W. (1995, July). Light moves. *LAN Magazine*, 123-127.

Yearwood, J., & Pham, B. (2000). Case-based support in a cooperative medical diagnosis environment. *Telemedicine Journal*, *6*, 243-250.

Yourdon, E. (1996). Java, the Web and software development. *IEEE Computer*, *29*(8), 25-39.

Electronic References

http://telemed.medicine.uiowa.edu—Introduction to Telemedicine by Susan Zollo.

http://www.nlm.nih.gov/research/telesymp.html—Provides information on the Telemedicine Symposium.

http://telemed.medicine.uiowa.edu—Provides slide shows on various aspects of telemedicine prepared by the National Laboratory for the Study of Rural Telemedicine, of the University of Iowa.

http://telemed.medicine.uiowa.edu—Telemedicine for Home Monitoring by Michael Kienzle.

http://www.ntia.doc.gov/reports/telemed/—Telemedicine Report to Congress.

http://telemed.medicine.uiowa.edu—University of Iowa's Telemedicine Resource Center (UI TRC).

http://tie.telemed.org/—Telemedicine Information Exchange.

http://www.nlm.nih.gov/research/telemedinit.html—National Library of Medicine's National Telemedicine Initiative.

http://www.tmgateway.org/—Federal Telemedicine Gateway.

This work was previously published in Mobile Commerce Applications, edited by N. S. Shi, pp. 95-116, copyright 2004 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 3.28
Identifying Optimal Chronic Kidney Disease Patient Education Web Sites:
Assessing E-Health Technology by Content Area Experts

Lisa M. Nanovic
University of Wisconsin, USA

Jonathan B. Jaffery
University of Wisconsin, USA

ABSTRACT

One in nine adults in the United States has chronic kidney disease (CKD). Randomized studies show electronic health (e-health) systems improve health outcomes in chronic disease. This study describes a systematic evaluation of available CKD Web sites. We evaluated Web sites for educating patients with CKD, focusing on three specific design issues: usability, quality, and content. Thirty Web sites were evaluated between April and July 2004. Cohen's kappa and intraclass correlation of quartile rankings for two independent evaluators were calculated. Mean score for evaluator 1 was 7.293 (standard error 0.511) and for evaluator 2 was 8.189 (0.413). Cohen's kappa for the 2 evaluators' total scores for all 30 Web sites is 0.7671, and intraclass correlation is 0.7703. In general, results show that it is possible to identify Web sites more likely to provide a positive educational experience for CKD patients. Further evaluation is needed to investigate the utility of variably ranked Web sites as educational interventions.

INTRODUCTION

Medical informatics has traditionally consisted of applications to improve information exchange, decision-making, and time management for health care professionals. As the fundamental relationship between patients and physicians has changed from paternalistic to partnership, the informatics focus is shifting towards the consumer (Eysenbach, 2000). The objectives of this transition are to empower patients with appropriate medical knowledge, education, and the tools for self-management. This may be particularly effective in enhancing patient and health care provider partnerships in the management of chronic disease (Celler, & Lovell, 2003). According to the Pew Research Center, a nonpartisan organization that provides information on national and global trends, 70% of the health care consumers, who are influenced by online information, report that the Internet affected their decision making about health care (Fox, & Rainie, 2002). Still, very little is known about how patients judge health information obtained on the Internet. Moreover, data suggest that while health care consumers are facile at obtaining answers to directed health questions, they seldom assess the quality or source of the information. Consumers also tend not to remember Web sites from where they acquired the information (Eysenbach & Kohler, 2002). Thus, health care provider-sanctioned Web instruction is needed to help alert and direct patients to information that is valid and accurate.

One in nine adults in the United States has chronic kidney disease (CKD) (Coresh, Astor, & et al. 2003). This creates a tremendous public health burden, with more than $22 billion in spending on the end-stage renal disease (ESRD) program in 2002 (USRDS, 2002). Slowing the rate of loss of kidney function in later stages of CKD by 10-30% could save $5 to $30 billion dollars in health care expenditures over the next decade (Trivedi, & Pang, 2002). Prior to end-stage disease, there may be opportunities to slow progression and delay disease complications, but patient education is vital. Time constraints prevent clinicians from taking on the task of repetitive patient education and counselling. However, customized Web sites available via the Internet may be able to fill this need by providing novel ways to reach patients with CKD.

Randomized studies have found that e-health systems can improve knowledge, self-care, quality of life, coping skills, health care participation, and health outcomes in chronic disease, as well as self-management behaviors, including diet, exercise, medication taking, and health utilization (Gustafson, & Hawkins, 1999a; Gustafson, & McTavish, 1999b). However, very little is known about the selection of e-health Web sites and the assessment of medical information obtained by patients (Eysenbach & Kohler, 2002). There are a number of CKD patient education sites available on the World Wide Web, some of which are sponsored by not-for-profit organizations and some by private industry. Few are written at a level so as to be useful to the general population, and many do not comply with available standards for health Web sites (Calderon, & Zadshir, 2004a; Calderon, & Zadshir, 2004b; Jaffery, & Becker, 2004).

Additionally, as research in the field of e-health expands, it is becoming increasingly clear that in order to capitalize on the potential power of the Internet as an educational tool, innovative methods for communicating with patients is needed. Marill, Miller, and Kitendaugh (2006) report on the ongoing challenges to the National Library of Medicine's Medline Plus, highlighting the dynamic process inherent in health information Web site design.

Although more literature is emerging on optimizing human-computer interactions for specialty clinical areas and the education of chronic disease patients, to date, much of this has been led by the cancer and mental health arenas (Bader, & Strickman-Stein, 2003; Chelf, & Deshler, 2002; Chernecky, & Macklin, 2006; Chou, & Lin, 2004), with limited contribution in the nephrology literature. The purpose of this manuscript is to describe a systematic evaluation of readily available CKD Web sites. A more comprehensive review of these sites revealed significant deficits that are likely to impair their usefulness as educational tools.

METHODS

Based on our previous work, relevant literature in other disease states, and published guidelines for evaluation of e-health Web sites (Kim, & Eng, 1999; Robinson, & Patrick, 1998; Winker, & Flanagin, 2000), a system was developed for evaluating Web sites aimed at educating patients with CKD. Three distinct Web-based design issues were addressed: usability, quality, and content.

1. **Usability:** Defined on the basis of the ease with which a user can access clinically relative material (ease of navigation) and readability (reading level), was assessed with the following tools:
   a. Reading level, which was assessed using the Flesch-Kincaid grade level formula (Kincaid, Rogers, & Chissom, 1975).
   b. Ease of navigation, which was assessed by the number of links needed to get to clinically relevant material. Clinically relevant material is defined as containing information directly related to aspects of CKD management most likely to improve clinical outcomes, as elaborated below. Number of links was defined as the number of mouse-clicks needed to arrive at such material from the home page.
2. **Quality** is a complex measure. Defining and assessing the quality of health education Web sites remains an elusive goal (Gattoni & Sicola, 2005). Various criteria have been used to evaluate the quality of health information on Web sites. These have arisen through expert consensus or arbitrary design. In this study, we defined quality specifically on the basis of adherence to those shared areas that these groups describe as essential when judging the quality of e-health sites (Kim et al., 1999; Robinson et al., 1998; Winker et al., 2000). In this sense, the opinions of area experts are key in judging the quality of relevant CKD Web sites.

Quality questions *(Yes/No)*:

1. Authorship/source: Are the sources of health information mentioned with credentials listed?
2. Quality/best-available evidence: Is it clear whether information is based on scientific evidence (studies), expert consensus, or personal/professional experience/opinion?
3. Privacy: Is there a privacy policy?
4. Updated: Is the date of last modification listed?
5. Contact: Is a link to a valid contact form/e-mail available?
6. Ownership/sponsorship: Are all sources of funding listed?
7. Advertising: Is there an advertising policy?
8. Editorial content: Is there an editorial policy/description of content review?

3. Content had no standard measures to evaluate what information is appropriate to convey to health care consumers visiting a CKD Web site. Evaluations of other e-health education sites for other disease entities have tended to focus on five to eight clinical areas (Croft & Peterson 2002; Impicciatore, Pandolfini, 1997). We judged a CKD Web site's content according to how accurately it addresses six areas. These are based on specific management goals according to the kidney disease outcome quality initiative (K/DOQI) (National Kidney Foundation K/DOQI, 2004), emphasizing aspects of CKD management most likely to improve clinical outcomes. At this point, we focus on each of the areas that are grouped under high blood pressure, CKD complications, and nutrition.
   a. High blood pressure. Hypertension is both a significant underlying cause and

complication of CKD. Blood pressure control is recognized as among the most important factors for slowing the progression of CKD and is associated with other positive outcomes, including the prevention of cardiovascular disease.

Questions *(Yes/No)*:

1. Hypertension: Are specific blood pressure goals addressed?

b. *Complications of CKD.* Among the kidney's functions is the production of two hormones, erythropoietin, a glycoprotein hormone that stimulates the production of red blood cells by stem cells in bone marrow, and 1,25 dihydroxy vitamin D_3, a steroid compound necessary for normal bone growth. Consequently, two of the most common complications of CKD are anaemia and bone disease, or renal osteodystrophy.

Questions *(Yes/No)*:

2. Anemia: Are target hemoglobin/hematocrit goals addressed?
3. Bone disease: Is vitamin D replacement therapy addressed?

c. *Nutrition.* Diet is one of the most modifiable factors that can affect the control of all of the above, thereby offering the potential to have profound effects on both CKD progression and morbidities. Therefore, our content tool evaluates whether CKD Web sites addresses some of the nutritional areas most relevant to CKD, namely sodium, phosphorus, and protein intake.

Questions *(Yes/No)*:

4. Sodium: Are specific dietary sodium goals addressed?
5. Phosphorus: Is appropriate phosphate binder use addressed?
6. Protein: Are specific dietary protein intake goals addressed?

It is important to note that as the appropriateness of CKD Web site content is something that cannot be assumed by the Web site owner nor assessed by the consumer (CKD patient), there is the need for content validation by experts, irrespective of target audience. Therefore, in designing our study, two nephrologists were to perform the Web site evaluation, drawing on their expertise in assessing CKD Web site content. Other aspects of evaluation are objective including "reading level" and "number of clicks," but it was vital that experts in nephrology assess the appropriateness of content.

CKD WEB SITE IDENTIFICATION AND STATISTICAL ANALYSES

CKD Patient Education Web site Identification

Accordingly, the investigators also felt that it was crucial to attempt as exhaustive a search of CKD Web sites as possible, using multiple common search engines. For instrument testing and validation, however, the evaluation was led to exploring the National Kidney Disease Education Program Chronic Kidney Disease Compendium (NKDEP) (2004) for additional relevant Web sites to support instrument testing and validation.

Using five conventional search engines (Excite, Google, HotBot, Lycos, and Yahoo), we entered four search terms (kidney disease, chronic kidney disease, chronic kidney failure, and renal failure) and identified the first 40 uniform resource locators (URLs), the World Wide Web address of a site on the Internet, from each search engine. Of these, 110 were redundant, leaving 90 discrete sites. Twenty-one sites were directed toward the care of animals, 18 were obsolete or pertained to nondisease states, 19 were oriented toward providers rather than patients, and two were focused solely on end-stage renal disease. For the final evaluation, this left 30 Web sites designed for educating patients with CKD.

In the data analysis, we calculated quartile rankings for each of the three areas for two independent evaluators for all 30 Web sites using Microsoft 2000 Excel. Cohen's kappa and intraclass correlation were calculated using the online statistics toolbox of the Department of Obstetrics and Gynaecology at the Chinese University of Hong Kong (Department of Obstetrics and Gynaecology 2005 [accessed 2005 March 23]).

RESULTS

Preliminary Content Validity and Reliability Testing

Initially, five Web sites not among those identified by the search described above were identified using the NKDEP (2004). Two nephrologists from University of Wisconsin-Madison evaluated these sites. Revisions were made based on these pilot evaluations, and this process was repeated with additional Web sites identified in the same manner, allowing for very good correlation among evaluators (r=0.925).

Specifically, revisions to the "usability" construct included the "ease of navigation" concept in addition to "readability." During piloting, readability was assessed using both the Fry readability scale and the Flesch-Kincaid grade level formula. We found these tools to correlate very well (r=0.857), which is consistent with previous correlations between the Flesch-Kincaid formula and other readability scales (Kincaid et al., 1975; Paasche-Orlow, Taylor, 2003). As well, the Flesch-Kincaid formula is significantly easier to use, as it is an embedded tool in Microsoft Word. Therefore the Flesch-Kincaid formula was subsequently chosen for the 30 Web site evaluation. Accommodation for visual impairment was removed as none of the Web sites allowed for this. Content was revised to better reflect specific goals and published guidelines of the National Kidney Foundation, which in turn are based on best evidence or, in the absence of evidence, expert opinion (K/DOQI, 2004). Finally, quality was revised to allow for greater objectivity. All sections were simplified dramatically. Piloting the system in this manner has enabled us to render it a more objective instrument, improving inter-rater reliability and has allowed us to address content validity.

Following piloting of the evaluation system, scoring was weighted to allow each area of evaluation to have similar weight for a median score (Figure 1). Usability scores were the composite of readability and ease of navigation, with readibility, which is key to usability, given greater weight. Because reading grade levels are inversely related to readibility (i.e., the higher the reading grade level, the lower the readibility) and the number of links needed to locate useful material is likewise inversely related to ease of navigation (i.e., the greater the number of links needed, the lower the ease of navigation), the following adjustments were made: the grade level (out of 12) was multipled by a factor of 0.2, and then subtracted from 4. The number of links needed was multiplied by a factor of 0.1, and then subtracted from 1. This formula allowed for the usability score, like the content and quality, to be higher with a lower reading grade level/number of links needed, as well as providing greater relative weighting of readability.

All 30 Web sites were evaluated between April and July 2004. Total scores were calculated for all 30 Web sites for each of the two evaluators. Mean score for evaluator 1 was 7.293 (standard error 0.511) and for evaluator 2 was 8.189 (0.413). Each evaluator's score was assigned scores accordingly to quartiles (i.e., 1 through 4; see Figure 2). Two indices were calculated to assess for interranking reliability.

Cohen's kappa measures concordance between two rankings on the same objects using an ordinal scale. In calculation, differences in scoring are usually weighted according to the differences between the two evaluators. A correlation of 0.6-0.8 is typically considered to be highly significant, with a correlation of 0.8-1.0 almost perfect. The Cohen's kappa for the two evaluators' total scores for all 30 Web sites is 0.7671. Intraclass correlation evaluates the level of agreement between rankings in measurements. The coefficient represents concordance, where 1 is perfect agreement and 0 is no agreement at all. The intraclass correlation for the two evaluators' total scores for all 30 Web sites is 0.7703.

DISCUSSION

The young field of e-health is changing rapidly. As Danaher, McKay, and Seeley (2005) point out in a recent analysis of information architecture design of health behavior Web sites) that despite increasing sophistication, there is as yet no standardized, universally accepted or validated process for designing the structure of health-related Web sites. Despite this, e-health Web sites designed to educate patients with CKD may be a powerful tool in the nephrologists' armamentarium. Currently available and easily accessible, CKD education Web sites present significant limitations as they do not appear to comply with available standards of e-health and are not constructed to be helpful for the majority of end-users who might benefit from them. Despite this, it is possible to identify Web sites that may be more likely to provide a positive educational experience for patient and families.

In previous work, we identified differences in readability and overall e-health Web site quality between nonprofit and industry-owned organizations (Jaffery & Becker, 2004). The present study is the first to our knowledge to attempt an exhaustive evaluation of the quality of CKD education Web sites in multiple realms. Calderon and colleagues identified 12 such sites that fulfilled at least three of six "domains of CKD information" and also found significant limitations of readability (Calderon et al., 2004a, b). While there have been many different criteria used by diverse groups to evaluate Web sites in a variety of specialties and subspecialties, readability is usually included and is nearly universally felt to be too high (high-school level or above) (Berland, Elliott, 2001; Friedman, Hoffman-Goetz, 2004; Griffin, McKenna, 2004; Kusec, Brborovic, 2003). Not surprisingly, Birru, et al. (2004) showed that adults with low health literacy are limited in their ability to gain information from health Web sites.

Our study has several limitations. Only two evaluators performed the Web site evaluations. While correlations were strong, it is possible that different results could be seen with additional evaluations. This study also did not address the issue of languages other than English; this is clearly a potential area for improvement, most notably for Spanish -peaking minorities, a disproportionate number of whom have CKD.

Many of the readily identified sites on the Web are either directed toward health care professionals, rather than consumers, or geared toward specific kidney diseases or distinct CKD complications. Because of this, the present evaluation may be biased against some of the more likely sites to be visited by patients, such as those arising more frequently from common search engines searches, but eliminated from our preestablished exclusion criteria. However, this bias should actually favor the Web sites that were chosen, as they fulfilled criteria more relevant to patients with CKD.

Finally, the present analysis does not specifically address the utility of these sites as educational interventions. While it is tempting to use information such as that found in this evaluation to make recommendations to patients and providers about specific Web site usage, in fact, none of these sites has been proven to enhance patient knowledge of CKD or aid in patient care. Further evaluation will be needed to investigate the relative utility of variably ranked Web sites as educational interventions, as well as strategies for more effectively steering patients and families toward Web sites most useful to them.

ACKNOWLEDGMENT

The authors would like to acknowledge Dr. Arjang Djamali, Dr. Loren Denlinger, and Dr. Bryan Becker for help with manuscript review as well as the anonymous reviewers of IJHISI for their helpful comments and editorial suggestions. Parts of this work were supported by an NIH K12-RR017614 grant (JBJ).

REFERENCES

Bader, JL., & Strickman-Stein, N. (2003). Evaluation of new multimedia formats for cancer communications. *Journal of Medical Internet Research 5(16)*,

Berland, GK., Elliott, MN., Morales, LS., Algazy, JI., Kravitz, RL., Broder, MS., et al. (2001). Health information on the Internet: accessibility, quality, and readability in English and Spanish. *Journal of the American Medical Association 285*, 2612-2621.

Birru, MS., Monaco, VM., Charles, L., Drew, H., Njie, V., Bierria, T., et al.(2004). Internet usage by low-literacy adults seeking health information: an observational analysis. *Journal of Medical Internet Research 6(25).*

Calderon, JL., Zadshir, A., and Norris, K. (2004). Structure and content of chronic kidney disease information on the World Wide Web: barriers to public understanding of a pandemic. *Nephrology News and Issues 18(76)*, 78-79, 81-74.

Calderon, JL., Zadshir, A., & Norris, K. (2004). A survey of kidney disease and risk-factor information on the World Wide Web. *Medscape General Medicine 6(3).*

Celler, BG., Lovell, NH., & Basilakis, J. (2003) Using information technology to improve the management of chronic disease. *The Medical Journal of Australia 179*, 242-246.

Chelf, JH., Deshler, AM., Thiemann, KM., Dose, AM., Quella, SK., & Hillman, S. (2002). Learning and support preferences of adult patients with cancer at a comprehensive cancer center. *Oncology Nursing Forum 29.* 863-867.

Chernecky, C., Macklin, D., & Walter, J. (2006). Internet design preferences of patients with cancer. *Oncology Nursing Forum 33*, 787-792.

Chou, MH., Lin, MF., Hsu, MC., Wang, YH., & Hu, HF. (2004). Exploring the self-learning experiences of patients with depression participating in a multimedia education program. *The Journal of Nursing Research 12*, 297-306.

Coresh, J., Astor, BC., Greene, T., Eknoyan, G., & Levey, AS. (2003). Prevalence of chronic kidney disease and decreased kidney function in the adult US population: Third National Health and Nutrition Examination Survey. *American Journal of Kidney Diseases 41*, 1-12.

Croft, DR., & Peterson, MW. (2002) *An evaluation of the quality and contents of asthma education on the World Wide Web.* Chest 121: 1301-1307.

Danaher, BG., McKay, HG., & Seeley, JR: The information architecture of behavior change websites. *Journal of Medical Internet Research 7(12)*, 2005

Department of Obstetrics and Gynaecology TC-UoHK: Home Page. URL:. http://department.obg.cuhk.edu.hk/index.asp [accessed 2005 Mar 23]

Eysenbach, G. (2000) Consumer health informatics. *British Medical Journal* 320: 1713-1716.

Eysenbach, G., & Kohler, C. (2002) How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *British Medical Journal 324,* 573-577.

Fox, S., & Rainie, L. (2002). E-patients and the online health care revolution. Physician Exec 28: 14-17. URL: http://www.pewinternet.org [accessed 2003 Dec 2002]

Friedman, DB., Hoffman-Goetz, L., Arocha, JF. (2004) Readability of cancer information on the internet. *Journal of Cancer Education 19.* 117-22.

Gattoni, F., and Sicola, C. (2005) How to evaluate the quality of health related websites. *La Radiologia Medica 109,* pp. 280-287.

Griffin, E., McKenna, K., & Worrall, L. (2004) Stroke education materials on the World Wide Web: an evaluation of their quality and suitability. *Topics in Stroke Rehabilitation 11,* 29-40.

Gustafson, DH., Hawkins, R., Boberg, E., Pingree, S., Serlin, RE., Graziano, F., et al. (1999) Impact of a patient-centered, computer-based health information/support system. *American Journal of Preventive Medicine 16*, 1-9.

Gustafson, DH., McTavish, FM., Boberg, E., Owens, BH., Sherbeck, C., Wise, M., et al. (1999) Empowering patients using computer based health support systems. *Quality in Health Care 8*, 49-56.

Impicciatore, P., Pandolfini, C., Casella, N., & Bonati, M. (1997) Reliability of health information for the public on the World Wide Web: Systematic survey of advice on managing fever in children at home. *British Medical Journal 314,* 1875-1879.

Jaffery, JB., & Becker, BN. (2004) Evaluation of eHealth web sites for patients with chronic kidney disease. *American Journal of Kidney Diseases 44*, 71-76.

Kim, P., Eng, TR., Deering, MJ., & Maxfield, A. (1999) Published criteria for evaluating health related web sites: review. *British Medical Journal 318*, 647-649.

Kusec, S., Brborovic, O., & Schillinger, D. (2003) Diabetes websites accredited by the Health On the Net Foundation Code of Conduct: readable or not? *Studies in Health Technology and Informatics 95*, 655-660.

Kincaid, JP., FR, Rogers, RL., Chissom, BS. (1975) *Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy enlisted personnel*. *Research Branch report* 8-75. Memphis: Naval Air Station.

Marill, JL., Miller, N., & Kitendaugh, P. (2006) The MedlinePlus public user interface: studies of design challenges and opportunities. *Journal of the Medical Library Association 94,* 30-40.

National Kidney Disease Education Program Chronic Kidney Disease Compendium. *Online Resources About Chronic Kidney Diseas*e URL: http://www.nkdep.nih.gov/pdf_ppt/1OnlineResources.pdf [accessed 2004 Mar 2003],

National Kidney Foundation Kidney Disease Outcome Quality Initiative (K/DOQI). *Clinical Practice Guidelines* URL: http://www.kidney.org/professionals/kdoqi/guidelines.cfm [accessed: 2004 Mar 2003],

Paasche-Orlow, MK., Taylor, HA., & Brancati, FL. (2003) Readability standards for informed-consent forms as compared with actual readability. *New England Journal of Medicine 348,* 721-726.

Robinson, TN., Patrick, K., Eng, TR., & Gustafson D. (1998) An evidence-based approach to interactive health communication: a challenge to medicine in the information age. Science Panel on Interactive Communication and Health. *Journal of the American Medical Association 280*, 1264-1269.

Trivedi, HS., Pang, MM., Campbell, A., &Saab, P. (2002) Slowing the progression of chronic renal failure: economic benefits and patients' perspectives. *American Journal of Kidney Diseases 39*, 721-729.

United States Renal Data System, Minneapolis, USA. URL: http://www.usrds.org [accessed 18 Nov 2003], 2003

Winker, MA., Flanagin, A., Chi-Lum, B., White, J., Andrews, K., Kennett, RL., et al. (2000) Guidelines for medical and health information sites on the internet: principles governing AMA web sites. *Journal of the American Medical Association 283*, 1600-1606.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 2, Issue 1, edited by J. Tan, pp. 36-43, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 3.29
Evaluation of a Fuzzy Ontology-Based Medical Information System

David Parry
Auckland University of Technology, New Zealand

ABSTRACT

Evidence-based medicine (EBM) requires appropriate information to be available to clinicians at the point of care. Electronic sources of information may fulfill this need but require a high level of skill to use successfully. This paper describes the rationale and initial testing of a system to allow collaborative search and ontology construction for professional groups in the health sector. The approach is based around the use of a browser using a fuzzy ontology based on the National Library of Medicine (NLM) Unified Medical Language System (UMLS). This approach may provide high quality information for professionals in the future.

INTRODUCTION

Evidence-based medicine (EBM) (Sackett, Richardson, Rosenberg, & Haynes, 1997) has become increasingly important in the modern healthcare industry. Indeed, the concept of basing practice on evidence is even extending to the software engineering domain (Kitchenham, Dyba, & Jorgensen, 2004). Care that is not based on evidence has become increasingly indefensible from professional, safety, and economic points of view. Electronic access to high quality information can improve the professional knowledge of clinicians (Leung et al., 2003), and is very popular (Westbrook, Gosling, & Coiera, 2004). However, there are a number of difficulties associated with providing high-quality information to support EBM.

Assessing and finding appropriate information is difficult and can be time-consuming. This is partly due to the continuing difficulty users have in navigating the interfaces used by various systems and also because of the lack of training available. Indeed, if the concept of just-in-time information retrieval, as an aid to clinical decision-making at the point of care is to be realised (Gardner, 1997), then complex time-consuming strategies performed by trained users are not possible. Recent work, looking at the usage of the Clinical Information Access programme (CIAP) in New South Wales (Gosling, Westbrook, & Coiera, 2003) has emphasised cultural barriers to using online sources of information in a clinical setting, and this includes a perceived lack of skill in information retrieval by clinicians.

In assessing the usefulness of information sources, a framework to identify the aspects that are important needs to be established. Three dimensions have been identified, including information quality, clinical relevance, and clinical usefulness, based partly on the work of Sackett et al. (1997), and some of the limits used in PubMed and other information sources. The aspects of each dimension are outlined in Tables 1 to 3.

Diversity

Both the users and sources of information are characterised by diversity, and existing examples of information portals reflect this. The CIAP system, described by Moody and Shanks (1999), is particularly interesting as a "top-down" approach to providing evidence at the point of care, that is, the project was driven by the funding authority by the New South Wales health department rather than a "bottom-up" approach driven by clinical units. Having multiple database systems with many different interfaces and means of searching can only increase the obstacles to effective use of these tools. Even the CIAP system has over 40 different, searchable, databases available, each with its own interface, not to mention the individual journals, and tools such as Google.

Table 1. Information quality

Aspect	Comments
Peer-review	World wide Web (WWW) sites as well as journals may now have peer-review in place.
Randomised Controlled Trial (RCT)	This is the gold standard for clinical interventions although many interventions have not been subjected to this process. There are also issues of the quality and power of a trial. In some cases meta- analysis can cause smaller trials to lose credibility.
High citation number	This is more of a rule of thumb than an absolute factor. If the source is frequently cited then it indicates that large numbers of authors have found it relevant. It is perfectly possible that a particularly bad study may have a high citation index, or that the index may be inflated for other reasons such as age of the reference. It is possible to infer that references cited in 'good' documents are more likely to be good themselves but this is dangerous to extend too far.
Recent	This depends on the rate of change of the field. Documents in very active research areas are likely to have a shorter useful life than those in inactive areas.
Significant result	A document containing information that a treatment or diagnostic method is effective, and that this effect is large, is likely to be more useful than one that does not. If there is a traditional treatment that is shown to be ineffective then this also is significant.
Authoritative Source	For electronic sources of information the Health on the Net Code of conduct can give some guidance – otherwise, inclusion by indexes or directories e.g. MEDLINE or Cochrane can lend authority. The author affiliation can be an important issue here. An automated system for "authoritativeness" is described by (Farahat, Nunberg, Chen, & Heylighen, 2002).
Usability	Traditional web usability, for example Neilson's heuristics(Neilson, 2000), and also in terms of technical issues such as plug-ins media etc.

Aside from the differences in professional education — which will influence the use of preferred search terms — along with the clinical usefulness indicators, users may also have fundamental differences in their understanding of the meaning of terms. Then to share understanding of the meaning of search terms has been a driver in the use of ontologies (Noy & McGuinness, 2001), and indeed Musen (2001) assigned ontology use and creation as the central role in medical informatics. A general view of a system to support reaching for useful medical information is illustrated in Figure 1. Key elements include the use of multiple information sources accessed through a single browser, an ontology-supported scheme for query expansion and refinement, and the identification of users as members of a professional group with expertise in particular domains and five levels of expertise based on the Dreyfus (Dreyfus, Dreyfus, & Athanasiou, 1986) classification (novice, advanced beginner, expert).

The next section deals with the methods used to construct a system to see if such an approach is valuable. The third section describes the case study and prototype usability testing. The fourth section includes the results of the evaluation, and

Table 2. Clinical relevance

Aspect	**Notes**
Human	Although animal studies, or theoretical ones may be of great use – for example in the case of poisoning or electric shock, human studies are often essential
Correct Sex	Included in this is whether the interventions are safe for pregnant women, and the variation in body sizes and compositions between the sexes, along with other issues related to gender.
Age group relevance	Various age bands are used, or bands that reflect characteristics of the individual rather than his or her age.
Speciality is appropriate	Information designed for one medical specialty may not be appropriate for others, for example between pathologists and other clinicians. Similarly the information requirements of different clinical groups e.g. Physiotherapists and Surgeons treating a patient with an artificial hip may have different needs.
Appropriate language	Is this information in a suitable language for use by clinicians, or is it designed for lay people? The requirements for precision and readability will vary according to the intended audience.

Table 3. Clinical usefulness

Aspect	**Comments**
Appropriate to stage of encounter (e.g. therapy, diagnosis, etc.)	This also excludes information that is purely of a research nature, if better information for the clinical decision is available. However such information can be useful if it casts doubt on current clinical practice, or can help explain otherwise unexpected results.
Deals with available tools	This includes such aspects as whether the drugs or procedures involved are licensed or available in the location, and acceptable in terms of cultural factors and cost.
Suitable format	Are the documents or information sources able to be read by the user; correct language, is a machine reader available. Concrete example of this includes different varieties of microfiche, or PDF files that may require large bandwidths for download.
Available in a timely fashion	Broadly the information may be available immediately (read off the screen- a time period of seconds), quickly (within the library or searching area - a time period of minutes), after a short pause (if documents need to be retrieved from a nearby site- a time period of hours) or after a long time (if the document needs to be specially ordered or generated- a time period of days)
Useful for exclusion	I.e. the information source confirms that a potential diagnosis or treatment is not correct

Figure 1. The overall system

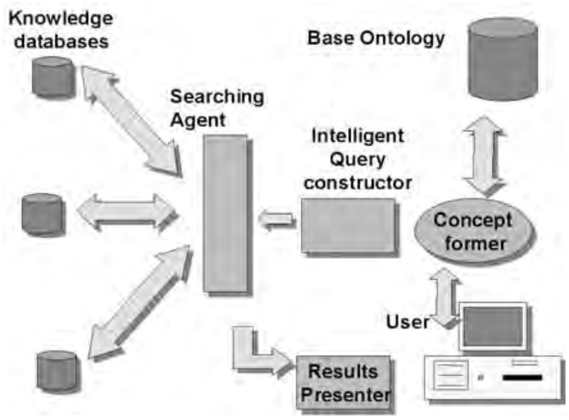
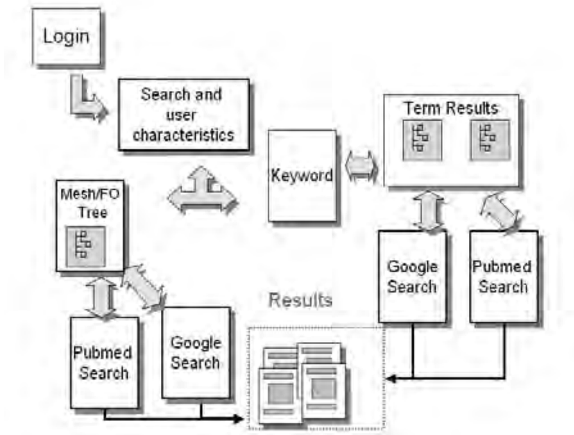


Figure 2. The system components



the final section discusses the significance of such an approach.

METHODS

A prototype system was designed, and tested for usability and usefulness.

System Design

The components of the prototype are illustrated in Figure 2. The case study was performed using a prototype version of the system, built in Visual Basic.NET, using an SQL server 2000 as the database (Microsoft). The Google application program interface (Google, 2004) was used along with the entrez eUtilities of PubMed (National Library of Medicine, 2003) to provide two data sources. These are part of an increasing number of Web Services that are being made available via Simple Object Access Protocol (SOAP) that allows XML-based communication and service control over the Internet.

Fuzzy Ontology

Ontologies are extremely important in medical informatics (Musen, 2001), but multiple ontologies (Noy & Musen, 2004) can be difficult to maintain and combine. In addition, terms are often multiplying within ontologies, thus making the mapping between a query term and the intended location in an ontology difficult.

The concept of a "fuzzy ontology" was introduced in Parry (2004b). Effectively, this approach reuses a current ontology, in this case, the MeSH hierarchy (U.S. National Library of Medicine, 2001), and assigns a particular membership value to each multiple-occurring term in each location.

Table 4 demonstrates how these issues arise in existing ontologies. "Pain" occurs in five locations in the MeSH ontology. Because the term is located in a number of different places, query expansion for this term is difficult, because there is a wide range of numbers of "related" terms. For this reason, the MeSH hierarchy was adapted in order to allow users to assign membership values to term, location pairs via a machine learning system described in Parry (2004a). The case study was designed partly to investigate different methods of learning these relations, but all of the searches were performed using the original MeSH hierarchy terms without the use of fuzzy, ontology support.

Many issues arise from the use of multiple ontologies, including the difficulties associated

Table 4. Multiple occurrence examples —"Pain"

Term	Concept ID	Parent	Depth	Root term
Pain	G11.561.796.444	Sensation	4	Musculoskeletal, Neural, and Ocular Physiology
Pain	F02.830.816.444	Sensation	4	Psychological Phenomena and Processes
Pain	C23.888.646	Signs and Symptoms	3	Pathological Conditions, Signs and Symptoms
Pain	C23.888.592.612	Neurologic Manifestations	4	Pathological Conditions, Signs and Symptoms
Pain	C10.597.617	Neurologic Manifestations	3	Nervous System Diseases

with communicating between ontologies and the need for maintenance of large numbers of ontologies. The fuzzy ontology as described is partly suggested in order to allow a common framework, or base ontology, with different membership values associated with different users and groups. It should be noted that because of the learning methods involved, only "is-a" type relations are currently used, based on the currently existing MeSH hierarchy.

Another advantage of this approach is completeness. Rather than impose an arbitrary standard of the importance of a particular location in the ontology, which is required in a crisp ontology to avoid too many examples of a term appearing in the ontology, the term or object can be located in all relevant locations

Most importantly, for searching processes, the use of a fuzzy ontology for the mapping of search terms allows the relative weight of each term in the required output to be calculated. By allowing these weights to be calculated accurately, it removes the bias associated with multiple-located terms being used for searching. If a term is located in multiple locations in a crisp ontology, and is used for query expansion purposes, say by including offspring, then the danger is that the large number of relatively irrelevant expansion terms outweigh those which are useful.

In particular, the use of a fuzzy ontology approach allows the convenient representation of the relationships in a domain according to a particular view without sacrificing commonality with other views; the ontology framework is common, just the membership values are different.

Finally, this approach holds out the possibility that the representation of a potentially very large ontology, can be compressed. If whole areas are not required, the relations to the core can be set to zero. Unwanted intermediate levels can also be removed, with lower-level terms only communicating directly with higher levels. This aspect removes the need to create artificial groupings to avoid orphaned terms. At the limit, a fuzzy ontology, with all membership values = 0 or 1, will have each term or object located in one location only and will behave in exactly the same way as a crisp ontology. A scheme for visually describing the fuzzy ontology is shown in Figure 3.

CASE STUDY

The setting was an academic department of Obstetrics and Gynecology, and only elements of the MeSH tree relevant to this domain were included. Ethical approval was obtained, and eight users were allocated an hour each to use the system. During user number 8's study, the database was corrupted, and the subject was unable to complete

Figure 3. The fuzzy ontology

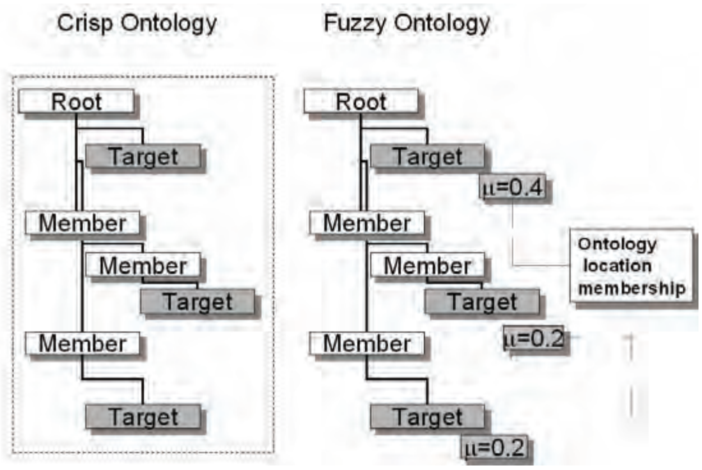


Table 5. Comparison of crisp and fuzzy ontologies

Aspect	Fuzzy ontology	Crisp ontology
Multiply-located terms	Does not occur	Issue for disambiguation
Query expansion	Depends on membership value.	Depends on location only
Customisation	Simple, based on modification of membership values	Requires new ontology and/or ontology sharing.
Intermediate locations for grouping	Unnecessary	Needed for construction – may be useful
Storage required	Depends on the number of terms in the ontology and the membership values of the relations, can be smaller or larger than crisp.	Depends on number of terms in the ontology
Knowledge representation	Related to use	Related to structure.

any testing. As such, the subject was therefore not included in any analysis. This is a small number of users but Neilson (2000) points out that usability testing can often be successful with small groups. In fact, the whole department only comprised less than 10 faculty staff; as with many systems designed for professional use, the pool of users is actually quite small. Results of their searches were presented in the results browser shown in Figure 4.

The users were asked to perform the following tasks:

1. Log into the system and select the appropriate demographic and area of interest.
2. Perform a search using the "obstetric" keyword on the Google interface. This was done mostly to familiarise the users with the system, in particular, the appropriate use of the mouse and the use of the + anchor in lists to expand them, as in Windows Explorer.
3. They then performed another search using terms of their own choosing, again using the Google interface.
4. With the open browser windows, they were asked to rate a number of the pages shown

Figure 4. The results browser

in terms of usefulness via the slider. They were also asked to perform an analysis on pages that they rated highly. In most cases, this amounted to around five pages.

5. They were then asked to perform a similar task using the full MeSH tree.

RESULTS

The results of the study are based around the usability questionnaire responses and user comments. The questionnaire used in the study was adapted from one of those generated from the site provided by Perlman (2001). This questionnaire was originally reported in Davis (1989) and has subsequently been used in a number of studies. This questionnaire focuses on the use of a system for work-related tasks, and the scale runs from -2 to +2, to allow a 0 for neutrality. All of the questions are phrased so that a positive result implies satisfaction with the system. One of the most interesting aspects of this questionnaire is that it specifically links ease of use and usefulness. The original work suggested that increased perceived ease of use has a causal influence on perceived usefulness. However, more recent work (Segars & Grover, 1993) appears to suggest that this analysis is not complete. It is suggested that, in turn, an information system that is perceived as useful must be retrieving useful information.

Overall, the perceived usefulness was rated as X = 1.16 (SD 1.21), and the perceived ease of use at X = 1.53 (SD 0.29). Ease of use could be expected to rate more highly as the situation for testing was somewhat artificial.

Comments about the system were recorded, and general satisfaction seemed quite high. Of particular interest was the ease of use of the analysis system, despite the fact that there were bugs in this version, which allowed duplicate words to occur in the pick list. There was certainly a preference towards identifying positive (very, somewhat relevant) rather than negative (irrelevant, unwanted) words. The users preferred to analyse those documents they found useful, and tended to ignore those they found useless.

One aspect of particular benefit was the presentation of the derived MeSH keywords, which allowed a user to reconsider his or her search before it began. General observations of users included the fact that they found dealing with large numbers of windows a little confusing. By attempting to improve visibility, the use of multiple windows tended to remove the obvious focus. Mouse movement became more uncertain when

Table 6. User group details

Number	Job description	Professional group	Computer experience	Gender	Age Range
1	Senior Academic	Doctor, interest in MFM	Moderate	Male	50+
2	Senior Academic	Doctor interest in MFM	Moderate	Female	50+
3	Junior Academic	Doctor, interest in REI	High	Female	30+
4	Research Midwife	Midwife background, clinical researcher	Moderate	Female	50+
5	New Consultant	Doctor, General Obstetrics and Gynaecology	Moderate	Female	30+
6	Senior Academic	Doctor, interest in Contraception	High	Female	40+
7	Junior Academic	Doctor, interest in Infertility	Moderate	Female	30+
8	New Consultant	Doctor, General Obstetrics and Gynaecology	Moderate	Female	30+

there were overlapping windows, and the users were often uncertain as to the difference between closing and minimising windows. In many cases, the users maximised the active window.

One of the recurring themes was the uncertainty of whether such a system was primarily for medical professionals or for patients. When browsing the documents recovered via Google, the users were sometimes surprised to find what they regarded as legitimate medical pages among the obviously patient-centred ones. This is an unexpected benefit of using multiple search engines — multiple search strategies are used simultaneously. Various meta-engines already use this approach, but they currently do not appear to use non-commercial data sources such as PubMed.

DISCUSSION

Finding and applying appropriate information is one of the key tasks of the knowledge worker (Kidd, 1994). There exists a vast body of knowledge in electronic form for workers and patients in the health sector. However, finding appropriate knowledge is difficult and time-consuming. Fears of inappropriate information being provided abound (Eysenbach, 2002). In order to fully realize the potential benefits of electronic knowledge sources, the sources must be appropriate for their use and usable by the potential beneficiaries. Understanding the knowledge requirements of users in this domain and providing appropriate tools for such users remain great challenges for informatics professionals. This paper has attempted to set up a framework for future research in the area of appropriate knowledge sources based around a user perspective. The importance of delimiting different user groups within the health sector has also been identified. In addition, a prototype system for combining knowledge from different sources in an integrated way has been tested for usability and potential usefulness. The challenges of using diverse information sources from the Web have been raised in Allan et al. (2003), and this area remains a particularly important area of information retrieval research. Other work has

Table 7. Initial group satisfaction

Question	User 1	User 2	User 3	User 4	User 5	User 6	User 7	Mean
Perceived Usefulness								
1 (Quick)	0	2	1	1	1	2	2	**1.26**
2 (Performance)	0	1	0	0	1	2	2	**0.86**
3 (Productivity)	2	1	1	0	1	2	2	**1.29**
4 (Effectiveness)	1	1	0	0	1	2	2	**1.00**
5 (Easier)	1	1	0	1	1	2	2	**1.14**
6 (Useful)	1	1	2	1	1	2	2	**1.43**
Perceived Ease of Use								
7 (Easy to Learn)	2	1	2	2	2	2	2	**1.86**
8 (Easy to Control)	-1	2	1	2	2	1	2	**1.29**
9 (Clear Interact)	2	2	1	2	1	1	2	**1.57**
10 (Flexible)	-1	2	1	2	1	1	2	**1.14**
11 (Skill)	2	2	2	2	1	2	2	**1.86**
12 (Easy to Use)	0	2	1	2	1	2	2	**1.43**

been done recently on the usability of medical information sources (Alexander, Hauser, Steely, Ford, & Demner-Fushman, 2004), and improvements are certainly possible. The results of this study suggest that an integrated knowledge discovery system for a medical professional is desirable and that the prototype represents a useful start in this direction. The results for ease of use compare favourably with similar scores in the technology acceptance model, that is, in Henderson and Divett (2003), dealing with electronic shopping, where ease of use was 1.25 and usefulness was 0.96 when converted to the same scale as used in this work. It is hoped that further research in this area will continue, in particular, in the following areas; the replacement of the executable form of the system with a browser-based client server system that will allow much larger user groups to interact with it and the provision of a substantial base for learning about group preferences. Mobile and wireless information retrieval may be more appropriately integrated into clinical workflow especially by means of "information appliances" (Eustice, Lehman, Morales, Munson, Edlund, & Guillen, 1999), recent work in this area (Burdette, Herchline, & Richardson, 2004) suggests that these devices may be particularly suitable for hospital use. The integration of information sources which provide their data via Web Services is also rapidly becoming accepted in the world of digital libraries (Fu & Mostafa, 2004). Integrating information from diverse sources via ontologies is also becoming increasingly important especially in the context of the "Semantic Web" (Berners-Lee, Hendler, & Lassila, 2001). The Ontology Web Language (OWL) (Smith, Welty, & McGuinness, 2004) could also be modified to support a fuzzy ontology, and it has already been recognised that storage of such ontologies on the Web can allow effective knowledge sharing (Haarslev, Lu, & Shiri, 2004). Finally, more research needs to be undertaken in the use and standardization of aspects of information reliability, usefulness, and relevance to improve research and classification in this area especially from the perspective of the clinical worker.

REFERENCES

Alexander, G., Hauser, S., Steely, K., Ford, G., & Demner-Fushman, D. (2004, September 7-11). A usability study of the PubMed on tap user interface for PDAs. In M. Fieschi, E. Coiera, & Y.-C.J. Li (Eds.), *Proceedings of the 11th World Congress on Medical Informatics, Medinfo 2004,* San Fransisco (pp. 1411-5). Amsterdam: IOS Press.

Allan, J., Aslam, J., Belkin, N., Buckley, C., Callan, J., Croft, B., et al. (2003). Challenges in information retrieval and language modeling. *SIGIR Forum, 37*(1), 31-47.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001, May). The Semantic Web. *Scientific American*, 29-37.

Burdette, S., Herchline, T., & Richardson, W. S. (2004). Killing bugs at the bedside: A prospective hospital survey of how frequently personal digital assistants provide expert recommendations in the treatment of infectious diseases. *Annals of Clinical Microbiology and Antimicrobials, 3*(1), 22.

Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340.

Dreyfus, H. L., Dreyfus, S. E., & Athanasiou, T. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer.* Oxford: Blackwell.

Eustice, K. F., Lehman, T. J., Morales, A., Munson, M. C., Edlund, S., & Guillen, M. (1999). A universal information appliance. *IBM Systems Journal, 38*(4), 575-601.

Eysenbach, G. (2002). Infodemiology: The epidemiology of (mis)information. *The American Journal of Medicine, 113*(9), 763-765.

Farahat, A., Nunberg, G., Chen, F., & Heylighen, F. (2002). AuGEAS: Authoritativeness grading, estimation, and sorting: Collective Intelligence and its Implementation. *Computational & Mathematical Organization Theory, 5*(3), 194-202.

Fu, Y., & Mostafa, J. (2004, June 7-11). Toward information retrieval Web Services for digital libraries In *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries,* Tucson, AZ (pp. 370-371). Tuscon, AZ: ACM Press.

Gardner, M. (1997). Information retrieval at the point of care. *British Medical Journal, 314*, 950-953.

Google. (2004, 2005). *Google Web API.* Retrieved on July 1, 2004, from http://www.google.com/apis/index.html

Gosling, A. S., Westbrook, J. I., & Coiera, E. W. (2003). Variation in the use of online clinical evidence: A qualitative analysis. *International Journal of Medical Informatics, 69*(1), 1-16.

Haarslev, V., Lu, Y., & Shiri, N. (2004, September 20-24). ONTOXPL — Intelligent exploration of OWL ontologies. Paper presented at the *Web Intelligence, 2004 (WI 2004) Proceedings of the IEEE/WIC/ACM International Conference,* Beijing, China (pp. 624-627). Washington, DC: IEEE Computer Society.

Henderson, R., & Divett, M. J. (2003). Perceived usefulness, ease of use and electronic supermarket use. *International Journal of Human-Computer Studies, 59*(3), 383-395.

Kidd, A. (1994). *The marks are on the knowledge worker.* Paper presented at the Human Factors in Computing Systems: Celebrating Interdependence, Boston (pp. 186-191). New York: ACM.

Kitchenham, B. A., Dyba, T., & Jorgensen, M. (2004). *Evidence-based software engineering* (pp. 273-281).

Leung, G. M., Johnston, J. M., Tin, K. Y. K., Wong, I. O. L., Ho, L.-M., Lam, W. W. T., et al. (2003). Randomised controlled trial of clinical decision support tools to improve learning of evidence based medicine in medical students. *BMJ, 327*(7423), 1090-1090.

Moody, D., & Shanks, G. (1999, December 1-3). *Using knowledge management and the Internet to support evidence based practice: A medical case study*. Paper presented at The 10th Australasian Conference on Information Systems, Victoria University Wellington (pp. 660-667). New Zealand: Victoria University of Wellington.

Musen, M. (2001, September 2-5). *Creating and using ontologies: What informatics is all about*. Paper presented at the Medinfo 2001, London (p. 1514). Amsterdam: IOS Press.

National Library of Medicine. (2003, July 21). *Esearch entrez utility*. Retrieved October 2003, from http://eutils.ncbi.nlm.nih.gov/entrez/query/static/esearch_help.html

Neilson, J. (2000). *Designing Web usability*. Indianapolis, IN: New Riders.

Noy, N. F., & McGuinness, D. (2001). *Ontology development 101: A guide to creating your first ontology*. Palo Alto, CA: Stanford University.

Noy, N. F., & Musen, M. A. (2004). Ontology versioning in an ontology management framework. *Intelligent Systems and Their Applications, IEEE [see also IEEE Intelligent Systems], 19*(4), 6-13.

Parry, D. (2004a, November 1-3). *Fuzzification of a standard ontology to encourage reuse*. Paper presented at The 2004 IEEE International Conference on Information Reuse and Integration (IEEE IRI-2004), Las Vegas, NV (pp. 582-587). Washington, DC: IEEE.

Parry, D. (2004b). A fuzzy ontology for medical document retrieval. In M. Purvis (Ed.), *The Australasian workshop on data mining and Web intelligence* (Vol. 32, pp. 121-126). Dunedin: Australian Computer Society.

Perlman, G. (2001). *Web-based user interface evaluation with questionnaires*. Retrieved on January 3, 2003, from http://www.acm.org/~perlman/

Sackett, D., Richardson, W., Rosenberg, W., & Haynes, B. R. (1997). *Evidence based medicine — How to practice and teach EBM*. Edinburgh, UK: Churchill Livingstone.

Segars, A., & Grover, V. (1993). Re-examining perceived ease of use and usefulness: A confirmatory factor analysis. *MIS Quarterly, 17*(4), 517-526.

Smith, M. K., Welty, C., & McGuinness, D. L. (2004, October 2). *OWL Web ontology language guide*. Retrieved on March 1, 2004, from http://www.w3.org/TR/2004/REC-owl-guide-20040210/#OntologyMapping

U.S. National Library of Medicine. (2001, November 15). *Medical subject headings*. Retrieved on January 11, 2004, from http://www.nlm.nih.gov/mesh/

Westbrook, J., Gosling, A. S., & Coiera, E. (2004). Do clinicians use online evidence to support patient care? A study of 55,000 clinicians. *Journal of the American Medical Informatics Association, 11*(2), 113-121.

This work was previously published in the International Journal of Healthcare Information Systems and Informatics, Vol. 1, Issue 1, edited by J.Tan, pp. 40-51, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 3.30
Enhancing Cognitive Screening in Geriatric Care:
Use of an Internet-Based System

Peter A. Lichtenberg
Wayne State University, USA

Amanda Schafer Johnson
Wayne State University, USA

David M. Erlanger
Headminder™ Corporation, USA

Tanya Kaushik
Headminder™ Corporation, USA

Michael E. Maddens
William Beaumont Hospital, USA

Khaled Imam
William Beaumont Hospital, USA

Jeffrey Barth
University of Virginia School of Medicine, USA

Frank M. Webbe
Florida Institute of Technology, USA

ABSTRACT

Cognitive screening measures for age-related cognitive impairment have been found to have only fair validity, and the risks of harm even may outweigh the benefits at this time (U.S. Preventative Service Task Force, 2003). A large-scale project designed to assess elder care in Primary Care Physician offices noted that dementia evaluation and treatment was one of the most overlooked aspects of care. Taken together, these studies cited the lack of time and technical expertise in test administration as the most prominent barriers to the accurate detection of dementia in Primary Care Physician offices. It was for these reasons that the Cognitive Screening Test (CST) was created. The CST requires no physician time or training to administer or interpret it. The current study investigated the clinical utility of this cognitive screening system by comparing the results of 102 patients to those of expert geriatricians, using consensus conference methods for diagnosis. Overall clinical utility demonstrated scores at .80 or above for sensitivity, specificity, and positive and negative predictive power. In contrast, the MMSE had only a .38 sensitivity. A Receiver Operating Curve (ROC) analysis indicated a .863 accuracy rating for the predetermined cut score on the CST.

VALIDATION OF A DEMENTIA SCREENING TEST IN GERIATRIC PRACTICE

The growing prevalence of dementia makes it imperative that Primary Care Physicians become more able and more comfortable to diagnose dementia, particularly in its earlier stages. Early diagnosis gives physicians the opportunity to initiate treatment with acetylcholinesterase inhibitors, which have shown to delay cognitive and functional declines (Mohs et al., 2001) and to delay nursing home placement (Schneider, 2000; Turner, 2003), all the while producing infrequent, temporary side effects (Boise, Morgan, Kaye,& Camicioli, 1999). It allows patients and families to plan for future care needs, financial needs, and legal needs associated with a lengthy, debilitating illness (Boise et al., 1999; Knopman, Donohue,& Gutterman, 2000), and families and patients may be given insight into the behavioral and personality changes that may occur and the safety precautions that eventually may be necessary (Knopman, et al., 2000; Sano & Weber, 2003). On the whole, primary care physicians continue to have difficulty detecting the majority of dementia cases among their patients. In previous studies, less than 30% of patients have been correctly identified as cognitively impaired, including even those with poor cognitive performance (Boise, Neal, & Kaye, 2004; Olafsdottir, Skoog, & Marcusson, 2000; Valcour, Masaki, & Blanchette, 2002). In an effort to increase the detection of cognitive impairment and generally to increase the quality of patient care, variables have been identified that impede this process. In a review, Reuben, Roth, Kamberg, and Wenger (2003) identified physician time constraints and a lack of physician technical expertise with cognitive screening administration and interpretation as major factors that impede effective dementia screening. Additionally, compounding the issue is that the standard primary care model of medical practice relies heavily on the patient's symptoms report. However, lack of awareness of symptoms is a common feature of Alzheimer's disease; thus, even when queried directly, patients often deny or minimize cognitive problems (Zanetti et al., 1999).

Conducting and interpreting brief cognitive screening is problematic in Primary Care Practices, even in practices where screening tests such as the Mini-Mental Status Examination (MMSE) are used. It is well established that the MMSE is sensitive to the effects of education and age (Anthony, LeResche, Niaz, Von Korff, & Folstein, 1982; Crum, Anthony, Bassett, & Folstein, 1993; Monsch et al., 1995; Spreen & Strauss, 1998; Tombaugh & McIntyre, 1992; Tombaugh, McDowell, Krisjansson, & Hubbley, 1996; Uhlmann & Larson, 1991). As a result of these findings, normative data have been published that make such considerations (Crum et al., 1993; Tombaugh et al., 1996). However, interpretation that considers age and education rarely occurs in clinical practice. More often than not, single-cut scores are used as a basis to distinguish the impaired from intact patients; for example, a cut score of 23/24 on the MMSE has been suggested to indicate cognitive impairment (Folstein, Folstein, & McHugh, 1975). This cut score also has been found appropriate in population-based studies (Ganguli et al., 1993). However, depending on the age and educational characteristics of the patient, this cut score consistently has demonstrated low sensitivity or specificity (Monsch, et al., 1995; Spreen & Strauss, 1998; Tombaugh & McIntyre, 1996). Due to these test constraints, cognitive screening measures for age-related cognitive impairment have been found to have only fair validity, and the risks of harm even may outweigh the benefits at this time (U.S. Preventative Service Task Force, 2003).

Ideally, a screening device for dementia will contain many memory items, since this is the domain of cognition typically impaired; it will require little in the way of administration and interpretation by those not trained in psychometrics and will be based on normative data that are meaningful to clinical interpretation. In response to both the problematic characteristics of traditional screening measures and the qualities desired in screening devices, a computerized screening battery was created. The Cognitive Screening Test (CST), an Internet-based program, is easily administered and requires no interpretation by the administrator. Conceptually, CST may provide a superior method of screening for cognitive impairment, because it instantly can take into consideration the factors that have shown to significantly affect screening interpretation in order to maximize sensitivity and specificity. Thus, even given equal qualities of the test instrument itself, an Internet-based system has the advantage of being able to utilize corrections for variables such as age and education in sophisticated ways. The purpose of this is to present preliminary data analyses using the CST system in order to compare its validity to other traditional screening devices and to demonstrate other qualities of the CST that lend to its incremental validity over other devices.

Two hypotheses were proposed for the study:

1. The CST will demonstrate a high level of clinical agreement (over .8 on combined indices of sensitivity, specificity, and positive and negative predictive power) with expert physician consensus diagnoses.
2. The CST will demonstrate superior clinical utility data than will the MMSE using a single cut score.

METHODS

Participants and Clinic Procedures

The sample consisted of 102 patients from a specialized geriatric clinic; characteristics of the sample are listed in Table 1. The clinic is located in a suburban setting and typically attracts European American patients who have above-average educational levels. The attending physicians at this clinic are all geriatric fellow trained and board certified. The only patients who were not approached about the study were those who had significant motoric, visual, or cognitive impairments that prevent them from using a computer or understanding directions, and individuals who were not fluent in English.

There are two groups of patients at this clinic, both of which were represented in this study. The first includes those who have established the clinic for primary medical care; the second are individuals who have been referred to the geriatric

Table 1. Sample demographics

	n	Mean/Percent	SD
Age (years)	102	79.3	6.6
Education (years)	102	13.5	2.9
Ethnicity:			
African American	6	5.9%	
European American	96	94.1%	
Gender:			
Male	47	46.1%	
Female	55	53.9%	

memory clinic by their own physicians in order to address specific geriatric issues. Evaluation of all patients is always very thorough. When first admitted to the clinic, there is a comprehensive review of symptoms and a physical examination, including a detailed neurological examination. This would be repeated for primary care patients if cognitive status were in question. Medication review and an informant interview are conducted. Medical, family, social, and occupational histories are taken. Functional assessment, laboratory tests, and neuroimaging are part of the clinical workup for cases of suspected dementia. Included in the screening of patients who are screened for cognitive decline is the Mini-Mental State Examination.

MEASURES

CST Normative Sample and Scoring Rules

The CST norms and cut scores were derived from normative data collected completely independently of the current study. The characteristics of the normative data were a sample size of 105 with a mean age of 73 (s.d.=7) and a mean educational level of 12 years .Seventy percent of the normative sample was made up of non-Hispanic whites, and 30% was minority elders (primarily African Americans).The normative data were employed in developing the cut scores as follows. Multiple regression analyses were conducted for each subtest of interest (learning total, memory total, executive function—total errors and mean response time), with subtest performance as the dependent variable and many potentially significant determinants added as predictors in the equation (e.g., health history variables, demographics, etc.). Of all of the variables included, age, ethnicity, education, and gender alone significantly predicted subtest performance.

Using the normative data beta weights and constants, the significant demographic variables were plugged into the general linear model (a+bx) to derive an expected score for each individual in the current sample. This expected score then was subtracted from the individual's raw score, which was divided by the standard error of the estimate to produce a z-score that describes the patient's deviation from the mean score. An Impairment Index then was derived for each subject as follows: z scores were converted to Impairment Index scores by summing subtest results according to the following: z scores > -2.0 were assigned a subtest score of 0; z scores > -2.5 (but less than -2.0) were assigned a subtest score of .5; z scores < -2.5 were assigned a subtest score of 1. A total score of 1.5 or more (summed across the three subtests) indicates a high probability of cognitive impairment and represents the cut score for impairment in these analyses. See Table 2 for a visual illustration of the translation of z-scores into this impairment index. In summary, the analyses from the cases reported here provide a cross validation of the equations obtained from the normative derivation sample.

Table 2. Translation of z-scores into the cognitive impairment index

Z-Score	Cognitive Impairment Index
-2.0 or greater	0
-2.0 > z > -2.5	0.5
Less than –2.5	1

Cognitive Screening Test (CST)

The CST is composed of several subtests. The following describes each one and explains the rationale for its development. Total administration time is approximately 15 minutes.

- **Keyboard Proficiency Subtests (2 minutes).** The first of these two subtests consists of a green ball presented on the screen. The participant is instructed to press the spacebar repeatedly until the green ball turns red. The second involves a series of numbers appearing on the screen; the participant is instructed to press the number key that corresponds with the number shown. These subtests are administered solely for the purpose of acclimating the patients to the nature of the tests.
- **Learning and Memory Subtests (7-8 minutes).** Individuals learn the placement of nine household objects placed in a virtual cabinet. In the course of three learning trials, they are quizzed about the location of each object. The number of correct responses summed over all three trials is referred to as the Learning subtest. After a delay involving the interpolation of the Executive Function subtest, a final inquiry is made about the location of the objects. This final recall is referred to as the Memory subtest. See Figure 1 for an illustration of this task.
- **Executive Function Subtest (6-8 minutes).** There are two components to this task. For the first, numbers from 0 to 9 are presented one at a time. Individuals are instructed to press the 0 or 1 key when the respective number is presented. Test takers are asked not to respond to other numbers. The second part consists of the numbers again being presented, but the participants are instructed to press the 0 key in response to the appearance of a 1 and to press the 1 key in response to the appearance of a 0. Again, subjects are told not to respond to numbers other than 0 or 1. The second part of this test taps executive function, with its emphasis on self-monitoring, and response inhibition and flexibility (Lezak, 1995). See Figure 2 for an illustration of the Executive Function subtest. Both the number of errors and the mean response time for the second part of this task are used to compute the CST decision.

PROCEDURE

Recruitment and CST Administration

During medical appointments, patients were approached by their physicians about volunteering for the study. Those who agreed to participate then were provided informed consents and ad-

Figure 1. Illustration of the "memory cabinet" used in the memory and learning subtests of the CST

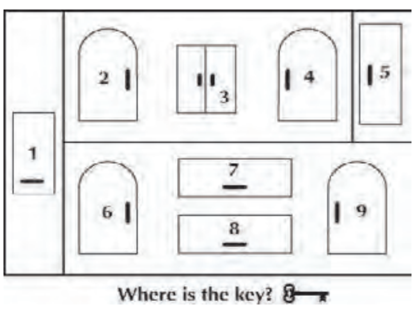


Figure 2. The "executive function" subtest of the CST; this figure illustrates what patients view and which responses are appropriate in each task

Part 1 of Executive Function Subtest *Part 2 of Executive Function Subtest*

ministered the CST protocol by a graduate student in psychology. Administration typically took 15 minutes. A double-blind procedure was employed, as physicians did not disclose participant diagnoses to the CST administrator, nor did the administrator ever inform physicians about their performance. All participants were administered the CST (n=102).

Consensus Diagnosis Conference

Consensus diagnosis conferences by the attending physicians were conducted, during which clinical impressions, lab record reviews, and the MMSE scores were utilized in order to determine diagnoses; physicians were blinded to the patients' performances on the CST. DSM-IV criteria were utilized for diagnosis of dementia (DSM-IV, 1994); those who were diagnosed with Mild Cognitive Impairment (MCI) met Petersen's consensus criteria (Petersen et al., 1999). Using these criteria and methods of assessment, 35.3% of the sample was determined to be demented, 10.8% to have MCI, and 53.9% to be cognitively intact.

Statistical Analyses

In order to determine the concordance between all of the CST and the consensus diagnoses, scores above and below the cut points were plotted, and accuracy was computed manually.

RESULTS

The demographic characteristics of this sample are presented in Table 1. The final sample of 102 participants administered the CST was examined. An overall 83% concordance was found between the geriatrician consensus diagnosis and the CST predetermined cut score (see previous section titled CST Scoring and Decision Rules). Specifically, there was 80% sensitivity, 87% specificity, 88% positive predictive power (ppp), and 79% negative predictive power (npp) (see Table 3). In order to determine further whether the predetermined cut score (Impairment Index of 1.5 summed across subtests) was the most accurate in predicting cognitive status, a receiver operating characteristic (ROC) curve was examined. The predetermined cut score that had been employed did, in fact, produce the highest overall accuracy among the range of possible scores. The area under the curve (AUC) value for all possible accuracy values as determined by the ROC curve was .861 (SE=.039) (see Figure 3). Thus, there was support for hypothesis 1: clinical accuracy of the CST.

In order to compare the clinical utility of the Cognitive Screening Test to another screening test that utilized a single cut score, comparisons were made between the CST and the Mini Mental State Examination (MMSE). Using the single cut-score method, sensitivity of 37.7%, specificity of 97.6%, ppp of 95.2%, and npp of 55.4% were found for the MMSE. Taken together, these indices demonstrate that a single cut score in this clinic population leads to an unacceptable number of false negatives and very poor sensitivity.

Table 3. Screening device accuracy

	Sensitivity	Specificity	PPP	NPP
CST	.80	.87	.88	.79
MMSE	.38	.98	.95	.55

Figure 3. Receiver operating characteristic (ROC) curve for the cognitive impairment index of the CST; Area under the curve (AUC)=.861 (SE=.039), p<.001

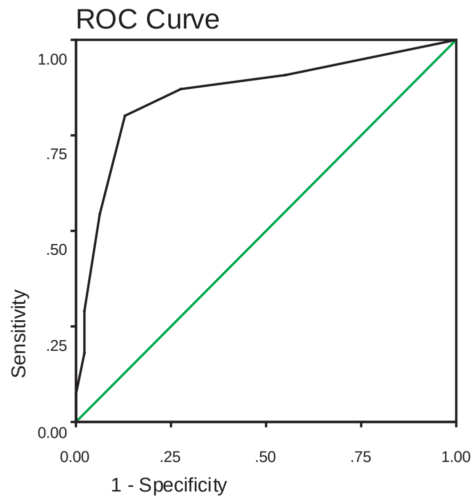


Table 4. Mean scores on cognitive tests by impairment status

	Impairment Status		
	Intact (n=55)	MCI (n=11)	Demented (n=36)
CST Impairment Index Total*†	.76 (.92)	1.82 (1.03)	1.82 (.98)
MMSE†	28.53 (1.44)	25.73 (3.64)	23.63 (4.30)

*Note. *Signifies that the intact and MCI groups are significantly different. †Signifies that the intact and demented groups are significantly different. No tests were significantly different between the MCI and demented groups.*

Figure 4. Mean IADL scores plotted as a function of the cognitive impairment index scores

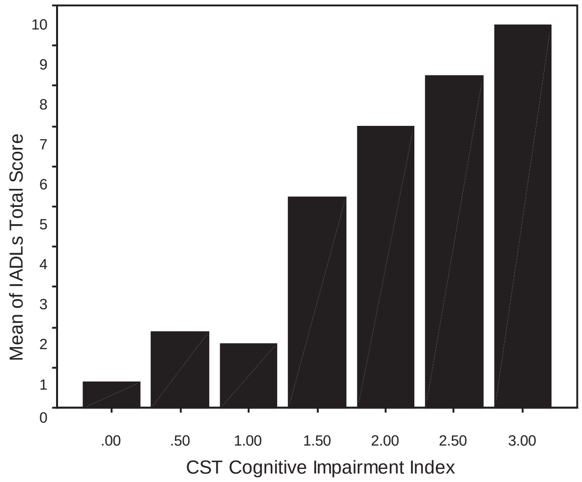
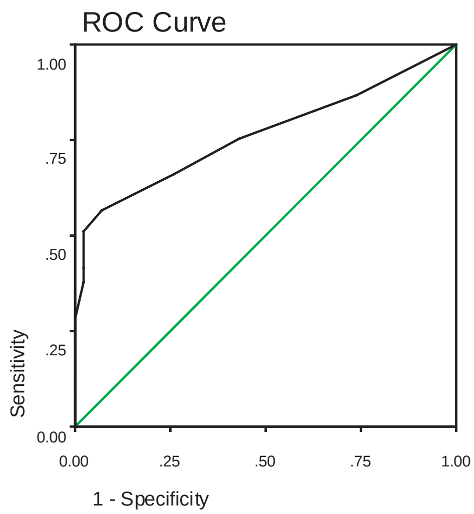


Figure 5. ROC curve for MMSE scores. AUC=.764 (SE=.049), p<.001



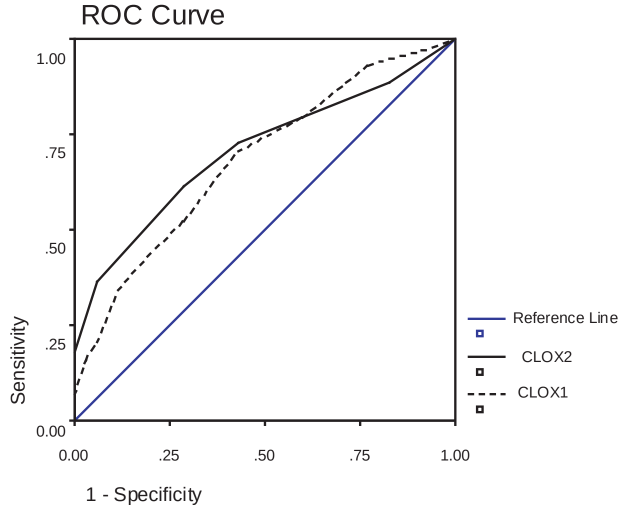
In order to better determine the ability of the CST and MMSE to detect early dementia, an analysis of variance (ANOVA) function was performed among those diagnosed as demented, those diagnosed with Mild Cognitive Impairment (MCI), and those considered cognitively intact by the geriatricians. Significantly different scores were found between those considered intact and those considered demented by all tests. The only test that was significantly different among those who were intact and those with MCI was the CST. A follow-up analysis was conducted among only the MCI patients to determine the effectiveness of the different screens in detecting early dementia. Of the 11 MCI patients, nine (81.8%) were correctly identified by the CST as likely impaired, whereas only three (27.3%) of these were correctly identified by the MMSE.

DISCUSSION

This study demonstrated that the Cognitive Screening Test was administered easily in a physician's clinic setting and had a high rate of concordance with expert geriatrician diagnosis with an overall hit rate of 83%. The exciting implications of these data are related to the use of technology for cognitive screening in a way that reduces clinician burden (i.e., time and expertise) while improving accuracy of the screening tool. As already stated, physicians and other healthcare professionals do not need to be concerned with administering the measure or interpreting results. Studies have reported that physicians may choose not to take part in increased care innovations if they do not feel that they possess the skills to do it correctly (Reuben et al., 2003); thus, a procedure that does not require physicians to learn new skills (i.e., cognitive testing) may more likely be utilized. Additionally, the CST does not take significantly more time than other pencil-and-paper measures and can be administered with little help by a medical technician or other person who can navigate to the Web site and respond to any questions or anxieties of the patient. This translates to little time investment on the part of the physician. Finally, there is no additional equipment to purchase, as the entire system will be available on the Internet. A standard CPU, keyboard, modem, and monitor are used.

There are limitations to this model. The CST is not a diagnostic tool and should not be used for diagnostic purposes. In order to establish a true cognitive impairment, follow-up assessment should be thorough, and diagnoses should use accepted clinical criteria. It is important to note that the major thrust of this study was not to compare the quality of the CST and the MMSE. All cognitive screening tests are affected by variables such as age, education quantity, and education quality. The advantage of the CST is that the inclusion of these variables in a sophisticated way can be brought into the interpretation of a given patient's single test score. This ability is simply not possible with any paper or pencil test, unless the test interpreter is very sophisticated in his or her approach to interpretation. Given these

Figure 6. ROC curve for CLOX1 and CLOX2 scores. .CLOX1 AUC=.683 (SE=.060), p=.005; CLOX2 AUC=.707 (SE=.058), p=.002



advantages of the CST, it was not surprising that the clinical utility of the CST exceeded that of the MMSE when using a single cut score.

REFERENCES

Anthony, J.C., LeResche, L., Niaz, U., Von Korff, M.R., & Folstein, M.F. (1982). Limits of the "mini-mental state" as a screening test for dementia and delirium among hospital patients. *Psychological Medicine*, *12*, 397-408.

Boise, L., Morgan, D.L., Kaye, J., & Camicioli, R. (1999). Delays in the diagnosis of dementia: Perspectives of family caregivers. *American Journal of Alzheimer's Disease and Other Dementias*, *14*, 20-26.

Boise, L., Neal, M.B., & Kaye, J. (2004). Dementia assessment in primary care: Results from a study in three managed care systems. *Journals of Gerontology: Series A: Biological Sciences & Medical Sciences Special Issue: Anti-Aging Medicine: The Hype and the Reality Part 1, 59A*, 621-626.

Crum, R.M., Anthony, J.C., Bassett, S.S., & Folstein, M.F. (1993). Population-based norms for the mini-mental state examination by age and educational level. *Journal of the American Medical Association*, *12*, 2386-2391.

Diagnostic Criteria from DSM IV. (1994). American Psychiatric Association. Washington, DC.

Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189-198.

Ganguli, M., Belle, S., Ratcliff, G., et al. (1993). Sensitivity and specificity for dementia of population-based criteria for cognitive impairment: The MoVIES project. *Journals of Gerontology*, *48*, M152-M161.

Knopman, D., Donohue, J.A., & Gutterman, E.M. (2000). Patterns of care in the early stages of Alzheimer's disease: Impediments to timely diagnosis. *Journal of the American Geriatric Society*, *48*, 300-304.

Lezak, M.D. (1995). *Neuropsychological assessment* (3rd edition). New York: Oxford University Press.

Mohs, R.C., Doody, R.S., Morris, J.C., et al. (2001). A 1 year, placebo-controlled preservation of function survival study of donepezil in AD patients. *Neurology*, *57*, 481-488.

Monsch, A.U., Foldi, F.S., & Ermini-Fuenfschilling, D.E., et al. (1995). Improving the diagnostic accuracy of the mini-mental state examination. *Acta Neurologica Scandinavica*, *92*,145-150.

Olafsdottir, M., Skoog, I., & Marcusson, J. (2000). Detection of dementia in primary care: The Linköping study. *Dementia & Geriatric Cognitive Disorders*, *11*, 223-229.

Petersen, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G., & Kokmen, E. (1999). Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology*, *56*, 303-308.

Reuben, D.B., Roth, C., Kamberg, C., & Wenger, N.S. (2003). Restructuring primary care practices to manage geriatric syndromes: The ACOVE-2 intervention. *The American Geriatrics Society*, *51*, 1787-1793.

Sano, M., & Weber, C. (2003). Psychological evaluation and nonpharmacologic treatment and management of Alzheimer's disease. In P.A. Lichtenberg, D.L. Murman, & A.M. Mellow (Eds.), *Handbook of dementia* (pp. 25-47). Hoboken, NJ: John Wiley & Sons.

Schneider, L. (2000). Treatment of Alzheimer's disease. In I. Katz, & D. Oslin (Eds.), *Annual review of gerontology and geriatrics: Focus on* *psychopharmacologic interventions in late life* (pp. 120-135). New York: Springer Publishing.

Spreen, O., & Strauss, E. (19998). *A compendium of neuropsychological tests.* New York: Oxford University Press.

Tombaugh, T.N., McDowell, I., Krisjansson, B., & Hubbley, A.M. (1996). Mini-mental state examination (MMSE) and the modified MMSE (3MS): A psychometric comparison and normative data. *Psychological Assessment, 8,* 48-59.

Tombaugh, T.N., & McIntyre, N.J. (1992). The mini-mental state examination: A comprehensive review. *Journal of the American Geriatric Society, 4,* 922-935.

Turner, R.S. (2003). Neurologic aspects of Alzheimer's disease. In P.A. Lichtenberg, D.L. Murman, & A.M. Mellow (Eds), *Handbook of dementia* (pp.1-24). Hoboken, NJ: John Wiley & Sons.

Uhlmann, R.F., & Larson, E.B. (1991) Impact of mild to moderate hearing loss on mental status testing. *Journal of the American Geriatric Society, 39*, 876-880.

U.S. Preventative Service Task Force. (2003). Screening for dementia: Recommendations and rationale. In *Guide to clinical preventive services, third edition: Periodic updates; 2003.*

Valcour, V.G., Masaki, K.H., & Blanchette, P.L. (2002). Self reported driving, cognitive status, and physician awareness of cognitive impairment. *Journal of the American Geriatrics Society, 50,* 1265-1267.

Zanetti, O., Vallotti, B., Frisoni, G.B., Geroldi, C., Bianchetti, A., Pasqualetti, P., et al. (1999). Insight in dementia: When does it occur? Evidence for a nonlinear relationship between insight and cognitive status. *Journals of Gerontology: Series B: Psychological Sciences & Social Sciences*, P100-P106.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 1, Issue 3, edited by J. Tan, pp. 47-57, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 3.31
The Impact of Information Technology in Healthcare Privacy

Maria Yin Ling Fung
University of Auckland, New Zealand

John Paynter
University of Auckland, New Zealand

ABSTRACT

The increased use of the Internet and latest information technologies such as wireless computing is revolutionizing the healthcare industry by improving services and reducing costs. The advances in technology help to empower individuals to understand and take charge of their healthcare needs. Patients can participate in healthcare processes, such as diagnosis and treatment, through secure electronic communication services. Patients can search healthcare information over the Internet and interact with physicians. The same advances in technology have also heightened privacy awareness. Privacy concerns include healthcare Web sites that do not practice the privacy policies they preach, computer break-ins, insider and hacker attacks, temporary and careless employees, virus attacks, human errors, system design faults, and social engineering. This chapter looks at medical privacy issues and how they are handled in the U.S. and New Zealand. A sample of 20 New Zealand health Web sites was investigated.

INTRODUCTION

Advances in information technology have increased the efficiency of providing healthcare services to patients. Using Web-based technology, the healthcare team can also include the patient, who must be an informed decision maker and active participant in his or her care. These same advances also improve the features, functions, and capabilities of the electronic medical record systems and potentially increase the number of parties, namely hospitals, insurance companies, marketing agencies, pharmaceutical companies, and employers that may have unauthorized access to private medical information. These systems are justifying themselves in terms of cost and life savings. Accessibility to mobile computing devices in the healthcare industry is also evolving. Wireless computing devices enable physicians, clinicians, and nurses to enter patient data at the point of care (Kimmel & Sensmeier, 2002). Disease management systems provide caregivers with information on efficacy of drugs and treatments at various stages of a medical condition. Using bar-coding technology together with decision support, systems can ensure that patients can receive the correct medication or treatment.

Healthcare organizations must manage a tremendous amount of information, from clinical test results, to financial data, to patient tracking information. While most healthcare organizations have policies and procedures in place to guarantee at least minimum levels of privacy protection, they are not core features of most technology systems in the healthcare industry. This is true despite the fact that unauthorized disclosure of an individual's private medical information can affect one's career, insurance status, and even reputation in the community. Without adequate privacy protection, individuals must take steps to protect themselves from what they consider harmful and intrusive uses of their health information, often at significant costs to their health.

Healthcare privacy is an increasingly complex legal and operational issue facing the healthcare industry. For example, in the areas of mental health, HIV, pharmaceuticals, and genetic information, issues of privacy and the appropriate use of health information have already shown themselves to be particularly sensitive. The public has also become increasingly conscious of privacy issues, such as protection of electronic medical records, commercial uses of health information, and insurer and employer access to patient-identifiable information. The increasing use of the Internet also brings a corresponding need for privacy awareness. The very nature of electronic records makes them more easily transportable and thus accessible.

Healthcare professionals face many challenges as they seek ways to deliver quality healthcare while maximizing efficiency and effectiveness and at the same time ensuring privacy. A substantial barrier to improving the quality of and access to healthcare is the lack of enforceable privacy rules. Individuals share a great deal of sensitive, personal information with their doctors. This information is then shared with others, such as insurance companies, pharmacies, researchers, and employers, for many reasons. Yet unlike other personal information, there is very little legal protection for medical records.

This chapter focuses mainly on the impact that information technology has on healthcare privacy and the ways in which privacy can be achieved. We examine this in the context of the situation in the U.S.A. and in New Zealand, which has supposedly the world's strictest privacy legislations in the Privacy Act (1993). Comparisons to other countries are also made where information security technology has been applied in the medical domain.

WHAT IS HEALTH INFORMATION?

The American Health Information Management Association (AHIMA) (The American Health Information Management Association and the Medical Transcription Industry Alliance, 1998) defines health information as:

- Clinical data captured during the process of diagnosis and treatment.
- Epidemiological databases that aggregate data about a population.
- Demographic data used to identify and communicate with and about an individual.
- Financial data derived from the care process or aggregated for an organization or population.
- Research data gathered as a part of care and used for research or gathered for specific research purposes in clinical trials.
- Clinical data and observations taken by trainees in a teaching hospital.
- Reference data that interacts with the care of the individual or with the healthcare delivery systems, like a formula, protocol, care plan, clinical alerts, or reminders.

- Coded data that is translated into a standard nomenclature or classification so that it may be aggregated, analyzed, and compared.

AHIMA further states that healthcare information and data serve important functions, including:

- Evaluation of the adequacy and appropriateness of patient care.
- Use in making decisions regarding healthcare policies, delivery systems, funding, expansion, education, and research.
- Support for insurance and benefit claims.
- Assistance in protecting the legal interests of the patients, healthcare professionals, and healthcare facilities.
- Identification of disease incidence to control outbreaks and improve public health.
- Provision of case studies and epidemiological data for the education of health professionals.
- Provision of data to expand the body of medical knowledge.

WHAT IS HEALTHCARE PRIVACY?

Healthcare is a service industry that relies on information for every aspect of its delivery. Health information is important to the patients, the medical practitioners, the healthcare professionals, and institutions, in addition to society as it directs the health of the population. It must be protected as a valuable asset, and in its primary form as the medical record of a unique individual, it must be safeguarded.

Privacy of health information is a legitimate concern. Such concerns grow as technology is in place to allow confidential medical records and sensitive health information such as: mental illness, HIV, substance abuse, sexually transmitted disease, and genetic information, to be made available to employers, bankers, insurers, credit card companies, and government agencies for making decisions about hiring, firing, loan approval, and for developing consumer marketing.

The application of information technology to healthcare, especially the development of electronic medical records and the linking of clinical databases, has generated growing concerns regarding the privacy and security of health information. The security and integrity of electronic health data must be protected from unauthorized users. However, in the medical field, accessibility for certain authorized functions must overrule any other concerns, that is, when a doctor needs to access the information about a patient in order to provide emergency treatment, it is imperative that the data become available without delay (Ateniese, Curtmola, de Medeiros, & Davis, 2003).

While patients have a strong interest in preserving the privacy of their personal health information, they may also have an interest in medical research and other efforts by healthcare organizations to improve the quality of medical care they receive.

Categories of Healthcare Privacy

In addition to technological revolutions, which are the main cause for privacy concerns, there are three distinct kinds of violations of health information privacy according to the congressional testimony of Janlori Goldman, director of the Health Privacy Project at Georgetown University (Starr, 1999):

- Individual misappropriation of medical records;
- Institutional practices — ambiguous harm to identifiable individuals; and
- Institutional practices — unambiguous harm to identifiable individuals.

Individual Misappropriation of Medical Records

Starr (1999) states that this category involves individuals who misuse medical data, often publicly disclosing sensitive information and typically violating both the policies of the institutions that kept the records and the laws of their state. It is by far the most common type of violation of health information privacy that can be corrected by stronger penalties and more aggressive enforcement of privacy laws and policies. According to *Health Privacy Project: Medical Privacy Stories* (2003), examples include:

- Following the rape accusations against basketball player Kobe Bryant, the alleged victim's medical records were subpoenaed by Bryant's defense lawyers from a Colorado hospital. After a hospital employee released the records to a judge, attorneys for the hospital have asked that judge to throw out the subpoenas and destroy the records already received by him, citing state and federal medical privacy laws. Attorneys for the victim are also attempting to prevent Bryant's defense team from gaining access to her medical records from two other hospitals. However, a number of news stories have published sensitive medical information that reporters allege came from hospital employees (Miller, 2003).
- A hospital clerk at Jackson Memorial Hospital in Miami, Florida stole the social security numbers of 16 patients named Theresa when they registered at the hospital. The hospital clerk then provided the social security numbers and medical record information to a friend, also named Theresa, who opened up over 200 bank and credit card accounts and bought six new cars (Sherman, 2002).

Institutional Practices: Ambiguous Harm to Identifiable Individuals

This category consists of the use of personal health data for marketing and other purposes where the harm to the individual is ambiguous or relatively small. For example, a chemist or pharmacist sells patient prescription records to a direct mail and pharmaceutical company for tracking customers who do not refill prescriptions, and sending patients letters encouraging them to refill and consider alternative treatments. The problem is not so much harmful to the customers, who might have appreciated the reminders; what worries them most is the hands into which such lists might fall. This may also raise the question of the merchandising of health data for purposes unrelated to those for which patients provided the original information.

Institutional Practices: Unambiguous Harm to Identifiable Individuals

This category consists of institutional practices that do cause harm to identifiable individuals. Different from the other two categories, this one raises much more serious privacy issues and needs correction and reform. Starr stresses that the commingling of the insurance and employment functions in the United States has led to serious abuse of confidential medical information; and the development of genetics has made possible a new and insidious form of discrimination. He recommends security measures such as encryption, the use of a universal health identifier, segmentation of medical records, and biometric identifiers for and audit trails of those accessing medical records (Starr, 1999). Examples from *Health Privacy Project: Medical Privacy Stories* (2003) and Starr (1999) include:

- Two hundred and six respondents in a survey reported discrimination as a result of access to genetic information, culminating in loss of employment and insurance coverage or ineligibility for benefits (Science and Engineering Ethics, 1996).

- A survey found that 35% of Fortune 500 Companies look at peoples' medical records before making hiring and promotion decisions (Unpublished study, University of Illinois at Urbana-Champaign, 1996).
- An Atlanta truck driver lost his job in early 1998 after his employer learned from his insurance company that he had sought treatment for a drinking problem (J. Appleby, "File safe? Health Records May Not Be Confidential," *USA Today*, March 23, 2000, p. A1).

TECHNOLOGICAL CHANGES

Information technologies, such as the Internet and databases, have advanced substantially in the past few years. With the adoption of these new technologies, the healthcare industry is able to save billions of dollars in administrative and overhead costs. These savings can be used to discover new drugs or expand coverage for the uninsured. Through these new technologies, patient care will also be improved; for example, telemedicine allows medical specialists to "examine" and "treat" patients halfway around the world. Perhaps most importantly, information technologies help to empower individuals to understand and to take charge of their own healthcare needs. Patients become active participants in the healthcare process through secure electronic communication services. Wilson, Leitner, and Moussalli (2004) suggest that by putting the patient at the center of the diagnosis and treatment process, communication is more open, and there is more scope for feedback or complaint. This enhances and supports human rights in the delivery of healthcare.

The Internet and Patients

The use of Internet in the healthcare industry involves confidential health information being developed and implemented electronically. There are already several applications available on the Internet for caregivers and patients to communicate and for the electronic storage of patient data. These applications include: electronic mail, online conversations and discussion lists (online chat and NetMeeting), information retrieval, and bulletin boards. Caregivers and patients use electronic mail and online chat to communicate. Patients can search the Web for information about symptoms, remedies, support groups, and health insurance rates. They can also obtain healthcare services, such as second opinions and medical consultations, and products, such as prescription drugs, online (Choy, Hudson, Pritts, & Goldman, 2002).

Patient databases are stored on the Internet, with some providers storing complete patient records in Internet-accessible sites. Patients can interact with databases to retrieve tailored health information (selection-based on personal profile, disease, or a particular need such as travel or cross-border healthcare) (Wilson et al., 2004). However, the availability of medical information in an electronic form (whether or not available over the Internet) raises privacy issues.

The Internet and Health Professionals

Through the use of the Internet, health professionals will have the most up-to-date information available at the click of a mouse. Hospitals, clinics, laboratories, and medical offices will be digitally linked together, allowing for the quick, efficient transfer and exchange of information. The test results will be digitized, allowing for a speedy transfer from labs to hospitals while gaining back the valuable time lost in physical transport. For example, Telehealth in Canada will make geography disappear on a large scale (Siman, 1999). It is a new initiative that significantly improves health services, particularly to remote and rural areas. It also allows physicians to do a complete physical examination of the patient via a digital link. Diagnosis can be made over long-distance telephone lines, rather than after long-distance travel, thus saving the patient the strain and cost of travel. Physicians and other caregivers may use the Internet to discuss unusual cases and obtain advice from others with expertise in treating a particular disease or condition (Siman, 1999).

The Internet and Health-Related Activities

The Internet can support numerous health-related activities beyond the direct provision of care. By supporting financial and administrative transactions, public health surveillance, professional education, and biomedical research, the Internet can streamline the administrative overhead associated with healthcare, improve the health of the nation's population, and lead to new insight into the nature of disease. In each of these domains, specific applications can be envisioned in which the Internet is used to transfer text, graphics, and video files (and even voice); control remote medical or experimental equipment; search for needed information; and support collaboration, in real time, among members of the health community (Committee on Enhancing the Internet for Health Applications: Technical requirements and implementation strategies, 2000). For example, the Internet could do the following (Committee on Enhancing the Internet for Health Applications: Technical requirements and implementation strategies, 2000):

- Enable consumers to access their health records, enter data or information on symptoms, and receive computer-generated suggestions for improving health and reducing risk;
- Allow emergency room physicians to identify an unconscious patient and download the patient's medical record from a hospital across town;
- Enable homebound patients to consult with care providers over real-time video connections from home, using medical devices capable of transmitting information over the Internet;
- Support teams of specialists from across the country who wish to plan particularly challenging surgical procedures by manipulating shared three-dimensional images and simulating different operative approaches;
- Allow a health plan to provide instantaneous approval for a referral to a specialist and to schedule an appointment electronically;
- Enable public health officials to detect potential contamination of the public water supply by analyzing data on nonprescription sales of antidiarrheal remedies in local pharmacies;
- Help medical students and practitioners access, from the examining room, clinical information regarding symptoms they have never before encountered; and
- Permit biomedical researchers at a local university to create three-dimensional images of a biological structure using an electron microscope 1,000 miles away.

Also called: "Medicine of the Millennium," telemedicine is connecting geographically separate healthcare facilities via telecommunications, video, and information systems. The purpose of telemedicine is for remote clinical diagnosis and treatment, remote continuing, medical education, and access to central data repositories for electronic patient records, test requests, and care outcomes.

However, the increasing use of the Internet brings a corresponding need for privacy awareness. The very nature of electronic records makes them more easily transportable and, thus, accessible. Privacy on the Internet is becoming more and more of a concern as confidential information transmitted via the Internet may be intercepted and read by unauthorized persons. Some commonly used Internet protocols may allow information to be altered or deleted without this being evident to either the sender or receiver. Patients may be totally unaware that their personally identifiable health information is being maintained or transmitted via the Internet, and worse still, they may be subject to discrimination, embarrassment, or other harm if unauthorized individuals access this confidential information. While technology can and should be used to enhance privacy, it can also be used to undermine privacy.

WHY ARE THERE HEALTHCARE PRIVACY CONCERNS?

Undoubtedly, the Internet is a valuable tool for improving healthcare because of its ability to reach millions of Internet users at little or no additional cost and absence of geographic and national boundaries. Unfortunately, the Internet is also an ideal tool for the commission of fraud and other online crime. Examples of such fraud include healthcare scams such as the selling of misbranded and adulterated drugs, and bogus miracle cures.

Many of the bigger healthcare Web sites collect information by inviting users to create a personalized Web page where they can acquire medical information tailored specifically to their age, gender, medical history, diet, weight, and other factors. Some sites offer alerts on special medical conditions, health and fitness quizzes, and even the opportunity to store one's own medical records and prescriptions online in case of emergency (Medical privacy malpractice: Think before you reveal your medical history, 2001). Other Web sites collect information using cookies. Cookies are small pieces of data stored by the user's Internet browser on the user's computer hard drive. Cookies will be sent by the user's browser to the Web server when the user revisits the same Web site. Hence the user's information such as number of visits, average time spent, pages viewed, and e-mail address will be collected and used to recognize the user on subsequent visits and provide tailored offerings or content to the user.

The California HealthCare Foundation recently examined the privacy policies and practices of 21 popular health sites including: DrKoop.com, Drugstore.com, and WebMD.com (Medical privacy malpractice: Think before you reveal your medical history, 2001). They found that visitors to the sites are not anonymous, and that many leading health Web sites do not practice the privacy policies they preach. In some cases, third-party ad networks run banner ads on the sites that collect information and build detailed profiles of each individual's health conditions.

In New Zealand, no published survey has been previously conducted. In order to examine the privacy policies and practices of the New Zealand health sites, 20 medical related Web sites were chosen from the electronic yellow pages (www.yellowpages.com.nz) and studied for the purpose of this chapter. At this site, the individual listings are arranged such that those that have Web sites appear first. Only unique sites were examined. That is, those with multiple listings or branches were ignored. Those with only a simple banner ad in the electronic yellow pages were also excluded. Of these 20 Web sites, three were medical insurers, or offered a medical insurance policy as one of their services; one was for health professionals to use to support traveling patients; and the rest were medical clinics and hospitals. The result shows that all but one of these Web sites collected personal information, but only one had a privacy statement, and it was very obscure; three used cookies, and none mentioned the purpose for which the information was collected. The New Zealand Information Privacy Principle 3 requires that a well-expressed Web site should have a privacy statement. A privacy statement tells consumers that their privacy right is being considered (Wiles, 1998). The Web sites studied all failed to meet such a requirement. The results of the above studies indicate that healthcare privacy concerns are not just problems in New Zealand, but universal ones.

According to Anderson (1996), many medical records can be easily obtained by private detectives, who typically telephone a general practice, family health services authority, or hospital and pretend to be the secretary of a doctor giving emergency treatment to the person who is the subject of the investigation

Although privacy is a concern as electronic information is vulnerable to hackers and system errors that can expose patients' most intimate data, the most persistent risk to security and privacy is through the people who have authorized access, much more so than the hackers or inadvertent system errors. As medical information systems are deployed more widely and made more interconnected, security violations are expected to increase in number.

WHAT ARE THE CONCERNS?

The American Health Information Management Association (1998) estimates that when a patient enters a hospital, roughly 150 people have legitimate access to that person's medical record, including food workers, pharmacists, lab technicians, and nursing staff, each with a specific authority to view components of the record necessary for their job and each with unique ability to act within a system.

The increasing use of the Internet in the healthcare industry has also heightened concerns on privacy. The CERT Coordination Center at Carnegie Mellon University, a national resource for collecting information about Internet security problems and disseminating solutions (Committee on Enhancing the Internet for Health Applications: Technical requirements and implementation strategies, 1997), lists seven general areas of vulnerability:

- Compromised system administrator privileges on systems that are unpatched or running old OS versions.
- Compromised user-level accounts that are used to gain further access.
- Packet sniffers and Trojan horse programs.
- Spoofing attacks, in which attackers alter the address from which their messages seem to originate.
- Software piracy.
- Denial of Service.
- Network File System and Network Information System attacks and automated tools to scan for vulnerabilities.

In addition to the above vulnerabilities, other concerns are:

- To whom should organizations be allowed to disclose personal health information with and without patient consent? Under what conditions may such disclosures be made?
- What steps must organizations take to protect personal health information from loss, unauthorized editing, or mischief?
- What types of security technologies and administrative policies will be considered sufficient protection?

Additional Common Threats and Attacks

A threat is any of the capabilities, intentions, and attack methods of adversaries to exploit or cause harm to information or a system. Threats are further defined as being passive (monitoring but no alteration of data) and active (deliberate alteration of information). King, Dalton, and Osmanoglu (2001) define four common threat consequences and the sources of threats in the following sections:

- **Disclosure:** If information or data is revealed to unauthorized persons (breach of confidentiality).
- **Deception:** If corporate information is altered in an unauthorized manner (system or data integrity violation).
- **Disruption:** If corporate resources are rendered unusable or unavailable to authorized users (denial of service).
- **Usurpation:** If the corporate resources are misused by unauthorized persons (violation of authorization).

Temporary or Careless Employees

Electronic health records stored at healthcare organizations are vulnerable to internal or external threats. Although with the protection of firewalls, careless employees, temporary employees, or disgruntled former employees cause far more problems than do hackers. As a company's employees have tremendous access to the company's resources, it is possible that the computer system could be hacked into internally, as well as by third parties. For example, an employee attaches a database of 50,000 names to an e-mail and sends it to a business partner who is working on a marketing campaign at another company. It would be very likely that data could be intercepted or harvested by a third party and used for improper or unauthorized purposes (Silverman, 2002).

Human Errors and Design Faults

A serious threat to the confidentiality of personal health information in hospitals and health authorities is the poor design and lax administration of access controls (Anderson, 1996). In many hospitals, all users may access all records; it is also common for users to share passwords or to leave a terminal permanently logged on for the use of everyone in a ward. This causes a breakdown of clinical and medico-legal accountability and may lead to direct harm. Other design errors include improperly installing and managing equipment or software, accidentally erasing files, updating the wrong file, or neglecting to change a password or backup a hard disk.

Insiders

Another source of threat comes from the trusted personnel (the insiders) who engage in unauthorized activities (copying, stealing, or sabotaging information, and yet their actions may remain undetected) or activities that exceed their authority (abusing their access). The insiders may disable the network operation or otherwise violate safeguards through actions that require no special authorization.

Crackers, Hackers, and Other Intruders

While internal threats consist of authorized system users who abuse their privileges by accessing information for inappropriate reasons or uses, external threats consist of outsiders who are not authorized to use an information system or access its data, but who nevertheless attempt to access or manipulate data or to render the system inoperable. Computer break-ins are proven to have occurred in the healthcare industry. The Health Care Privacy Project, a non-profit corporation in Washington DC, reported that a hacker found a Web page used by the Drexel University College of Medicine in Pennsylvania that linked to a database of 5,500 records of neurosurgical patients (*Health Privacy Project: Medical privacy stories*, 2003). The records included patient addresses, telephone numbers, and detailed information about diseases and treatments. After finding the database through the search engine Google, the hacker was able to access the information by typing in identical usernames and passwords. Drexel

University shut down its database upon learning of the vulnerability, and a university spokeswoman stated that officials had been unaware that the database was available online, as it was not a sanctioned university site.

A "2002 Computer Crime and Security" survey conducted by the Computer Security Institute (CSI) with the participation of the San Francisco Federal Bureau of Investigation's (FBI) Computer Intrusion Squad found that the threat from computer crime and other information security breaches continues unabated and that the threat from within the organization is far greater than the threat from outside the organization. Results show that 74% cited their Internet connection as a frequent point of attack than cited their internal systems as a frequent point of attack (33%); 28% suffered unauthorized access or misuse on their Web sites within the last twelve months; 21% said that they did not know if there had been unauthorized access or misuse; 55% reported denial of service; and 12% reported theft of transaction information (*Cyber crime bleeds U.S. corporations, survey shows; Financial losses from attacks climb for third year in a row*, 2002).

Social Engineering

According to King et al. (2001), a social engineering attack involves impersonating an employee with known authority, either in person (disguised) or by using an electronic means of communication (e-mail, fax, or the telephone). For example, an attacker places a phone call to the system administrator claming to be a corporate executive who has lost the modem pool number and forgotten the password. In the hospitals, an outsider places a phone call to an authorized insider, pretending to be a physician in legitimate need of medical information.

Information Warfare

A RAND Corporation study of information warfare scenarios in 1995 suggests that terrorists using hacker technologies could wreak havoc in computer-based systems underlying emergency telephone services, electric power distribution networks, banking and securities systems, train services, pipeline systems, information broadcast channels, and other parts of our information infrastructure (*Committee on Enhancing the Internet for Health Applications: Technical requirements and implementation strategies*, 2000).

Although the above examples do not specifically describe threats to healthcare organizations, they do indicate the growing vulnerability of information systems connected to public infrastructure such as the Internet. As such, the drive for increased use of electronic health information linked together by modern networking technologies could possibly expose sensitive health information to a variety of threats that will need to be appropriately addressed.

Healthcare Privacy Concerns in the United States

According to Ball, Weaver, and Kiel (2004), a national survey of e-health behavior in the U.S. found that 75% of people are concerned about health Web sites sharing information without their permission and that a significant percentage do not and will not engage in certain health-related activities online because of privacy and security concerns. For example, 40% will not give a doctor online access to their medical records; 25% will not buy or refill prescriptions online; and 16% will not register at Web sites. However, nearly 80% said that a privacy policy enabling them to make choices about whether and how their information is shared would make them more willing to use the Internet for their private health information.

A Pew report (2005) documented that 89% of health seekers were concerned about privacy issues, with fully 71% very concerned. When people were made aware of the possibility of the issuance of universal medical ID numbers, a Gallup poll found that 91% opposed the plan; 96%

opposed the placement on the Web of information about themselves held by their own doctor (The Gallup Organisation, 2000). On the other hand, the healthcare administrators are aware of security issues and have many safeguards in place. In a recent survey of healthcare information technology executives, participants ranked the protection of health data as their primary concern (Reid, 2004). Hospitals, for example, indicate that current security technologies in use include anti-virus software (100%), firewalls (96%), virtual private networks (83%), data encryption (65%), intrusion detection (60%), vulnerability assessment (57%), public key infractions (20%), and biometrics (10%). Virtually all respondents expected to use all these technologies to some degree during the next two years (The Gallup Organisation, 2000).

Recent evidence indicates that many medical organizations are lagging behind in their implementation of appropriate security systems. A study of 167 U.S. hospitals conducted by research firm HCPro found that 76% had not conducted an information security audit, and only half planned to do so by April 2001 (Johnson, 2001). Of the hospitals that had performed an audit, 51% said that they would need major improvements to, or a complete overhaul of, their security systems, and 49% claimed that they would have to significantly change or replace their security policies. Alarmingly, only 5% said they had an annual budget for HIPAA compliance.

The inadequacy of some medical providers' security systems was recently underscored by the hacking of the University of Washington Medical Center (UWMC) computers (*Thomson Corporation and health data management*, 2005). SecurityFocus.com reported that an intruder was able to break into the UWMC computers and view the name, address, and Social Security number and medical procedures of over 4,000 cardiology patients. Theoretically, the UWMC could face potential lawsuits by distressed patients.

Healthcare Privacy Concerns in New Zealand

A survey was conducted in 1998 to study the practice and plans in New Zealand for the collation and retention of health records about identifiable individuals, with particular reference to the implications for privacy arising from the increased use of National Health Index Numbers (NHI) (Stevens, 1998).

What is NHI? The NHI provides a mechanism to uniquely identify healthcare users. It was developed to help protect personally identifying health data held on computer systems and to enable linkage between different information systems whilst still protecting privacy. The NHI database records contain information of each person to whom an NHI number has been allocated, their name, date of birth, date of death, address, gender, ethnicity (up to three entries allowed), residence status, and other names by which they may be known (Stevens, 1998). It, however, does not contain any clinical information, and its availability for research purposes tends to be limited chiefly to a peripheral role in cohort studies and clinical trials (Stevens, 1998).

Alongside the NHI database is the Medical Warnings System (MWS) database, which can only be accessed via the individual's NHI number. The MWS is designed to warn healthcare providers of the presence of any known risk factors that may be important in making clinical decisions about individual patient care.

The MWS database records contain individuals' NHI numbers, donor information (e.g., heart or kidney), contact details for next of kin (name, relationship, and phone number), medical warnings (typically allergies and drug sensitivities, classified as "danger" or "warning" or unverified "report"), medical condition alerts (such as diabetes), and summaries of healthcare events (so far these have been limited to hospital admissions, showing dates of admission and discharge, hospital, and diagnosis or procedure code).

These two databases are maintained by New Zealand Health Care Information Services (NZHIS) formed in 1991, a division of the Ministry of Health. NSHIS is responsible for the collecting, extraction, analysis, and distribution of health information.

In this survey, it is found that a statement on the Ministry Web site states that access to the MWS is "restricted solely for the use of providers in the context of caring for that individual." However, it is estimated that there are some 20,000 people who have direct access to the MWS and a further 70,000 who potentially have access to it, so that in practice, the security of the system probably relies heavily upon the difficulty of getting a hold of the NHI for the individual subject of an unauthorized enquiry (Johnson, 2001).

The survey further reveals that the same Web site document states, in respect of the NHI and MWS systems, "The Privacy Commissioner will be continuously involved in ensuring that the very highest possible standards of integrity and probity are maintained." Yet NZHIS do not appear to have taken any steps either to check with the Privacy Commissioner before making that statement or, having made the statement without the Commissioner's knowledge or agreement, to involve him/her at all in checking arrangements for operation of these databases. At the least, therefore, the statement is misleading in suggesting a form of endorsement by the Privacy Commissioner.

During the survey, more than one doctor contacted admitted that they use a different name for transactions involving their own healthcare, because they do not trust the security of records held by hospitals, laboratories, and other healthcare agencies with which they deal. This implies that the more health records that are to be integrated, the more users that must be concerned about the possibility of security breaches in any one part of the larger system. This also implies that the functions of an information system can be subverted if it does not gain and keep the confidence of both users and subjects.

WHO HAS ACCESS TO THE HEALTHCARE INFORMATION?

There are a variety of organizations and individuals who have an interest in medical data, and they are both within and outside of the healthcare industry. Usually access to the health information requires a patient's agreement by signing a "blanket waiver" or "general consent forms" when the patient obtains medical care. Signing of such a waiver allows healthcare providers to release medical information to employers, insurance companies, medical practitioners, government agencies, court orders or legal proceedings, direct marketers, medical institutions, hospitals, and newsgroups/chat rooms on the Internet.

Employers

Employers have an interest in an employee's fitness to work and fitness to perform particular tasks such as flying airplanes, controlling air traffic, and driving trains, buses, trucks, and cars. Some self-insured businesses establish a fund to cover the insurance claims of employees, which requires employees' medical records to be open for inspection by employers instead of an insurance company.

Insurance Companies

Insurance companies seek to combat rising costs of care by using large amounts of patient data in order to judge the appropriateness of medical procedures. They may also have an interest in healthcare data about a person's injuries and illnesses in relation to medical claims.

In New Zealand, the Accident Rehabilitation and Compensation Insurance Corporation (ACC), whose accident records are used for calculating workplace premium, will be shared with healthcare organizations. For example, to be eligible for weekly compensation, an injured person must be (a) incapacitated through injuries and (b) an earner at the time of the incapacity. ACC obtains medical opinion to clarify incapacity. It also obtains information from Inland Revenue, employers, and accountants to satisfy the second criteria.

Medical Practitioners

The medical practitioners have an explicit statutory obligation to disclose information on patients who have a serious physical condition, notifiable disease, or impairment that the doctor knows is likely to result in significant danger to the public (Clarke, 1990). In some cases it may be important that sensitive health data to be conveyed as part of information provided about a referral, in particular if the patient has been diagnosed as HIV-positive.

Government Agencies

In the U.S., government agencies may request citizens' medical records to verify claims made through Medicare, MediCal, Social Security Disability, and Workers Compensation. In New Zealand, government agencies such as Inland Revenue Department may share the information with healthcare organizations and ACC for tax and benefits purposes.

Medical Institutions and Clinical Researchers

Medical institutions such as hospitals or individual physicians require health information for evaluation of quality of service. This evaluation is required for most hospitals to receive their licenses. Clinical researchers and epidemiologists need health information to answer questions about the effectiveness of specific therapies, patterns of health risks, behavioral risks, environmental hazards, or genetic predisposition for a disease or condition (e.g., birth defects).

Direct Marketers

Drug companies want to know who is taking which drug so that they can conduct post-marketing surveillance to develop marketing strategies. Direct marketers use health-screening tests to collect medical information and build up data banks of businesses for promoting and selling products that are related to the information collected.

Court Orders/Legal Proceedings

In the U.S., medical records may be subpoenaed for court cases for people who are involved in litigation, an administrative hearing, or workers' compensation hearing. In the (less litigatious) New Zealand context, this is more likely to involve the granting of powers of attorney to make decisions on medical matters for patients who are not capable of making such decisions.

Internet Service Providers/Users

The Internet is available for individuals to share information on specific diseases and health conditions. While the Web sites dispense a wide variety of information, there is no guarantee that information disclosed in any of these forums is confidential.

MECHANISMS FOR ADDRESSING HEALTHCARE PRIVACY

Today, healthcare organizations are confronting the challenge of maintaining easy access to medical/clinical data while increasing data security. Technology is only part of the solution. Many healthcare organizations have deployed, to varying degrees, mechanisms to protect the security and integrity of their medical records, such as the use of strong enterprise-wide authentication, encryption, several levels of role-based access control, auditing trails, computer misuse detection systems, protection of external communications, and disaster protection through system architecture as well as through physically different locations. Among other strategies, databases are also used to address security and access control. One database will have consumer identification (ID) geographic information linked to an ID number. The second database will have actual clinical information indexed by patient ID number but no personal data (Ball et al., 2004). However, there are obstacles to the use of security technologies which are yet to be resolved.

Technological Solution

Technological security tools are essential components of modern distributed healthcare information systems. At the highest level, they serve five key functions, as seen in Table 1 (Committee on Enhancing the Internet for Health Applications: Technical requirements and implementation strategies, 1997): availability, accountability, perimeter identification, controlling access, and comprehensibility and control.

However, these types of controls focus more on protecting information *within* healthcare provider institutions and do not address the problems of unrestricted exploitation of information (e.g., for data mining) after it has passed *outside* the provider institution to secondary players or to other stakeholders in the health information services industry (Committee on Enhancing the Internet for Health Applications: Technical requirements and implementation strategies, 1997). In New Zealand, the Health Intranet, a communications infrastructure that allows health information to be exchanged between healthcare providers in a secure way, defines six key elements that any security policy must address (New Zealand Health Information Service, 2001): confidentiality, integrity, authenticity, non-repudiation, auditing and accountability.

Security Architecture

The primary goal of a security architecture design in the healthcare industry is the protection of the healthcare provider's assets: hardware, software, network components, and information resources.

Healthcare Finance Administration (HCFA) (CPRI toolkit: Managing information security in heath care, 2005) suggests that technical protection measures are traditionally grouped into three high level categories:

Table 1. Functions of technological security tools

Principles	Implementation
Availability	Ensuring that information is accurate and up to date when needed
Accountability	Ensuring the access to and use of information is based on a legitimate need and right to know
Perimeter identification	Knowing and controlling the boundaries of trusted access to the information system, both physically and logically
Controlling access	Ensuring the access is only to information essential to the performance of jobs and limiting the access beyond a legitimate need
Comprehensibility and control	Ensuring that record owners, data stewards, and patients understand and have effective control over appropriate aspects of information privacy and access

- **Confidentiality** measures provide the mechanism to ensure that the privacy of information is maintained. Mechanisms include encryption (e.g., virtual private networks, end-to-end, and link level encryption).
- **Integrity** measures enhance the reliability of information by guarding against unauthorized alteration. Protection measures include: digital signature and strong authentication using certificates provided through the Public Key Infrastructure (PKI) initiative.
- **Availability** measures seek to ensure that information assets are accessible to internal and external users when needed and guard against "denial of service" attacks. Protection measures include: firewalls and router filters for mitigating availability risks created by denial of service attacks.

While developing guidelines for the clinical security system for BMA (British Medical Association), Ross Anderson (1996) identified a few shortcomings of the NHS (UK National Health Services) wide network, which are useful for any security architectures to be built for the healthcare industry:

- The absence of an agreed common security policy enforced by all the systems that will connect to the network.
- The lack of confidence in the technical security measures such as firewalls.
- Many of the NHS wide network applications are unethical, which make personal health information available to an ever-growing number of administrators and others outside the control of both patient and clinician. Such availability contravenes the ethical principle that personal health information may be shared only with the patient's informed and voluntary consent. For example, the administrative registers will record patients' use of contraceptive and mental health services, while the NHS clearing system will handle contract claims for inpatient hospital treatment and contain a large amount of identifiable clinical information.
- Item of service and other information sent over existing electronic links between general practitioners and family health services authorities. While registration links are fairly innocuous, at least two suppliers are developing software for authorities that enables claims for items of service, prescrip-

Table 2. Key elements of a security policy

Principles	Implementation
Confidentiality	Ensuring that the message is not readable by unauthorized parties whilst in transit by applying strong encryption tools, such as Digital Certificates
Integrity	Ensuring that the message is not damaged or altered whilst in transit by using secure private networks and Digital Signatures
Authenticity	Ensuring that the user is a trusted party by using user ID/password and/or Digital Certificates
Non-repudiation	Ensuring that the sender cannot claim the message is counterfeit, or deny sending and receiving it by using secure private networks and Digital Signatures
Auditing	Recording user connectivity and site access for audit purposes
Accountability	Identifying clear responsibilities of organizations and individual users through compliance with Legislation and Security Policies

tions, and contract data to be pieced together into a "shadow" patient record that is outside clinical control (Advanced information system, Family Health Services computer unit, 1995; Data Logic product information at http://www.datlog.co.uk/).

Table 3 is a typical security architecture, the components of which are formed based on the ten basic security services (physical security; firewalls; intrusion detection; access control; authentication; privacy and integrity (encryption); electronic signature/non-repudiation; virus protection; audit trail creation and analysis; and database security) identified by HCFA and a list of application-specific baseline requirements for the healthcare industry proposed by King et al. (2001). Some of the components and guidelines are also adopted from the Anderson's UK NHS model.

Encryption

There is an increasing number of health practitioners transferring patient health information

Table 3. Security architecture principles and guidelines

Principles	Security Services	Mechanisms
Confidentiality	Encryption is required over all communications channels (e.g., Internet, ISP-based connections, dial up etc.). Confidential data must be kept encrypted on user laptops and workstations. Such information is to be disclosed only to named individuals on a need-to-know basis.	• Firewalls—Use at connection to Internet and boundary points. • Physical Control—Central office and Data Center continued physical security; integrated smart card access control. • Encryption—Application-specific, primarily DES-based and PKI-based key mgmt; SSL. • Database Security—Proprietary, DBMS-specific; DAC, PKI-enabled; RBAC (role-based access control) integrated; DAC.
Integrity	Business unit managed change control is required. Field-level change history must be maintained. Rollback functionality is required.	Encryption—Application-specific, primarily DES-based and PKI-based key management; SSL (Secure Socket Layer).
Availability	Virus scanning and redundant and high availability solutions are required. Strong system configuration, change control, and regular backup/restore processes are required.	Virus Prevention—Workstation-based and server-based program; signed applications.
Identification and authentication	Strong authentication (encrypted username and password, token, certificate).	Authentication—User ID and password-based with limited smart card pilots; Private key-based with multi-factor identification.

Continued on following page

using electronic mail across wide area networks, for example, using mailbox systems to transfer registration data and item of service claims to family health services authorities, links between general practitioners and hospitals for pathology reports, and the use of Internet electronic mail to communicate with patients that require continuing management. Anderson (1996) suggests that the problem may be tackled using cryptography: encryption and digital signatures can protect personal health information against disclosure and alteration, whether accidental or malicious, while in transit through a network.

Encryption is a tool for preventing the possibility of attack and interception during transmission and storage of data, for assuring confidentiality and integrity of information, and for authenticating the asserted identity of individuals and computer systems by rendering the data meaningless to anyone who does not know the "key." Information that has been cryptographically wrapped cannot be altered without detection. For example, the integrity of a health message is destroyed by removal of personal identifiers or by encryption of crucial pieces of the message. At the destination, the receiver decrypts the message using the same key (symmetrical encryption) or a complementary but different key (asymmetrical encryption) (New Zealand Health Information Service, 2001). Pretty Good Privacy (PGP) and GNU-PGP are commonly used third-party encryption software, which are available free for most common makes of computer.

There are two types of encryption systems: Public-key encryption and private-key encryption. The most commonly used and secure private-key encryption system is the Data Encryption Standard (DES) algorithm developed by IBM in the 1970s, which is gradually replaced by the newer and more efficient algorithm, the Advanced Encryption Standard (AES), which was chosen by the U.S. government after a long, open contest.

Table 3. (continued)

Principles	Security Services	Mechanisms
Authorization and access control	Authorization by business unit or function and detailed role-based access control are required.	Access Control—Platform-specific access control lists; RBAC-based, centrally managed access.
Non-repudiation	Strict change controls are required. Field-level file change history must be maintained. Digital signatures for creator and the checker are required.	Electronic Signature—FIPS 140-1 digital signature; Escrow for encryption keys (not signing keys).
Auditing and monitoring	System-level for user-access, file changes, failed login attempts, alarms.	• Audit Trail Creation & Analysis—Logs generated on a platform-specific basis; Consistent log content, directive data reduction and analysis. • Intrusion Detection—Automated monitoring of limited entry/exit points; Pro-active with integrated action plan.
Compliance with regulations	Compliance with Legislation.	For example, HIPAA in the U.S., European Union Data Protection Directive or the Health Information Privacy Code 1994 in New Zealand.

According to Wayner (2002), the basic design of DES consists of two different and complementary actions: confusion and diffusion. Confusion consists of scrambling up a message or modifying it in some nonlinear way. Diffusion involves taking one part of the message and modifying another part so that each part of the final message depends on many other parts of the message. DES consists of 16 alternating rounds of confusion and diffusion.

Public-key encryption is quite different from the DES. The most popular public-key encryption system is the RSA algorithm, developed in the late 1970s, which uses two keys. If one key encrypts the data, then only the other key can decrypt it. Each person can create a pair of keys and publicize one of the pair, perhaps by listing it in some electronic phone book. The other key is kept secret. If someone wants to send you a message, only the other key can decrypt this message, and only you have a copy of that key. In a very abstract sense, the RSA algorithm works by arranging the set of all possible messages in a long loop in an abstract mathematical space (Wayner, 2002). Public key cryptography is the underlying means for authenticating users, securing information integrity, and protecting privacy. For example, New Zealand North Health is planning to use encryption to encrypt the patients' NHI number and to deposit the information in a database. As such, information about any individual can only be retrieved by means of the encrypted identifier.

In the wide area networks, both secure socket layer (SSL) encryption and IP security (IPSec) should be deployed to allow the continued evaluation of different modes of securing transactions across the Internet. SSL is used to transport the encrypted messages on a communication channel so that no message could be "intercepted" or "faked." It provides authentication through digital certificates and also provides privacy through the use of encryption. (IPSec) protocol, a standards-based method of providing privacy, integrity, and authenticity to information transferred across IP networks, provides IP network-layer encryption.

Virtual Private Network (VPN)

Virtual private networks (VPNs) are standard secure links between companies and their resource users, which allow a company's local networks to be linked together without their traffic being exposed to eavesdropping. It can reduce the exposure to interception of international network traffic. With the increasing use of Internet in the healthcare industry, VPNs play a significant role in securing privacy. VPNs use tunneling and advanced encryption to permit healthcare organizations to establish secure, end-to-end, private network connections over third party networks. Some practical applications that will be used include accessing and updating patient medical records, Tele-consultation for medical and mental health patients, electronic transfer of medical images (x-ray, MRI, mammography, etc.), psychiatric consultations, distance learning, and data vaulting (ScreamingMedia, 1999).

The Hawaii Health Systems Corporation (HHSC) has created a Virtual Private Healthcare Network and Intranet solution that allows for collaboration between its 12 hospitals, 3,200 employees, and 5,000 partners located worldwide. By creating a sophisticated healthcare network that supports high speed, broadband data connectivity, doctors, specialists, and administrators can collaborate throughout the State of Hawaii just as if they were together at the same hospital. This scalable solution also allows existing and future partners, clients, and suppliers to connect to the HHSC network to collaborate and share data. By using a unique subscription profile concept, the network provides impenetrable security and allows for the free and secure flow of mission critical data (ScreamingMedia, 1999).

Firewalls

When private networks carrying confidential data are connected to the Internet, firewalls must be utilized extensively to establish internal security boundaries for protecting the internal private network, computers, data, and other "electronic assets" from tampering or theft by outsiders. Firewalls are a collection of network computing devices such as routers, adaptive hubs, and filters working in tandem and configured to ensure that only expressly permitted packets of data may enter or exit a private network. Firewalls will screen all communication between the internal and external networks according to directives established by the organization. For example, Internet access to an internal patient data system should be entirely prohibited or limited only to those people authenticated by a password or security token (Committee on Enhancing the Internet for Health Applications: Technical requirements and implementation strategies, 2000).

Communications security is also important. Some general practices have branch surgeries, and many hospitals have branch clinics, so the possibility of access via a dial up modem from branches is often raised (Anderson, 1996). In such cases, the main additional risk is that an outside hacker might dial up the main system and gain access by guessing a password. In order to avoid that, Anderson (1996) suggests that there should be no direct dial access to the main computer system. Instead, the main system should dial back the branches. Extra effort should also be made to educate users to choose passwords with care, and all incidents should be investigated diligently.

Audit Trails and Intrusion Detection Monitoring

Transaction logs and audit trails are important, as changes to the patient data can be closely monitored and traced. Audit trails record who and when alterations are made to particular files. The use of audit trails is invaluable, as they can be used as evidence in a court of law. The HCFA information systems create audit logs that record, in a centralized repository, logon and logoff; instances where a role is authorized access or denied access; the individual acting in that role; the sensitivity level of the data or other asset accessed; what type of access was performed or attempted (e.g., whether the nature of the requested action was to create, read, update, execute a program, or delete). Anderson (1996) suggests that periodic audits should be carried out, and from time to time these should include penetration tests. For example, a private detective might be paid to obtain the personal health information of a consenting patient. In this way, any channels that have developed to sell information on the black market may be identified and closed off.

Intrusion detection is primarily a reactive function that responds as attacks are identified. HCFA recommends the use of intrusion detection software to monitor network and host-based assets and employ a computer emergency response team to report and respond when incidents occur.

Biometric Systems

New technology called "biometric authentication" is being used to replace passwords and other security measures with digital recognition of fingerprints or other unique attributes. Biometrics uses individual physiological (finger-scan, iris scan, hand-scan, and retina-scan) or behavioral characteristics (voice and signature scans) to determine or verify identity. The most commonly used is the physiological biometrics. Because biometric security is based on a unique feature of an individual's body, for instance, a fingerprint, it is very difficult to copy, steal, or replicate this information (The Independent Research Group, 2002). Iris-scan is very suitable for use by healthcare institutions. Iris-scan can verify or identify a person based on the unique characteristics of the human iris. The strengths of iris-scan include its high resistance to false matching, the stability of the iris over time, and the ability to use this biometric to provide access to healthcare information or entry into physically secure locations, such as a medical record-keeping or information technology department.

A study done in Albuquerque, New Mexico indicates that the most effective technologies currently available for identification verification (i.e., verifying the claimed identity of an individual who has presented a magnetic stripe card, smart card, or PIN) are systems based on retinal, iris, or hand geometry patterns (Stevens, 1998). On the other hand, single-sign-on technology enables users to log on via user IDs and passwords, biometrics, or other means to gain immediate access to all information systems linked to a network (Clarke, 1990). Single sign-on (SSO) is the capability to authenticate to a given system/application once, and then all participating systems/applications will not require another authentication (King et al., 2001). Both technologies are designed to provide increased security in an unobtrusive manner (Clarke, 1990). St. Vincent Hospital and Health Care Services, Indianapolis had implemented a combined biometric and single-sign-on system in one of its acute care departments using different types of biometric readers to identify physicians and nurses.

Smart Cards

Internet commerce interests are pushing forward aggressively on standards for developing and deploying token-based cryptographic authentication and authorization systems (e.g., the Mastercard-Visa consortium and CyberCash Inc.) (Siman, 1999). Smart Card Token is a smart card about the size of a credit card and has a liquid crystal display on which a number appears that changes every minute or so. Each user card generates a unique sequence of numbers over time and, through a shared secret algorithm for which the user has been assigned access privileges, can generate the corresponding sequence of numbers. The number can be used as a session password. The write-controlled internal memory supports services such as user-specific information storage, authentication, and cryptographic certificate management. Some even have biometric access control features. Employees and appropriate contractors will be issued smart cards or tokens that store a private key and other essential authentication information.

Access Control

A serious threat to the confidentiality of personal health information in hospitals and health authorities is the poor design and lax administration of access controls (Anderson, 1996). Anderson stresses that, in particular, the introduction of networking may turn local vulnerabilities into global ones if the systems with ineffective access controls are connected together in a network, and then instead of the data being available merely to all staff in the hospital, they might become available to everyone on the network.

However, access controls must also be harmonized among networked systems, or moving information from one system to another could result in leaks. The solution for this is to have a common security policy that clearly states who may access what records and under what circumstances. Anderson emphasizes that the following are important to the implementation of effective access controls:

- A senior person such as a hospital manager or partner in general practice must be responsible for security, especially if routine administration is delegated to junior staff. Many security failures result from delegating responsibility to people without the authority to insist on good practice.
- The mechanisms for identifying and authenticating users should be managed carefully. For example, users should be educated to

pick passwords that are hard to guess and to change them regularly; and terminals should be logged off automatically after being unused for five minutes.

- Systems should be configured intelligently. Dangerous defaults such as maintenance passwords and anonymous file transfer access supplied by the manufacturer should be removed. User access should be restricted to departments or care teams as appropriate. With hospital systems that hold records on many people, only a few staff should have access to the files of patients not currently receiving treatment.

Password Management

In many hospitals all users may access all records and often share passwords and leave terminals permanently logged on for the use of everyone in a ward. Such behavior causes a breakdown of clinical and medicolegal accountability and may lead to direct harm: one case has been reported in which a psychiatric patient changed prescription information at a terminal that was left logged on (Anderson, 1996).

It is important for administrators to educate all users that passwords issued to an individual should be kept confidential and not be shared with anyone. When a user ID is issued to a temporary user who needs access to a system, it must be deleted from the system when the user has finished his or her work. All passwords should be distinctly different from the user ID, and ideally they should be alphanumeric and at least six characters in length. Also, passwords should be changed regularly, at least every 30 days. Rittinghouse and Ransome (2004) suggest that it is a good security practice for administrators to make a list of frequently used forbidden passwords. Standard passwords that are often used to get access to different systems for maintenance purposes are not recommended.

Database Security

Database authentication and access control will be public key enabled and role-based. This means that a user will employ a multi-factor authentication procedure based on knowledge of his/her private key to obtain access to a database. Once authentication is complete, access, sometimes down to the record level, will be granted or denied based on the user's roles and associated privileges. Database security will be implemented on a discretionary access control (DAC) basis.

Social Engineering and Careless Disclosure Safeguards

The weakest link in security will always be people, and the easiest way to break into a system is to engineer your way into it through the human interface (CPRI toolkit: Managing information security in heath care, 2005). The main threat to the confidentiality of clinical records is carelessness in handling telephone/e-mail/fax inquiries, instant messaging and on-site visits, and inadequate disposal of information.

According to King et al. (2001), social engineering safeguards consist of non-technical (procedural) means that include: security training for all corporate users; security awareness training for all system administration personnel with well-documented procedures, handling, and reporting; and security awareness training for personnel responsible for allowing outside visitors into restricted areas (such as assigned escorts).

With regard to careless disclosure, Anderson (1996) developed a set of common sense rules that the best practices have used for years and that are agreed by the UK NHS Executives. Whether records are computerized or not, these rules of best practice can be summed up as clinician-consent-call back-care-commit:

- Only a clinician should release personal health information. It should not be released by a receptionist or secretary.

- The patient's consent must be obtained, except when the patient is known to be receiving treatment from the caller or in the case of emergency or the statutory exemptions. In the latter two cases the patient must be notified as soon as reasonably possible afterward.
- The clinician must call back if the caller is not known personally, and the number must be verified, for example, in the Medical Directory. This procedure must be followed even when an emergency is claimed, as private investigators routinely claim emergencies.
- Care must be taken, especially when the information is or may be highly sensitive, such as HIV status, details of contraception, psychiatric history, or any information about celebrities.
- The clinician must commit a record of the disclosure to a ledger. This should have the patient's name; whether consent was sought at the time (and, if not, the date and means of notification); the number called back and how it was verified; and whether anything highly sensitive was disclosed.

In addition, the guidelines for disclosure by telephone should also apply to faxes. Verifying the identity or, failing that, the location of the caller is just as important as it is when disclosing personal health information over the telephone. It is important, and it is the BMA's established advice that personal health information should be faxed only to a machine that is known to be secure during working hours.

Equipment Theft, Loss, and Damage

Anderson (1996) considers the most serious threat to the continued availability of computerized clinical information in general practice to be theft of the computer that has been experienced by over 10% of general practices surveyed. Data can also be destroyed in other ways such as by fire, flood, equipment failure, and computer viruses. He suggests that physical security measures must be taken; hygiene rules to control the risk of computer virus infestation must be applied together with a tested recovery plan.

Since most organizations do not perform realistic tests of their procedures, with the result that when real disasters strike recovery is usually held up for lack of manuals and suppliers' phone numbers, it is important that a drill based on a realistic scenario, such as the complete destruction of a surgery or hospital computer room by fire must be carried out, and a full system recovery to another machine from back up media held off site must be performed. Another measure is to keep several generations of back ups in cases of equipment failure and virus attacks that it may take time to notice that something has gone wrong. A typical schedule in a well run establishment might involve back ups aged one, two, three, four, eight, and twelve weeks, as well as daily incremental back ups.

Limitations of Security Technologies

Despite an aggressive move toward computerized healthcare records in recent years and ongoing parallel technological improvements, there are still limitations of the security technologies to achieve usable and secure systems (Gillespie, 2001).

Firewalls

Firewalls do not offer perfect protection, as they may be vulnerable to so-called tunneling attacks, in which packets for a forbidden protocol are encapsulated inside packets for an authorized protocol, or to attacks involving internal collusion (Gillespie, 2001). One of the concerns with firewalls is that most firewalls pass traffic that appears to be Web pages and requests more and more, as it is the way to get things to work through the firewall. The solution is to re-implement the whole as Web services (Webmail being a good example). These pressures continually erode the effectiveness of firewalls (Ateniese et al., 2003). For example, the NHS Network in Britain is a private intranet intended for all health service users (family doctors, hospitals, and clinics — a total of 11,000 organizations employing about a million staff in total). Initially, this had a single firewall to the outside world. The designers thought this would be enough, as they expected most traffic to be local (as most of the previous data flows in the health service had been). What they did not anticipate was that as the Internet took off in the mid-1990s, 40% of traffic at every level became international. Doctors and nurses found it very convenient to consult medical reference sites, most of which were in America. Trying to squeeze all this traffic through a single orifice was unrealistic. Also, since almost all attacks on healthcare systems come from people who are already inside the system, it was unclear what this central firewall was ever likely to achieve (Ateniese et al., 2003).

Cryptography

The basis for many of the features desired for security in healthcare information systems depends on deploying cryptographic technologies. However, there are limitations to the use of cryptography. One problem is that security tools based on cryptography are still largely undeployed. One general weakness is poor usage of the system by individuals that includes: easily guessed passwords to the cryptographic system are chosen, or even written down on a sticker and stuck on the notebook, or people use the same password across different systems. The password then becomes as safe as the weakest system that is using it (which will often be something like a Web browser that has been told to remember the password) (Anderson, 2005; Gutmann, 2005). The other problem is that cryptography does not solve the security problem, that is, cryptography transforms the access problem into a key management problem, including authentication, digital signatures, information integrity management, session key exchange, rights management, and so on. It is observed that as the scope of key management services grows, trust in the integrity of key assignments tends to diminish, and the problems of revocation in the case of key compromise become much more difficult (Gillespie, 2001). Although public key infrastructure can help deal with the problem, it has also introduced complexities of its own. This has led to organizations effectively misusing cryptographic keys, as managing them appropriately has become too complex. The simplest example is that everyone in the organization really does get the same key.

Biometrics

The deployment of biometrics is proven to be advantageous to the healthcare providers because it provides added security, convenience, reduction in fraud, and increased accountability. It increases the level of security by providing access to health information to authorized individuals and locking out those with nefarious intent. However, there are drawbacks to the technology. For example, when performing an iris-scan, individuals must remain perfectly still during enrollment and presentation, or the system will not be able to scan the iris, therefore causing false non-matching and failure-to-enroll anomalies to occur. Reid (2004) further identified a few drawbacks to biometrics: hardware costs, user perception, placement, and size. For example, iris-scans require specialized cameras with their own unique light source that can be very expensive. The user perception on having infrared light shined into the eye is quite disconcerting. To get the iris in the proper position can be quite time consuming. Some cameras can use eye recognition techniques to try to auto-pan and focus the camera, but such solutions do increase the cost of the camera and may still require some user coordination. The current size of the camera, which has been reduced to that of a desktop camera on steroids, is still very large. It needs further reduction to be able to work efficiently on a desk.

Hardware and Software Costs

The costs of putting secure technologies in place can be tremendous. Very often the implementation of secured systems requires procurement of new software and hardware as the legacy system becomes obsolete. Unfortunately, there are not many commercial tools readily available in the market to integrate legacy systems into modern distributed computing environments. Furthermore, such integration will involve many database content inconsistencies that need to be overcome, including patient identifier systems, metadata standards, information types, and units of measurement.

Overall the lack of standards for security controls and for vendor products that interoperate between disparate systems will hinder the implementation and enforcement of effective security solutions.

Legislation

The importance of assuring the privacy and confidentiality of health information has always been acknowledged. However, up until recently the legal protection for personal information has been patchy and disorganized (ScreamingMedia, 1999).

U.S. Legislation

The healthcare industry is currently going through an overhaul to meet government-mandated regulations stemming from HIPAA to ensure patient confidentiality, privacy, and efficiency. HIPAA, which was passed in 1996 and effective in 2001, gives consumers the right to their medical records, to limit disclosure, and to add or amend their records. Providers must have complied by April 2003. Entities covered include health insurers, physicians, hospitals, pharmacists, and alternative practitioners such as acupuncturists.

HIPAA requires all healthcare providers, health insurers, and claims clearinghouses to develop and implement administrative, technical, and physical safeguards to ensure the security, integrity, and availability of individually identifiable electronic health data. Failure to comply with HIPAA can result in civil fines of up to $25,000 a year for each violation of a standard. Because HIPAA encompasses dozens of standards, the fines can add up quickly, and wrongful disclosure of health information carries a criminal fine of up to $250,000, 10 years imprisonment, or both (King et al., 2001).

N.Z. Legislation

In New Zealand, the Privacy Act, which came into force on July 1, 1993, provides a measure of legal protection for all personal information, including health information, and applies to the public and private sectors and to information held in both paper and electronic formats.

The Health Information Privacy Code 1994, which is consistent with the provisions of the Privacy Act 1993 (s.46), was issued by the Privacy Commissioner specifically to protect the privacy of personal health information. While the code protects personal health information relating to an identifiable individual, it does not apply to statistical or anonymous information that does not enable the identification of an individual.

The Medicines Act 1981 was issued by the Ministry of Health to penalize any unauthorized sale of prescription medicines, publication of advertisements containing insufficient information about precautions and side effects, and advertising the availability of new medicines before their approval for use in New Zealand. Under Section 20, the maximum penalty for an individual is up to six months imprisonment or a fine not exceeding $20,000. Sections 57 and 18 have a maximum penalty for an individual of three months imprisonment and a fine not exceeding $500.

The Privacy Commission also considered the application of the Privacy Act to the process of caching (Anderson, 2001). Caching occurs when a Web page accessed by users is temporarily stored by the user's computer (client caching) or by the network server that provides the user with Internet access (proxy caching). It also considered that the Privacy Act applied to the use of cookies within New Zealand and offered sufficient protection. It is proposed that using cookies for the purpose of collecting, holding, or giving access to personal information would be an offence unless the Web site indicated such information would be gathered. However, the Privacy Commissioner did not support the creation of such an offense.

European Union Data Protection Directive

International action may further affect the ways in which personal health information is transmitted over the Internet. The EU Data Protection Directive, which went into effect on October 25, 1998, requires EU member states to block outbound transmissions of data to countries that do not have laws providing a level of privacy protection similar to that in the country where the data originated (Siman, 1999). The directive affords the people to whom the data refer a host of rights, including the right to be notified of data collection practices, to access information collected about them, and to correct inaccuracies (Stevens, 1998). In 1998, New Zealand addressed three aspects of the Privacy Act to ensure it is adequate for the purposes of the EU directive. This is important for New Zealand businesses dealing in personal data originating from Europe because the directive limits the exportation of data to third countries (countries outside the EU) that do not have an adequate privacy protection (Wiles, 1998). The three aspects are:

- The channeling of data from Europe through New Zealand to unprotected data havens;
- Limits on who may exercise rights of access and correction under the Privacy Act; and
- The complaints process. "With a long queue of complaints awaiting investigation, the EU may have concerns that our complaints system is not sufficiently resourced to provide timely resolution of complaints."

In view of the above, the Privacy Commissioner addressed that the Privacy Act is built upon a desire that the collection, holding, use, and disclosure of personal information should be carefully considered and that all activities in this area should be as open as possible.

FUTURE TRENDS

The growth of wireless computing in healthcare will take place for two reasons (The Independent Research Group, 2002):

- For all electronic medical record systems to work, physicians cannot be tied down to wired PC workstations. They will need to use some type of wireless device that allows them access to the relevant hospital databases.
- As the cost of healthcare continues to rise, many individuals are being treated on an outpatient basis. To keep track of an outpatient's vital statistics or signal when the patient needs immediate medical attention, many pervasive devices, such as toilet seats, scales, smart shirts, smart socks, and pacemakers, are being developed that collect relevant patient information. Collected data can then be transmitted via a wireless device using a wireless or mobile network to the patient's physician, who can then decide on possible interventions.

In the recent years, the technological advancements in sophisticated applications and interoperability has increased the popularity of wireless LAN (WLAN) and the use of wireless technology in healthcare. Also with the faster connection speeds of broadband LANs, the healthcare providers have developed a number of applications to improve patient safety and the healthcare delivery process. According to Kourey (2005), the use of personal digital assistants (PDAs) has become increasingly popular. It is because, "PDAs provide access to data and e-mail, store and retrieve personal and professional information and facilitate communication in wireless environments, their use among healthcare professionals has skyrocketed. Industry experts predict the trend to continue. In 2001, 26% of American physicians used handheld devices for tasks related to patient care. While some experts predict this number to reach 50% by 2005." Increasingly, clinicians can check on patient data or order treatments through secure wireless networks from anywhere in the hospital. For example (Hermann & Norine, 2004):

- A nurse is automatically notified on a handheld wireless device that a patient's blood pressure is falling.
- A doctor on rounds receives the results of an important blood test on a wireless PDA instead of having to call the lab for the information.
- A telemetry system records the vitals signs of dozens of patients in critical care and sends them wirelessly to a central control station for continuous, around-the-clock monitoring.
- A surgeon completing a procedure writes after-care orders while still in the operating room and transmits them to the clinical information system, making them instantly part of the patient's electronic record.

Rittinghouse and Ransome (2004) stress that employees who have not been properly educated about wireless security may not realize the dangers a wireless network can pose to an organization, given wireless computing is still a very new technology. They classify WLAN security attacks into two types:

- **Passive attacks** — An unauthorized party simply gains access to an asset and does not modify its content (i.e., eavesdropping). While an attacker is eavesdropping, he or she simply monitors network transmissions, evaluating packets for specific message content. For example, a person is listening to the transmissions between two workstations broadcast on an LAN or that he or she is running into transmissions that take place between a wireless handset and a base station.
- **Active attacks** — An unauthorized party makes deliberate modifications to messages, data streams, or files. It is possible to detect this type of attack, but it is often not preventable. Active attacks usually take one of four forms (or some combination of such):
    1. *Masquerading*: The attacker will successfully impersonate an authorized network user and gain that user's level of privileges.
    2. *Replay*: The attacker monitors transmissions (passive attack) and retransmits messages as if they were sent by a legitimate messages user.
    3. *Message modification*: It occurs when an attacker alters legitimate messages by deleting, adding, changing, or reordering the content of the message.
    4. *Denial-of-service (DoS)*: It is a condition that occurs when the attacker prevents normal use of a network.

When patient information is sent wirelessly, additional security measures are advisable, although a well-defined wireless utility basically protects confidentiality and restricts where the signal travels (Hermann & Norine, 2004). Tabar (2000) suggests that the growth of new technology also creates a unique security threat and requires user authentication protocols. For example, PDAs, laptops, and even mobile carts can fall into unauthorized hands; the electronic ID must be stored elsewhere. Vendors are working on solutions such as: hardware ID tokens that are inserted into the mobile devices before use and radio transmitter-tracking devices. Other browser-based only applications on the mobile computing device are also used such that the patient data resides only on the server and cannot be accessed by the mobile computing device once it is outside the WLAN coverage area. Turisco and Case (2001) argue that while vendors are responsible for code sets, encryption, privacy, and audit trails, user organizations need to manage the device with extreme care or cautions. Physical security is of paramount concern in the wireless communications. The device needs to be turned off when not in use and be kept in a safe place. Tabar (2000) concurs that the greatest hurdle in information security still rests with the user, and no technology can make up for slack policies and procedures. "Changing perceptions, culture and behavior will be the biggest challenges," says Monica Summers, IS Director at Beaufort Memorial Hospital, Beaufort, S.C. "It's not just the technology. You could slap down $5 million in technology, and it won't stop people from giving out their password." (Tabar, 2000).

CONCLUSION

Privacy is not just about security measures, but is at least as much about what information is collected and collated and practically recoverable. Health information has always been regarded as highly sensitive, which must be protected by medical ethics and privacy legislations. The emergence of new technology and new organizational structures in the healthcare industry has opened up the means and the desire to collect and collate such information in ways never previously considered.

The increased use of the Internet and latest information technologies such as wireless computing are revolutionizing the healthcare industry by improving healthcare services, and perhaps most importantly, empowering individuals to understand and take charge of their own healthcare needs. Patients become involved and actively participate in the healthcare processes, such as diagnosis and treatment through secure electronic communication services. Patients can search healthcare information over the Internet and interact with physicians. This enhances and supports human rights in the delivery of healthcare. The same technologies have also heightened privacy awareness. Privacy concerns include: healthcare Web sites that do not practice the privacy policies they preach, computer break-ins, insider and hacker attacks, temporary and careless employees, virus attacks, human errors, system design faults, and social engineering. Other concerns are the collection, collation, and disclosure of health information. Healthcare providers and professionals must take into account the confidentiality and security of the information they collect and retain. They must also ensure that their privacy policies or secure technologies meet the public expectation and abide by the law. Such policies and technologies must also be implemented to ensure the confidentiality, availability, and integrity of the medical records. If this is not done, resources could be wasted in developing secure systems, which never reach fruition, and the new systems will never gain the confidence of the public or of the health professionals who are expected to use them.

Technology is, to a large extent, both the cause of and the solution to concerns about the protection of personal health information. However, there are limitations to the secure technologies that need on-going research and development. Technologies, if coupled with physical security control, employee education, and disaster recovery plans, will be more effective in securing healthcare privacy. Further advances of new information technologies, if designed and monitored carefully, will continue to benefit the healthcare industry. Yet patients must be assured that the use of such technologies does not come at the expense of their privacy.

REFERENCES

Anderson, R. (1996, January 2). *Clinical system security: Interim guidelines*. Retrieved June 2005, from http://www.ftp.cl.cam.ac.uk/ftp/users/rja14/guidelines.txt

Anderson, R. (2001). *Security engineering: A guide to building dependable distributed systems*. Wiley. Retrieved June 2005, from http://www.ftp.cl.cam.ac.uk/ftp/users/rja14/c18_anderson.pdf

Andreson, R. (n.d.). *Why cryptosystems fail*. Retrieved June 2005, from http://www.cl.cam.ac.uk/users/rja14/

Ateniese, G., Curtmola, R., de Medeiros, B., & Davis, D. (2003, February 21). *Medical information privacy assurance: Cryptographic and system aspects*. Retrieved June 2005, from http://www.cs.jhu.edu/~ateniese/papers/scn.02.pdf

Ball, M., Weaver, C., & Kiel, J. (2004). *Healthcare information management systems: Cases, strategies, and solutions* (3rd ed.). New York: Springer-Verlag.

Carter, M. (2000). Integrated electronic health records and patient privacy: Possible benefits but real dangers. *MJA 2000*, *172*, 28-30. Retrieved 2001, from http://www.mja.com.au/public/issues/172_01_030100/carter/carter.html

Choy, A., Hudson, Z., Pritts, J., & Goldman, J., (2002, April). *Exposed online: Why the* new *federal health privacy regulation doesn't offer much protection to Internet users* (Report of the Pew Internet & American Life Project). Retrieved June 2005, from http://www.pesinternetorg/pdfs/PIP_HPP_HealthPriv_report.pdf

Clarke, R. (1990). Paper presented to the Australian Medical Informatics Association, Pert. Australian National University. Retrieved 2001, from http://www.anu.edu.au/people/Rogger.Clarke/DV/PaperMedical.html

Committee on Enhancing the Internet for Health Applications: Technical requirements and implementation strategies. (1997). For the record: Protecting electronic health information. National Academy Press: Washington. Retrieved 2001, from http://bob.nap.edu/html/for/contents.html

Committee on Enhancing the Internet for Health Applications: Technical requirements and implementation strategies. (2000). Networking health: Prescriptions for the Internet. National Academy Press: Washington. Retrieved 2001, from htttp://www.nap.edu/books/0309068436/html

Constantinides, H., & Swenson, J. (2000). *Credibility and medical Web sites: A literature review*. Retrieved 2001, from http://www.isc.umn.edu/research/papers/medcred.pdf

CPRI toolkit: Managing information security in heath care. (n.d.). Target it architecture Vol. 6. Security architecture version 1. The Centers for Medicare & Medicaid Services (CMS) Information Security (previously Health Care Financing Administration). Retrieved June 2005, from http://www.cms.hhs.gov/it/security/docs/ITAv6.pdf

Cyber crime bleeds U.S. corporations, survey shows; financial losses from attacks climb for third year in a row. (2002, April 7). Retrieved June 2005, from http://www.gocsi.com/press/20020407.jhtml?_requestid=953064

DeadMan's handle and cryptography. (n.d.). Retrieved June 2005, from http://www.deadmanshandle.com/papers/DMHAndCryptology.pdf

E-commerce, privacy laws must mesh. (2001, February/April). *News from the Office of The Privacy*

Commissioner, (39). Retrieved 2001, from http://www.privacy.org.nz/privword/pwtop.html

Fox, S. (n.d.). *Vital decisions.* Pew Internet & American Life Project. Retrieved June 2005, from http://www.pewinternet.org

Fox S., & Wilson R. (n.d.). *HIPAA regulations: Final. HIPAA Regs 2003.* Retrieved June 2005, from http://www.hipaadvisory.com/regs/

Gillespie, G. (2001). *CIOs strive to increase security while decreasing the 'obstacles' between users and data.* Retrieved 2001, from http://www.health datamanagement.com/html/current/CurrentIssueStory.cfm?PostID=9059

Gutmann, P. *Lessons learned in implementing and deploying crypto software.* Retrieved June 2005, from http://www.cs.auckland.ac.nz/~pgut001/

Health Insurance Portablity and Accountability Act of 1996. Public Law No. 104-191. Section 1173, USC 101. (1996).

Health privacy (About 1.4 million computer records for in-home supportive service breached). (2004, October 21). *California Healthline.* Retrieved June 2005, from http://www.californiahealthline.org/index.cfm?Action=dspItem&itemID=10 6520&ClassCD=CL141

Health Privacy Project: Medical privacy stories. (2003). Retrieved June 2005, from http:/www.healthprivacy.org/usr_doc/Storiesupd.pdf

Hermann, J., & Norine P. (2004). *Harnessing the power of wireless technology in healthcare.* Retrieved April 2005, from http://www.johnson-controls.com/cg-

iHealthbeat. (2001). Business and Finance: Survey. Retrieved June 2005, from http://www.ihealthbeat.org/members/basecontentwireless/pdfs/healthcare.pdf

Johnson, A. (2001). *The Camelot Avalon: Healthcare procrastinates on HIPAA.* Retrieved 2001, from http://www.camelot.com/newsletter.asp?PageID =326&SpageId=504

Kimmel, K., & Sensmeier, J. (2002). *A technological approach to enhancing patient safety* (White Paper). The Healthcare Information and Management Systems Society (HIMSS) (sponsored by Eclipsys Corporation). Retrieved June 2005, from http://www.eclipsys.com/About/IndustryIssues/HIMSS-TechApproach-to-PatientSafety-6-02-FORMATTED.pdf

King, C., Dalton, C., & Osmanoglu, T. (2001). *Security architecture: Design, deployment & operations.* USA: Osborne/McGraw Hill.

Kourey, T. (n.d.). *Handheld in healthcare part two: Understanding challenges, solutions and future trends.* Retrieved March 2005, from http://www.dell4healthcare.com/offers/article_229.pdf

Medical Privacy Malpractice: Think before you reveal your medical history. (n.d.). Retrieved 2001, from http://www.perfectlyprivate.com/beware_medical.asp

Miller, M. (2003, September 8). Issues of privacy in Bryant case. *Los Angeles Times.*

Ministry of Health Press Release. (2001). *Conviction for Internet drugs warning to other.* Retrieved 2001, from http://www.http://www.moh.govt.nz/moh.nsf/aa6c02e6249e77359cc256e7f0005521d/5332afe9b6839587cc256ae800647e1b? OpenDocument

More prosecutions tipped for online drugs. (2001, October 19). Retrieved 2001, from http://www.stuff.co.nz/inl/index/0,1008,978172a1896,FF.html

New Zealand Health Information Service. (2001). *Health intranet: Health information standards.* Retrieved 2001, from http://www.nzhis.govt.nz/intranet/standards.html

Null, C. (2003, March 4). Google: Net hacker tool du jour. *Wired News.*

Paynter, J., & Chung, W. (2002, January). Privacy issues on the Internet. In *Proceedings of the Thirty-Fifth Hawaii International Conference on System Sciences (HICSS-35),* Hawaii, USA.

Reid, P. (2004). *Biometrics for network security.* New Jersey: Prentice Hall.

Rogers, R. D. (1998). *Information protection in healthcare: Knowledge at what price? We're drowning in information and starving for knowledge*. Retrieved 2001, from http://www.privacy.org.nz/media/aichelth.html

Rossman, R. (2001). *The Camelot Avalon: Despite HIPAA uncertainty, security must prevail.* Retrieved 2001, from http://www.camelot.com/newsletter.asp?Page ID=416&SpageID=504

Rottinghouse, J., & Ransome J. (2004). *Wireless operational security.* USA: Elsevier Digital Press.

ScreamingMedia, Business Wire. (1999). *Seven-Mountains Software and SCI Healthcare Group to deliver a secure virtual private networking solution to Hawaii Health Systems Corporation*. Retrieved 2001, from http://www.industry.java.sun.com/javanews/stories/story2/0,1072,18810,00.html

Sherman, D. (2002) *Stealing from the sick*. Retrieved May 21, 2002, from http://www.NBC6.net

Silverman, M. (2002). *Inside the minds: Privacy matters.* Retrieved June 2005, from http://www.duanemorris.com/articles/static/SilvermanBookExcerpt.pdf

Siman, A. J. (1999). *The Canada health infoway — Entering a new era in healthcare*. Retrieved 2001, from http://www.hc-sc.gc.ca/ohih-bsi/available/documents/ecompriv_e.html

Stevens, R. (1998). *Medical record databases. Just what you need?* Retrieved 2001, from http://www.privacy.org.nz/people/mrdrep.html

Starr, P. (1999). *Privacy & access to information: Striking the right balance in healthcare* (Health and the Right to Privacy, Justice Louis Brandeis Lecture). Massachusetts Health Data Consortium, Boston, USA. Retrieved 2001 and June 2005, from http://www.nchica.org/HIPAAResources/Samples1/privacylessons/P-101%20Massachusetts%20Health%20Data%20Consortium.htm

Tabar, P. (2000). Data security: Healthcare faces a tricky conundrum of confidentiality, data integrity and timeliness. *Healthcare Informatics.* Retrieved April 2005, from http://www.healthcare-informatics.com/issues/2000/02_00/cover.htm

The American Health Information Management Association and the Medical Transcription Industry Alliance. (1998). *AHIMA position statement: Privacy official.* Retrieved 2001, from http://www.ahima.org/infocenter/index.html

The Gallup Organisation. (2000). *Public attitudes towards medical privacy.* The Institute for Health Freedom. Retrieved June 2005, from http://www.forhealthfreedom.org/Gallupsurvey/

The Independent Research Group. (2002). SAFLINK Report. Retrieved 2001, from http://www.cohenresearch.com/reports/sflk_11-05-02.pdf

Thomson Corporation and health data management. Survey: Hospitals boosting data security. (n.d.). Retrieved June 2005, from http://www.healthdatamanagement.com/html/

Turisco, F., & Case J. (2001). *Wireless and mobile computing.* Retrieved April 2005, from http://www.chcf.org/documents/ihealth/WirelessAndMobileComputing.pdf

Wayner, P. (2002). *Disappearing cryptography. Information hiding: Steganography & watermarking* (2nd ed.). USA: Elsevier Science.

Wiles, A. (1998). *Integrated care and capitation: New challenges for information protection*.

Retrieved 2001, from http://www.privacy.org.nz/shealthf.html

Wilson, P., Leitner, C., & Moussalli, A. (2004). *Mapping the potential of eHealth: Empowering the citizen through eHealth tools and services.* Retrieved June 2005, from http://www.cybertherapy.info/pages/e_health_2004.pdf

This work was previously published in Privacy Protection for E-Services, edited by G. Yee, pp. 56-93, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Section IV
Utilization and Application

This section introduces and discusses the ways in which information technology has been used to shape the medical and biomedical sectors and proposes new ways in which IT-related innovations can be implemented within organizations and in society as a whole. These particular selections highlight, among other topics, the implementation of telehealth and the evolution of new, computerized systems in hospitals. Contributions included in this section provide excellent coverage of today's medical environment and insight into how medical informatics impacts the fabric of our present-day global village.

Chapter 4.1
Successful Health Information System Implementation

Kristiina Häyrinen
University of Kuopio, Finland

Kaija Saranto
University of Kuopio, Finland

INTRODUCTION

A Standish Group (1994) study showed that only 16% of all information technology projects come in on time and within budget. The situation is not better concerning health information systems. Many health information system implementations are less than completely successful (Berg, 2001; Giuse & Kuhn, 2003; Lorenzi & Riley, 2003). In this article, the health information system means "a system, whether automated or manual, that comprises people, machines and /or methods organized to collect, process, transmit, and disseminate" data that represent user information in healthcare (Kuhn & Giuse, 2001, pp. 275). What is successful implementation and whose success is measured? Successes can be measured in many ways. Delone and McLean have been finding out the success factors of management information system which are also applicable to health information system. The success factors are: system qualities, *e.g.*, the ease of use or time savings, information quality, *e.g.*, completeness or data accuracy, usage, *e.g.*, the frequency of use or the number of entries, user satisfaction, *e.g.*, user-friendliness or overall satisfaction, individual impact, *e.g.*, changed work practices or direct benefits and organizational impact, *e.g.*, communication and collaboration or impact on patient care. Furthermore, user involvement during system development, implementation and organizational culture have been identified as possible factors measuring the success. However, the need for further research to determine which attributes are the most useful ones in measuring success has also been revealed. (van der Meijden, Tange, Troost & Hashman, 2003).

The different phases in implementation process are, in general, user needs and requirements analysis (specification), system design, initial system implementation and testing (Ahmad, Teater, Bentley, Kuehn, Kumar, Thomas & Mekhjian, 2002; Schuster, Hall, Couse, Swayngim & Kohatsu, 2003; Souther, 2001). The system requirements analysis includes workflow analysis, and the initial system implementation includes the technical installation of the information system, integration of the information system to other information systems and users' training. Project management is an important factor in every phase of the implementation project.

The purpose of this article is to highlight the health information system implementation process from end-user perspective. Which factors are crucial in the implementation process from the point of view of the end-users? How does project management contribute to the implementation process, what is the role of the end-user in system designing and how does training effect the information system implementation?

BACKGROUND

The lack of financial support was the most significant barrier to successfully implementing information technology in healthcare from both clients' and vendors' perspective. The vendors' inability to deliver products, and difficulties in achieving end-user acceptance or use were the other barriers from the point of view of the clients. (HIMSS, 2002.) Costs are often underestimated because the cost of the software is only the beginning of other expenditures, *e.g.*, person-hours for training and support have been forgotten (Ash, Stavri & Kuperman, 2003).

The social and organizational issues, not only the technical ones, are the critical issues in the implementation of information systems. The health information systems do not effectively support the health processes, and terminology for the healthcare environment is needed. (Ahmad et al., 2002; Berg & Toussaint, 2003; Berg, 2001; Giuse & Kuhn, 2003; Kuhn & Giuse, 2001; Littlejohns, Wyatt & Garvican, 2003).

Human-computer interaction is also perceived as unsatisfactory. The human-computer interaction indicates the means by which humans interact with computers, *e.g.*, users enter and retrieve data. To optimize the design of the human-computer interaction, concepts are needed (Berg, 2001; Kuhn & Giuse, 2001). Technical issues, *e.g.*, integration with other information systems and the need for open systems are also issues which must be solved (Giuse & Kuhn, 2003; Kuhn & Giuse, 2001).

The reasons for failures were that the complexity of healthcare tasks and social and professional cultures of healthcare organizations was not taken into account and, furthermore, the education of the users was insufficient and the timing of the education was wrong (Littlejohns, Wyatt & Garvican, 2003). Lorenzi and Riley (2003) report that the failures of the implementation of the health information system can be classified into four categories: technical shortcomings, project management shortcomings, organizational issues and information explosion. The technical failures contain, *e.g.*, the old system maintenance and staff training. Project management issues are, *e.g.*, project management skills. Organizational issues are concerned with constant changes. Information explosion means that knowledge has increased exponentially and new technical tools have been developed to cope with the information. Berg (2001) notes that it is important to notice that the implementation is not only a technical installation, and also that the project is not only a technical project but also an organizational development project.

The three major reasons that a project will succeed are user involvement, executive management support and the clear statement of requirements (Standish Group, 1994). Doolan, Bates and James (2003) reported that the factors associated with successful implementation are unusually strong leadership, a clearly defined long-term commitment, clear focus on improving clinical processes and gaining clinical involvement and support improving productivity. Lorenzi and Riley (2003) included technical skills, project management skills and people and organizational skills to the success factors. The skills mean knowledge, experience and abilities in each area. Ahmad et al. (2002) stated that success factors are a continuous executive support, engagement of physicians, an effective implementation team, a consistent user-friendly interface and on-going user support.

User involvement during system development, implementation process and organizational culture may explain the failure of the information system. The attributes assigned to system development were the extent of user involvement, redesigning work practices and the reconstruction of content and technical limitations. Communication, training and technical support were attributes addresses to implementation process. Organizational aspects attributes were organizational culture, *e.g.*, control and decision-making, management support, professional values as well as support and maintenance. (van der Meijden, Tange, Troost & Hashman, 2003.)

THE ROLE OF THE PROJECT MANAGEMENT

A project approach is the most common way to implement health information systems. Project management is the process of planning for organizing and controlling projects. From the end-users' point of view, the objectives of health information system projects must make explicit, *i.e.*, improve patient care or efficiency. It is recommendable to set a stage for improvements, *e.g.*, reduce the number of phone calls or move manual files to on-line files. The objectives of the project must also improve workflows and work practice, in other words the hospital managements and also clinicians involved in the project must also upgrade their work performance. The implementation of information system must add value for the end-user. Clear objectives motivate the end-users for implementation. (Ash, Stavri & Kuperman, 2003; Berg, 2001; Berg & Toussaint, 2003; Doolan, Bates & James, 2003, FitzHenry & Snyder, 1996; Giuse & Kuhn, 2003; Littlejohns, Wyatt & Garvican, 2003; Lorenzi, Riley, Blyth, Southon & Dixon, 1997; Nikula, Elberg & Svedberg, 2000; Lechleitner, Pfeiffer, Wilhelmy & Ball, 2003)

The information system implementation process must be seen as an organizational change process (Anderson & Stafford, 2002; Berg, 2001; Lorenzi, Riley, Blyth, Southon & Dixon, 1997). Change management, which means "the process of assisting individuals and organizations in passing from an old way of doing things to a new way of doing things" (Lorenzi & Riley, 2003 pp.200), should be taken into account from the start of the implementation process. Organizational resistance always occurs during the implementation of new information systems. The change management is one reason why the leader has an important role in projects. (Lorenzi & Riley, 2003; Lorenzi, Riley, Blyth, Southon & Dixon, 1997; FitzHenry & Snyder, 1996) Furthermore, the implementation process itself requires effective leadership (Ash, Stavri & Kuperman, 2003;Lorenzi & Riley, 2003; Souther, 2001). Leadership is needed at multiple levels in organizations; high-level leadership was considered the single most important factor. It was demonstrated by the long-term commitment of resources. (Ahmad et al. 2002; Doolan, Bates & James, 2003; Littlejohns, Wyatt & Garvican, 2003). At the executive level, leadership is needed to promote a shared vision the purpose of health information system, which is e.g. to improve patient care. At the project management level, the leadership is needed to make practical, effective and useful decisions (Ash, Stavri & Kuperman, 2003). The need for bottom-up participative and top-down authoritarian approaches are also useful when changing organizational behavior. (FitzHenry & Snyder, 1996; Souther, 2001). External consults can also be used as change agents. (FitzHenry & Snyder, 1996; Lechleitner, Pfeiffer, Wilhelmy & Ball, 2003; Souther, 2001)

There is a need for different people, and many types of skills, both technical and medical, in implementation process. The persons who are assigned to the projects must be influential and knowledgeable and they must represent significant organizational groups and professionals. The challenge is to find the right persons. The multidisciplinary team approach has proven advantageous in health information system implementation.(FitzHenry & Snyder, 1996; Lechleitner, Pfeiffer, Wilhelmy & Ball, 2003; Littlejohns, Wyatt & Garvican, 2003; Souther, 2001) There must be different levels of teams: an executive steering team, a project steering team and a project work team. The end-users are representatives in different levels of teams. It is important that physicians and nurses belong to the project steering team, which makes major policy decisions based on user needs and requirements, organization infrastructure and general implementation strategies. The end-users also belong to the project work team, whose focus is training, user-interface, testing, security policy and pilot implementation. (FitzHenry & Snyder, 1996; Souther, 2001) The end-users' participation to the project enhanced their involvement.

The organizational readiness for information system innovation could also be assessed in advance. The information about the organizational readiness could be used in planning implementation strategy (Nikula, 1999; Snyder-Halpern, 2001; Yetton, Sharma & Southon, 1999), *e.g.*, in what order to engage different parts of the organization or in which phase of implementation process managerial influence must be at its highest. The healthcare organizational readiness for information system innovation could be assessed by the validated eight sub-dimensions of knowledge, end-users, technology, administrative support, management structures, processes, resources, values and goals. The end-user readiness themes include, *e.g.*, skills to use computers or involvement in the change process. The information about the end-user readiness could be used in planning their training. (Snyder-Halpern, 2001)

THE END-USERS' ROLE IN DESIGNING INFORMATION SYSTEM

For the design of information systems, the conventional approach has been to model a system completely before implementing it (Berg, 2001; Kuhn & Giuse, 2001; Lorenzi, Riley, Blyth, Southon & Dixon, 1997). Prototyping has proven a successful project (Ahmad et al., 2002; Doolan, Bates & James, 2003; Lechleitner, Pfeiffer, Wilhelmy & Ball, 2003; Lorenzi, Riley, Blyth, Southon & Dixon, 1997; Schuster, Hall, Couse, Swayngim & Kohatsu, 2003). Prototyping requires an iterative approach, in which information system analysis, design, implementation and evaluation merge. It is evident that the information system changes work practices and vice versa. Therefore, the first step is to model the environment of the system. A socio-technical approach is suitable for system design. It means that the work practices in which information systems will be used are the starting points for design and implementation of the information systems. This means defining the core actors, their core tasks, the flow of work and the flow of data. The present workflows are analyzed, the future workflows described and documented. (Berg & Toussaint, 2003; Berg, 1999; FitzHenry & Snyder, 1996; Schuster, Hall, Couse, Swayngim & Kohatsu, 2003; Souther, 2001) The interactive model increases end-user involvement. The need for optimally adapting information systems to the work practice in the healthcare organizations and the users' terminology are important, as well as reducing workload and simplifying the workflow (Anderson & Stafford, 2002; Berg & Toussaint, 2003; Doolan, Bates & James, 2003; Giuse & Kuhn, 2003, Kuhn & Giuse, 2001; Lechleitner, Pfeiffer, Wilhelmy & Ball, 2003; Lorenzi, Riley, Blyth, Southon & Dixon, 1997). Health professionals must develop the information system together with the vendors. The interaction between the vendor, the teams installing the information system and the users is a crucial point. Representatives must include all different health professionals and administrative personnel (Anderson & Stafford, 2002; Berg & Toussaint, 2003; Doolan, Bates & James, 2003; Kuhn & Giuse, 2001; Lechleitner, Pfeiffer, Wilhelmy & Ball, 2003; Lorenzi, Riley, Blyth, Southon & Dixon, 1997 Nikula, Elberg, Svedberg, 2000; Schuster, Hall, Couse, Swayngim & Kohatsu, 2003; Sleutel & Guinn, 1999)

TRAINING AND SUPPORT

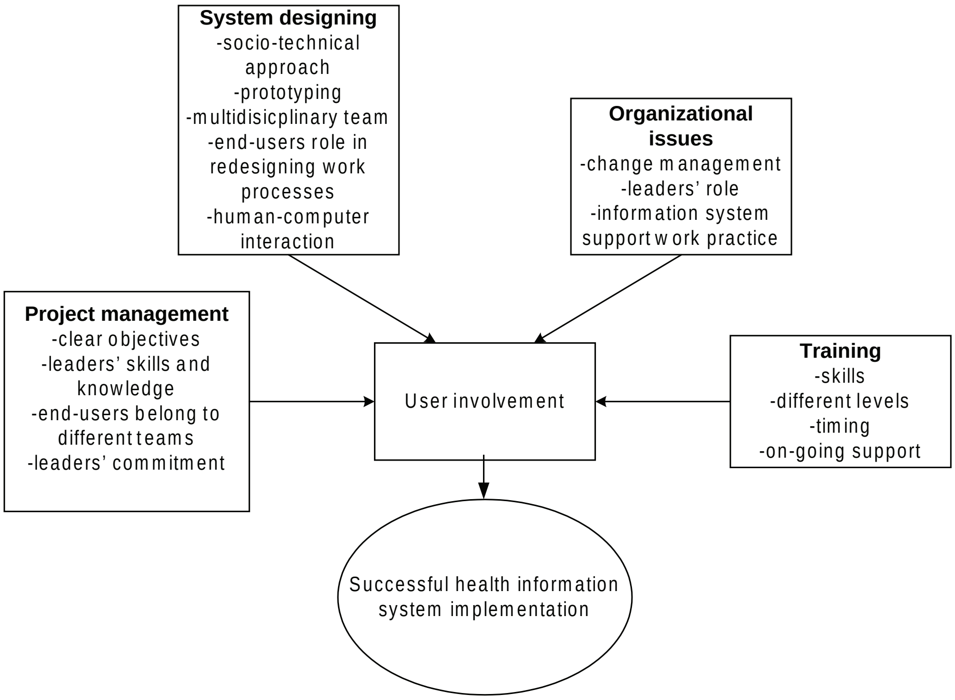
The proper training and support are an essential part of the health information system implementation process (Ahmad et al., 2002; Anderson & Stafford, 2002; Ash, Stavri & Kuperman, 2003; Doolan, Bates & James, 2003; Littlejohns, Wyatt & Garvican, 2003; Lorenzi, Riley, Blyth, Southon & Dixon, 1997; Schuster, Hall, Couse, Swayngim & Kohatsu, 2003; Souther, 2001). An implementation of the health information system requires extensive support and training in order to achieve user acceptance. The training plan must be completed before the initial implementation and it should also include intensive support at the time of the implementation. In the most successful implementation the end-users have more support after initial implementation than before it (Ash, Stavri & Kuperman, 2003).

There are also differences between computer literacy within healthcare professionals. (Saranto & Hovenga, 2004; Souther, 2001). Computer literacy means basic skills in computers use, *e.g.*, read and write. Information literacy consists of computer literacy, information retrieval abilities and communication skills. (Saranto & Hovenga, 2004) Therefore, the skills of the users, as well as their needs, must be taken into account when planning the training of a new information system. The quality of user training has a key role for a good user acceptance of information system (Anderson & Stafford, 2002; Quinzio, Junger, Gottwald, Benson, Hartmann, Jost, Banzhaf & Hempelmann, 2003). Disagreement on the efficiency of the system training has occurred (Lechleitner, Pfeiffer, Wilhelmy & Ball, 2003; Quinzio et al., 2003). The well-trained users were significantly more satisfied with the information system in routine use. A three-step training strategy has been recommended. First, the employees should be informed of the planned information system. The second phase includes general on-job training, and the third phase includes the opportunity to work on the individual problems of each user. (Quinzio et al., 2003) Whitman, Hamann & Vossler (1997) have also introduced a training plan. The strengths of the training plan are the contract that identified the roles and responsibilities of staff, management, and information services and staff development. Peer trainers have proven useful. Everyone must also have their own computer during training sessions. A competency test has been suggested as improvement to training plan. Ahmad et al. (2002) report a formalized training program for all users which precedes the implementation by two or four weeks. In addition, extensive on-site training occurred during the implementation period. The traditional help-desk phone has been found inadequate for user support.

FUTURE TRENDS

From the end-user perspective, the critical issues concerning the health information system implementation are project management, system designing, organizational issues and training and support. First of all, the project must have clear objectives. The teams must consist of different types of healthcare professionals, vendors, management and so on. It is important to find the right people to engage in the project. Project management task and responsibilities must be clear for each level of the teams and the interaction between the teams will also be efficient. The teams must work actively with the vendors. The management must also commit to the implementation process so the end-users notice that the project is important. The project

Figure 1. The factors of successful implementation process- end-users perspective



management aids the change management, which is one of the critical issues of information system implementation process. The leaders can use the information about the organizational readiness in planning the implementation strategy.

It is crucial that both representatives from the future users and leaders are involved in the design process. They are the experts of their work. The iterative development process will be used and the information systems adapting to workflows is important. When the information systems' requirements and future workflows are modeled properly it becomes evident that the work processes change, because the healthcare professionals must consider what they do and why they do it.

The training plan must be designed carefully. The information literacy of the end-users will be taken into account. A multilevel approach is good. First, the vendor can present the main functions of the information system. After that the tutors teach the users to use the information system and the users can train the system themselves. The users are also supported, they can ask questions. After that the users themselves can practice and also contact the tutors if they have any problems using the information system.

CONCLUSION

In the first stages the point of view in health information system implementation was only technical. The implementation process paid no attention to workflow patterns prior to the implementation, or to the computer experience among personnel or aspects of the organizational culture. Strategies for the successful management of health information system implementation must be planned carefully. The personnel of an organization must learn to use the information system and also change their working processes. Education is the critical point of the implementation process. Knowledge about end-user attitudes and skills prior to computerization is useful. The training will be tailored to meet the varying needs of user groups. There is a need for research in the area of evaluating health information system implementation.

REFERENCES

Ahmad, A., Teater, P., Bentley, T.D., Kuehn, L., Kumar, R.R., Thomas, A. & Mekhjian, H.S. (2002). Key attributes of a successful physician order entry system implementation in a multi-hospital environment. *Journal of the American Medical Informatics Association: JAMIA, 9*(1), 16-24.

Anderson, L. K., & Stafford, C. J. (2002). The “Big Bang”; implementation: not for the faint of heart. *Computers in Nursing, 20*(1), 14-20.

Ash, J. S., Stavri, P. Z., & Kuperman, G. J. (2003). A consensus statement on considerations for a successful CPOE implementation. *Journal of the American Medical Informatics Association: JAMIA, 10*(3), 229-234.

Berg, M. (1999). Patient care information systems and health care work: a sociotechnical approach. *International Journal of Medical Informatics, 55*(2), 87-101.

Berg, M. (2001). Implementing information systems in healthcare organizations: myths and challenges. *International Journal of Medical Informatics, 64*(2-3), 143-156.

Berg, M., & Toussaint, P. (2003). The mantra of modeling and the forgotten powers of paper: A sociotechnical view on the development of process-oriented ICT in health care. *International Journal of Medical Informatics, 69*(2-3), 223-234.

Doolan, D. F., Bates, D. W., & James, B. C. (2003). The use of computers for clinical care: a case series of advanced U.S. sites. *Journal of the American Medical Informatics Association: JAMIA, 10*(1), 94-107.

FitzHenry, F., & Snyder, J. (1996). Improving organizational processes for gains during implementation. *Computers in Nursing, 14*(3), 171-80.

Giuse, D. A., & Kuhn, K. A. (2003). Health information systems challenges: the Heidelberg conference and the future. *International Journal of Medical Informatics, 69*(2-3), 105-114.

HIMSS (2002). 13th Annual HIMSS Leadership Survey. Retrieved June 5, 2004, from *http://www.himss.org/2002survey/*

Kuhn, K.A. & Giuse, D.A. (2001). From hospital information systems to health information systems: Problems, challenges, perspectives. *Methods of Information in Medicine,* 40, 275-287.

Lechleitner, G., Pfeiffer, K. P., Wilhelmy, I., & Ball, M. (2003). Cerner millennium: the Innsbruck experience. *Methods of Information in Medicine, 42*(1), 8-15.

Littlejohns, P., Wyatt, J. C., & Garvican, L. (2003). Evaluating computerised health information systems: Hard lessons still to be learnt. *BMJ (Clinical research ed.), 326*(7394), 860-863.

Lorenzi, N. M., & Riley, R. T. (2003). Organizational issues = change. *International Journal of Medical Informatics, 69*(2-3), 197-203.

Lorenzi, N. M., Riley, R. T., Blyth, A. J., Southon, G., & Dixon, B. J. (1997). Antecedents of the people and organizational aspects of medical informatics: review of the literature. *Journal of the American Medical Informatics Association: JAMIA, 4*(2), 79-93.

Nikula, R. E. (1999). Organizational and technological insight as important factors for successful implementation of IT. *Proeedings of the AMIA Symposium,* , 585-588.

Nikula, R. E., Elberg, P. B., & Svedberg, H. B. (2000). Informed decisions by clinicians are fundamental for EPR implementation. *International*

Journal of Medical Informatics, 58-59, 141-146.

Quinzio, L., Junger, A., Gottwald, B., Benson, M., Hartmann, B., Jost, A., Banzhaf, A. & Hempelmann, G. (2003). User acceptance of an anaesthesia information management system. *European Journal of Anaesthesiology, 20*(12), 967-972.

Saranto, K., & Hovenga, E. J. (2004). Information literacy-what it is about?: Literature review of the concept and the context. *International Journal of Medical Informatics*, *73*(6), 503-513.

Schuster, D. M., Hall, S. E., Couse, C. B., Swayngim, D. S., & Kohatsu, K. Y. (2003). Involving users in the implementation of an imaging order entry system. *Journal of the American Medical Informatics Association: JAMIA*, *10*(4), 315-321.

Sleutel, M., & Guinn, M. (1999). As good as it gets? Going online with a clinical information system. *Computers in Nursing*, *17*(4), 181-185.

Snyder-Halpern, R. (2001). Indicators of organizational readiness for clinical information technology/systems innovation: a Delphi study. *International Journal of Medical Informatics, 63*(3), 179-204.

Souther, E. (2001). Implementation of the electronic medical record: the team approach. *Computers in Nursing*, *19*(2), 47-55.

Standish Group 1994. The CHAOS Report. Retrieved June 5, 2004, from *http://www.standishgroup.com/sample_research/chaos_1994_1.php*

van der Meijden, M. J., Tange, H. J., Troost, J., & Hasman, A. (2003). Determinants of success of inpatient clinical information systems: A literature review. *Journal of the American Medical Informatics Association: JAMIA, 10*(3), 235-243.

Yetton, P., Sharma, R., & Southon, G. (1999). Successful IS innovation: The contingent contributions of innovation characteristics and implementation. *Journal of Information Technology,* 14, 53-68.

Whitman, B. L., Hamann, S.K., & Vossler, B. L.(1997). A Training plan for Bedside Computers. *Journal of Nursing Staff Development*, *13*(1) 33-36.

KEY TERMS

Change Management: Change management is the process of assisting individuals and organizations in passing from an old way of doing things to a new way doing things.

Health Information System: Health Information System is the system, whether automated or manual, that comprises people, machines and /or methods organized to collect, process, transmit and disseminate data that represent user information in health care.

Implementation: The implementation of information system includes different phases: user needs and requirements analysis (specification), system design and initial system implementation and testing. The system requirements analysis includes workflow analysis; the initial system implementation includes technical installation of information system, integration the information system to other information systems and users' training.

Information Literacy: Information literacy consists of computer literacy and information retrieval abilities and communication skills. Computer literacy means basic skills in computers use, *e.g.*, being able to read and write with computers.

Project Management: Project management is the process of planning for organizing and controlling projects. The multidisciplinary team approach in project management means that there are different levels of teams: the executive steering team, the project steering team and the project work team. Different teams have their own tasks.

Prototyping: Prototyping is the development approach in which prototypes are produced. It is an iterative approach to the developing of the information system. The distinctions of the different phases of development of the information system, *e.g.*, analysis, design, implementation and evaluation are not clear.

Socio-Technical Approach: Socio-technical approach means that the work practices in which information systems will be used are the starting points for design and implementation of the information systems.

This work was previously published in Encyclopedia of Information Science and Technology, Vol. 5, edited by M. Khosrow-Pour, pp. 2678-2683, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 4.2
Perceived Level of Benefits and Risks of Core Functionalities of an EHR System

Diane C. Davis
Southern Illinois University-Carbondale, USA

Minal Thakkar
Southern Illinois University-Carbondale, USA

ABSTRACT

The need to adopt an electronic health record (EHR) system in United States (U.S.) hospitals seems to be more and more obvious when evaluating the benefits of improved patient safety, quality of care, and efficiency. The purpose of the study was to identify the status of EHR systems in U.S. hospitals in regard to the core functionalities implemented (as identified by the Institute of Medicine) and to determine if there was a significant relationship between perceived level of benefit and risk with the use of each core functionality, as well as if there was a significant relationship between the status of the EHR system and size of hospital. A national survey of U.S. hospitals was conducted to answer the research questions. The results showed that 37% had some components in **all** of the core functionalities of an EHR system, while 27% were using at least **some** functionalities. Health information and data, administrative processes, and results management were the three core functionalities that a majority of hospitals had as a part of their EHR system. A significant positive correlation between perceived benefits and risks was found in all of the eight core functionalities. There was no significant relationship found between status of EHR system and size of hospitals.

INTRODUCTION

With the growing need to provide the right information to the right person anywhere at anytime in today's global interconnected world, the U.S. healthcare industry has been moving toward an electronic health record (EHR) system. The need to adopt an EHR system in U.S. hospitals comes primarily from concerns regarding the quality of healthcare. Results of two studies of large samples using 1984 and 1992 data "imply that at least 44,000 and perhaps as many as 98,000 Americans die in hospitals each year as a result of medical errors" (Kohn, Corrigan, & Donaldson, 2000, p. 26). According to Aspden (as cited in Rippen & Yasnoff, 2004), universal availability of healthcare information and decision support through the national health information infrastructure can bring substantial improvements in patient safety and quality of care. The EHR, which is defined as "a secure, real-time, point-of-care, patient-centric information resource for clinicians" (HIMSS, 2003) is designed to provide this point of access to patient health information where and when it is needed by medical professionals.

EHR systems could save up to $81 billion in healthcare costs annually and improve health care quality (Hillestad et al., 2005). Financially sound hospitals and physician offices are leaping into adopting EHR systems (Goldschmidt, 2005). At the same time, some small hospitals and small physician offices are lagging behind in the use of EHR systems creating a digital divide (Goldschmidt, 2005). This may be due to lack of significant return on investment (ROI) in the short term, considering the high costs associated with the adoption of EHR systems.

Currently there are various initiatives carried out by governing agencies and healthcare associations in the area of promoting EHRs. David Brailer, national coordinator for health information technology, emphasized the important role EHR systems play in improving quality, increasing patient safety, increasing operational efficiency, and reducing costs. In his report on "The Decade of Health Information Technology: Delivering Consumer-centric and Information-rich Health Care: Framework for Strategic Action," Brailer said that reimbursing physicians for using EHR systems and reducing their risk of investing in them should accelerate the adoption of EHR systems in physicians offices (as cited in Mon, 2004).

At a conference in Baltimore on April 27, 2004, President Bush announced that most Americans will have EHRs within the next 10 years to allow doctors and hospitals to share patient records nationwide. To build upon the progress already made in the area of health information technology standards over the last several years, he proposed the FY2005 budget to include $100 million for demonstration projects (Cassidy, 2004). This will help test the effectiveness of health information technology and establish best practices for more widespread adoption in the healthcare industry (Administration Unveils 10 Year Health Information …., 2004).

Many organizations are working to develop initiatives and goals to help meet the needs of the healthcare industry. One initiative is Electronic Health Information Management (e-HIM) by the American Health Information Management Association (AHIMA). E-HIM goals are to: (1) promote the migration from paper to an electronic health information infrastructure, (2) reinvent how institutional and personal health information and records are managed, and (3) deliver measurable cost and quality results from improved information management ("AHIMA Mobilizes to Meet the e-HIM Call," 2003).

The Department of Health and Human Services (DHHS) charged a committee established by the IOM in May 2003 to:

- Provide guidance to DHHS on a set of "basic functionalities" that an electronic health record system should possess to promote patient safety.

The IOM committee considered functions, such as the types of data that should be available to providers when making clinical decisions (e.g., diagnoses, allergies, laboratory results); and the types of decision-support capabilities that should be present (e.g., the capability to alert providers to potential drug-drug interactions) (Institute of Medicine, 2003, p. 4).

Core functionalities of an EHR system and its components at that time as identified by the Institute of Medicine (IOM) committee were health information and data; results management; order entry/management; decision support; electronic communication and connectivity; patient support; administrative processes; and reporting and population health management. See the Appendix for a description of each of these functionalities.

PURPOSE AND GOAL OF THE STUDY

The overall goal of this study was to identify the core functionalities (as defined by IOM) being used by hospitals throughout the U.S. and the perceived level of benefits, costs, and risks associated with each of the functionalities. In order to meet this goal, a national survey was conducted to answer the following research questions:

- **RQ1:** What are the EHR core functionalities (such as health information and data, results management, order entry/management, decision support, electronic communication, patient support, administrative processes, reporting and population health management) utilized by healthcare systems?
- **RQ2:** Is there a significant difference between the status of the EHR system utilized and the size of the hospital (as measured by number of beds in the hospital)?
- **RQ3:** Is there a relationship between the perceived level of benefit and risk with the use of each of the core functionalities of an EHR system?

RELATED WORK

The status and the effects of using an EHR system are topics of growing interest to researchers. The most comprehensive study on the trends and usage of EHR has been conducted on a yearly basis by the Medical Records Institute (2005). The major role of the largest number of the 280 provider respondents of the 2005 study was information technology (IT) managers and professionals (42%) of which 11% were health information managers or MIS/CIS managers (p. 3). The motivating factors or driving forces for implementing EHR systems that were marked by 75% or more of these respondents were the need to (1) improve clinical processes or workflow efficiency, (2) improve quality of care, (3) share patient record information with healthcare practitioners and professionals, and (4) reduce medical errors (p. 6). The only major barrier to implementation marked by a majority of respondents (57%) was lack of adequate funding or resources (p. 22).

In regard to specific applications or functions, the ones used by a majority of respondents were in the Administrative and Financial Application; the Data Capture, Review, and Update Capabilities; and E-mail categories. Those used by 50% or more in the Administrative and Financial category were in billing and accounts receivable (66%), claims processing (63%), scheduling (61%), patient appointments, (59%), registration/admissions (59%), and charge capture and/or coding (53%) (MRI, 2005, p. 11). In Data Capture, Review, and Update Capabilities, demographics (67%), laboratory results (52%), medications being taken (50%), and allergies (50%) were the ones used by the majority of respondents (p. 13). In the e-mail category, only e-mail between practitioners was used by a majority of respondents (57%).

A study conducted in 2004 by AHIMA showed the industry is continuing to see more movement toward the EHR. The study (as cited in Zender, 2005) was conducted by Healthcare Informatics in collaboration with AHIMA at their 2004 national convention. It included 284 responses from respondents who primarily (80%) worked in clinical settings. Of these, 55% worked in hospitals while smaller groups worked in ambulatory care, long-term care, behavioral health, and others. A total of 83% of the respondents indicated they were HIM professionals (most having titles of director or manager). The study found that when organizations were asked to describe progress toward an EHR, 17% of respondents indicated extensively implemented; 26% indicated partially implemented; 27% said they were selecting, planning, or minimally implemented, and 21% indicated they were considering implementation and gathering information about it (Zender, 2005).

Medical personnel at the Adult Primary Care Clinic at the Medical University of South Carolina in Charleston, South Carolina, conducted a study to identify if the direct entry into the EHR system in the examination room by physicians had an effect on the physician-patient relationship as perceived by the patients themselves. They found the "use of the EHR had no negative impact on patients' perceived level of satisfaction with their physicians' interpersonal skills, the quality of the visit, or the perceived outcome of the care received" (Wagner et al., 2005, p. 38-39).

In a study conducted during the summer of 2004 by the American Academy of Family Physicians (AAFP), nearly 40% of respondents indicated they either completely converted to EHRs or were in the process of doing so and 24% with EHRs had purchased the system within the first half of the year (Carol, 2005). Cost seemed to be the primary barrier to EHR systems for physicians in small and medium-sized practices. According to Barbara Drury, an independent consultant of Pricare, a national consulting firm in Larkspur, CO (as cited in Carol, 2005), the median cost of implementing an EHR system was in the range of $30,000 per physician and could easily increase to $50,000 when including hardware, training, network upgrades, and IT services.

Previous research on EHR systems identified reducing medical errors, improving quality of care, conserving physician time, sharing patient information among healthcare practitioners, and workflow efficiency as the main benefits (Berman, 2004; Hier, Rothschild, LeMaistre, & Keeler, 2005; Valdes, Kibbie, Tolleson, Kunik, & Petersen, 2004). In regard to risks, privacy, security, and ROI were identified as major concerns (Bates, 2005; Hersh, 2004; Swartz, 2005). Few other studies conducted in the area of EHR reported the status of EHR systems in U.S. hospitals and physician offices (Gans, Kralewski, Hammons, & Dowd, 2005; "Small practices report success with EHRs," 2005; Valdes et al., 2004). These studies and other previous research conducted in the area of EHR systems studied the benefits, risks, and the extent of the usage of EHR applications and functions in U.S. hospitals and physician offices. They did not study the perceived level of benefits, risks, and costs associated with each of the core functionalities of an EHR system.

METHODOLOGY

In order to gain a better understanding of electronic record systems used in healthcare systems and terminology related to electronic health systems, the researchers conducted interviews at three hospitals within a 65-mile radius. Seven healthcare and information systems professionals from the three hospitals were interviewed. After these interviews and a thorough review of the literature, a draft of a survey instrument was developed. It was designed to gather answers to the research questions stated above. The survey instrument was reviewed by a panel of experts, which included an additional group of medical personnel and information systems personnel in local hospitals. Revisions were made to the survey instrument based on the comments of the reviewers. The survey was then approved by the Human Subjects Committee at the university employing the researchers prior to pilot testing. Next, the survey was sent to eight randomly selected hospitals from a national list of hospitals for pilot testing. Comments from these experts were reviewed and used as feedback for final revision of the instrument.

A database of randomly selected 1000 member hospitals of the American Hospital Association was purchased from Third Wave Research. A mailing including a cover letter, survey instrument, and drawing describing the core functionalities as identified by the IOM, and self-addressed return envelope was mailed in February 2005 to the Director of Health Information of these 1000 randomly selected hospitals. A follow-up mailing was sent in March. The response rate was slightly less than 10% and the findings described on the following pages are based on the 90 usable surveys that were returned. Responses to this study were coded onto Scantron sheets and analyzed using SPSS, version 14.

FINDINGS

Demographics of Respondents and Hospitals Surveyed

Fifty-eight percent of the respondents had 20 years or more of experience in the field; only 16% had less than 10 years. The mean for experience was 20 years. The majority of the respondents (61%) indicated their job title was health information manager or director of health information as seen in Table 1.

The specific titles in the miscellaneous "medical titles" category were Executive Director Physician Practice Network, Sr. VP Quality and Medical Staff Affairs, Transcription Coordinator, Coder in HIM, Patient Health Education Coordinator, and RN/DON. The titles in the miscellaneous "information systems" category were Systems Coordinator, Accounts Receivable /IT, Manager Patient Accounts/MIS, and Network Administrator. The "other" category contained respondents that indicated they had a title such

Table 1. Job titles of respondents

Title	Number	Percent
Director (or Manager) of Health Information	54	61.36
Director (or Manager) of Medical Records	13	14.77
HIM Manager/Director and Privacy Officer	3	3.41
Chief Information Officer/VP of Information Systems	3	3.41
Miscellaneous Medical Titles	6	6.82
Miscellaneous Information Systems' Titles	4	4.55
Other	5	5.68
Total	88	100.00

Table 2. Number of beds in facility

Beds	Number	Percent	Percent from AHA Statistics*
Less than 100	30	34.88	47.46
100 – 199	24	27.91	23.86
200 - 299	12	13.95	12.75
300 - 399	7	8.14	7.13
400 - 499	7	8.14	3.51
500 or more	6	6.98	5.29
Total	86	100.00	100.00

**Based on AHA 2003 statistics of 4,895 U.S. community hospitals.*

as Chief Financial Officer, Operations Manager, Team Leader, or Resource Manager.

The average (mean) for the number of beds in the facility was 209. The smallest hospital had 12 beds and the largest had 1460 beds. Table 2 shows the number of beds per facility as grouped according to the groupings used by the American Hospital Association (AHA). The average for the number of beds in the entire healthcare system was 546, with the smallest containing 20 beds and the largest containing 6000 beds.

The average number of hospitals in the healthcare system was 14 with a range from 1 to 172; 48% of the systems had only one hospital; 43% had 2 to 28 hospitals. The remaining 9% had more than 50 hospitals in their entire healthcare system.

Research Question #1

During the literature review and the interviews with healthcare and IT professionals, it was found that an EHR system could be purchased with some or all of the core functionalities depending on the funds available to purchase the software system. In an attempt to determine where hospitals were in regard to the use of EHR systems, the respondents were asked two questions. First, they were asked to mark the status of their facility in regard to the use of an EHR system from a list of options as shown in Table 3. The largest number of the respondents (37%) indicated they currently used some components in all of the eight core functionalities and 27% used only some of the eight core functionalities of an EHR system. Table 3 shows the number of responses and their corresponding percentages.

Second, they were asked to mark the core functionalities of the EHR system that their facility had or planned to have within the next five years (Research Question #1). Health Information and Data, Results Management, and Administrative Processes were the only core functionalities that were currently part of or interfaced with the EHR system in more than 50% of the respondents' facilities. The responses are shown in Table 4.

Research Question # 2

To answer the second research question, which was to determine if there was a relationship between the status of the EHR system and the size of the hospital as measured by the number of beds in the hospital, a one-way ANOVA was conducted. The overall F did not reveal any significant differences among the groups as seen in Table 5 (the groups are the levels of EHR status as listed in Table 3). In other words, there was no relationship between

Table 3. Status of facility in regard to use of EHR system

Status of Facility	Number	Percent
Currently using some components in all the core functionalities of an EHR system	33	37.08
Currently using some functionalities of an EHR system	24	26.97
Currently using some electronic records systems and plan to interface these into an EHR system within the next two years	9	10.11
Currently using some electronic records systems and plan to interface these into an EHR system sometime between 2 and 5 years from now	13	14.61
Not currently using any electronic records systems, but plan to use an EHR system sometime in the future	10	11.24
Total	89	100.01

Table 4. Status of the core functionality within facility

Core Functionality	Already a part of the EHR system or it is interfaced with the EHR system	Plan to interface within the next 5 years	Do not plan to interface or Not sure
Health Information and Data	64%	32%	4%
Results Management	56%	35%	9%
Order Entry/Management	46%	46%	7%
Decision Support	35%	39%	26%
Electronic Communication and Connectivity	48%	36%	16%
Patient Support	21%	36%	42%
Administrative Processes	57%	25%	17%
Reporting and Population Health Management	39%	27%	34%

the status of the EHR system and the size of the hospital. The significance of the results was assessed at the 0.05 alpha level.

General Benefits and Risks

The respondents were asked to evaluate the benefits, risk factors, and cost justification of the core functionalities for which they currently had interfaced or planned to interface into their EHR system within the next five years. They were to evaluate the core functionalities in regard to (1) the degree each one had benefits related to the delivery and quality of healthcare, (2) the degree each had risk factors in relation to cost, cultural change, security of information, dependency upon information, computer downtime, etc., and (3) the degree to whether or not each one was cost justifiable in regard to general terms such as whether the benefits in the long run outweighed the problems, risks, etc., and the anticipation that there would be some ROI in the future. The core functionality the respondents indicated as the most beneficial (one with the highest mean) was "health information and data," closely followed by "electronic communication and connectivity." The functionality with the greatest risk (highest mean for risk factors) was "electronic communication and connectivity," closely followed by "results management." The two functionalities that the respondents found to be the most cost justifiable were "health information and data" and "results management." The means for all three factors for each of the functionalities can be seen in Table 6.

Research Question #3

To answer the third research question, to determine if there was a relationship between perceived level

Table 5. Status of EHR system compared to size of the hospital

Status of EHR	Sum of Squares	df	Mean Square	F	p-value
Between Groups	364250.21	4	91062.553	2.011	.101
Within Groups	3622448.1	80	45280.601		
Total	3986698.3	84			

Table 6. Mean of benefits, risks, and costs associated with each of the core functionalities

Functionality	Benefits	Risk Factors	Cost Justification
Health Information and Data	8.83	6.71	7.84
Results Management	8.64	6.96	7.74
Order Entry/Management	8.49	6.77	7.59
Decision Support	7.71	5.95	6.45
Electronic Communication and Connectivity	8.69	7.03	7.21
Patient Support	7.71	6.67	6.83
Administrative Processes	8.25	6.59	7.46
Reporting and Population Health Management	7.58	6.23	6.83

of benefit and risk, a Pearson correlation was conducted. First, it was determined if an overall correlation existed between benefits and risks for each core functionality. Then the correlation was calculated again for each of the groups separately to test where the significance occurred and which group had a higher correlation between benefits and risks (those who had the functionality or with those who did not currently have it, but planned to within the next five years). Those who indicated they did not plan to have the core functionality or were not sure, were not included.

There was a significant positive correlation between benefits and risks for all eight core functionalities. The correlation was significant at the 0.01 alpha level. This positive correlation indicated that the respondents associated high benefits with high risks. For three of the eight core functionalities--Health Information and Data, Administrative Processes, and Reporting and Population Health Management--there was a significant positive correlation for both groups (those who already had it and those who planned to have it). For the other five functionalities, the significance existed only in the group that planned to have it. Therefore, for the group of hospitals who already had the core functionalities of an EHR system, the respondents did not seem to associate as high a risk with the benefits as those who were just in the planning stages.

In the area of Health Information and Data, the correlation was 0.544 between perceived benefits and risks. When looking at each group individually (those who had this functionality in their EHR system and those who planned to have it within the next five years), the correlation was twice as much for the group who planned to have this functionality (0.742) compared to those who already had it (0.361).

In the area of Results Management, there was again an overall high positive correlation between benefits and risks (0.578). However, this time when looking at the groups individually, a significant positive correlation (0.681) existed only in the group who planned to have this functionality (and not those who already had it). These results and those for the other core functionalities can be seen in Table 7.

DISCUSSION

The majority of the respondents (61%) indicated their job title was health information manager or director of health information. This is a much larger percent than the 11% of respondents to the 2005 Medical Records Institute's Survey of EHR trends and usage (which included physicians as well as those in other roles), but slightly less than

Table 7. Correlation between perceived benefit and risk for each functionality (overall and individual correlation for groups respectively)

Core Functionality	Overall Correlation for Both Groups	Calculated p value	Correlation for Group that Already Had Functionality	Calculated p value	Correlation for Group that Planned to Have Functionality	Calculated p value
Health Information and Data	.544*	.000	.361*	.011	.742*	.000
Results Management	.578*	.000	.247	.120	.681*	.000
Order Entry/ Management	.597*	.000	.134	.444	.738*	.000
Decision Support	.694*	.000	.205	.337	.816*	.000
Electronic Communication and Connectivity	.533*	.000	.087	.626	.723*	.000
Patient Support	.664*	.000	.414	.125	.722*	.000
Administrative Processes	.619*	.000	.364*	.021	.936*	.000
Reporting and Population Health Management	.666*	.000	.396*	.033	.908*	.000

**p = 0.01*

the 83% from the AHIMA study which indicated they were HIM professionals. The average (mean) for the number of beds in the facility's surveyed was 209; 35% of the hospitals had less than 100 beds. The most recent American Hospital Association statistics for bed size of hospitals in the U.S. was consulted to determine if the study sample provided a qualitative representation of U.S. hospitals. Thirty-five percent of the hospitals responding to the survey reported 1 to 99 beds while the 2003 AHA statistics reported approximately 47% hospitals in the category of less than 100 beds. Therefore, this study had responses from fewer small hospitals than there were in the actual population of hospitals. However, for all the other categories of hospitals by bed size, it was pretty comparable, varying only 2 to 5% within each bed size category (AHA, 2003). See comparisons in Table 2.

The largest number of the respondents (37%) indicated they currently used some components in all of the eight core functionalities and 27% used only some of the eight core functionalities of an EHR system. The study conducted by AHIMA, although worded slightly differently, found that 17% of organizations indicated they had extensively implemented EHR systems and 26% indicated partially implemented (Zender, 2005). These two studies had a slightly different focus and way of analyzing the status of the EHR system. The AHIMA statement asked respondents to describe their organization's progress toward implementation of an EHR, and this study sought to discover the status of their facility in regard to the use of the EHR system by looking at how many of the core functionalities were used or interfaced into the EHR system.

The overall F in this study did not reveal any significant differences among the status of the facilities' use of an EHR system in relation to the size of hospital. However, other research indicated that some small hospitals and small physician offices were lagging behind in the use of EHR systems creating a digital divide (Goldschmidt, 2005). In order for EHR systems to be seamlessly used and offer the intended benefits, they must become more affordable and see some return on investment (ROI).

Health Information and Data, Results Management (lab test results, radiology procedures, and other results reporting), and Administrative Processes were the only core functionalities that were currently part of or interfaced with the EHR system in more than 50% of the respondents' facilities. This seems to correspond fairly closely to the MRI Survey which found that specific applications or functions used by at least a majority of respondents were in the Administrative and Financial Application; the Data Capture, Review, and Update Capabilities; and E-mail categories. Laboratory results, which fell under the category of Data Capture, Review, and Update Capabilities, of the MRI survey were used by over 50% of respondents.

There was a significant positive correlation between perceived level of benefits and risks for all core eight functionalities. For the majority of the core functionalities, the positive associations were stronger for those in the group that planned to have it (than the group that already had it). Since all the overall correlations between risk and benefits were positive, we can conclude that the respondents associated high benefits with high risks. For the group of hospitals who already had the core functionalities of an EHR system, the respondents did not seem to associate as high a risk with the benefits as those who were just in the planning stages. This may imply that once the functionality is implemented the perceived level of risk decreases.

LIMITATIONS

The study may have some limitations. First, the low response rate may make it difficult to generalize the findings to the total population. Second, although the study identified that a correlation existed between the risks and benefits for each of the core functionalities among the groups, it did not identify any specific risks or benefits for each of the functionalities.

CONCLUSION AND FUTURE WORK

Each of the eight core functionalities can be adopted by hospitals individually or as an entire EHR system. This study identified the majority of the core functionalities used in the hospitals, as well as the correlation between the benefits and risks associated with these core functionalities. Other hospitals that are in the planning and information gathering stages can use this information to prioritize the core functionalities that they plan to adopt.

Based on the findings of this study, it is suggested that some of the high risks associated with EHR systems may be misconceptions by professionals in the field who have not yet adopted an EHR system. The benefits have been stated over and over again. However, little research has been done on the risks associated in using these systems. Future work in the area of EHRs should identify risks associated with the use of each of the core functionalities of an EHR system and ways to avoid these risks or come to the realization that they may not be as problematic as once thought.

This study identified a perceived level of cost associated with the benefits and risks in adopting each of the core functionalities. Since ROI is a barrier to adoption of EHR systems, future study of cost-benefit factors related to each of the core functionalities is recommended. An implication of this would be that hospitals and physicians will gain a better understanding of the cost factors related to each of the functionalities and be able to prioritize and procure the functionalities of an EHR system as needed.

REFERENCES

Administration unveils 10 year health information technology strategic framework. (2004). *E-health initiative*. Retrieved August 28, 2004, from http://www.ehealthinitiative.org/initiatives/policy/administration.mspx

AHIMA mobilizes to meet the e-HIM call. (2003). *AHIMA Advantage*, *7*(1), *1-3*.

American Hospital Association. (2005). *AHA hospital statistics: The comprehensive reference source for analysis and comparison of hospital trends*. Chicago: AHA, p. 10.

Bates, D. (2005). Physicians and ambulatory electronic health records. *Health Affairs, 24*(5), 1180-1189.

Berman, J. (2004). Safety centers and EMRs. *Health-IT world*. Retrieved August 11, 2004, from http://www.health-itworld.com/emag/050104/183.html

Carol, R. (2005). EHRs, the doctor will see you now. *Journal of American Health Information Management Association*, *76*(4), 24-28.

Cassidy, B. S. (2004). Skills for success in managing an EHR environment. *Advance Online Editions for Health Information Professionals.* Retrieved June 18, 2004, from http://www.advanceforhim.com/common/Editorial/PrintFriendly.aspx?CC=34975

Gans, D., Kralewski, J., Hammons, T., & Dowd, B. (2005). Medical groups' adoption of electronic health records and information systems. *Health Affairs*, *24*(5), 1323-1333.

Goldschmidt, P. G. (2005). HIT and MIS: Implications of health information technology and medical information systems. *Communications of the ACM*, *48*(10), 69-74.

Hersh, W. (2004). Health care information technology: Progress and barriers. *Journal of the American Medical Association, 292*(18), 2273-2274.

Hier, D., Rothschild, A., LeMaistre, A., & Keeler, J. (2005). Differing faculty and house staff acceptance of an electronic health record. *International Journal of Medical Informatics, 74*(7/8), 657-662.

Hillestad, R., Bigelow, J., Bower, A., Girosi, F., Meili, R. et al. (2005). Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. *Health Affairs, 24*(5), 1103-1117.

HIMSS. (2003). *Electronic health record definitional model version 1.0*. Retrieved July 26, 2004, from http://www.himss.org/content/mindmaps/EHR/multi-maps/resources/resources.htm

Institute of Medicine. (2003). Key capabilities of an electronic health record system. Retrieved July 26, 2004, from http://books.nap.edu/html/ehr/NI000427.pdf

Kohn, L. T., Corrigan J. M., & Donaldson, M. S. (2000). *To err is human: Building a safer health system*. Washington, DC: National Academy Press.

Medical Records Institute. (2005). *Medical records institute's seventh annual survey of electronic health record trends and usage for 2005*. Retrieved October 5, 2005, from http://www.medrecinst.com/files/ehrsurvey05.pdf

Mon, D. T. (2004). Next steps for the EHR draft standard: Core functionality and conformance criteria key for accreditation. *Journal of American Health Information Management Association, 75*(10), 50-51.

Rippen, H. E., & Yasnoff, W. A. (2004). Building the national health information infrastructure. *Journal of American Health Information Management Association, 75*(5), 20-26.

Small practices report success with EHRs. (2005). *Family Practice Management, 12*(4), 40.

Swartz, N. (2005). Electronic medical records' risks feared. *Information Management Journal, 39*(3), 9.

Valdes, I., Kibbe, D., Tolleson, G., Kunik, M., & Petersen, L. (2004). Barriers to proliferation of electronic medical records. *Informatics in Primary Care, 12*(1), 3-9.

Wagner, K. A., Ward, D. M., Lee, F. W., White, A. W., Davis, K. S., & Clancy, D. E. (2005). Physicians, patients, and EHRs. *Journal of American Health Information Management Association, 76*(4), 38-41.

Zender, A. (2005). Ready for the EHR? A new survey measures EHR implementation and individual readiness. *Journal of American Health Information Management Association, 76*(3), 54-55.

APPENDIX A.

1. **Health Information and Data:** A defined dataset that includes medical and nursing diagnoses, a medication list, allergies, demographics, clinical narratives, and laboratory test results for access by care providers when needed.

2. **Results Management:** A feature to manage the lab test results, radiology procedures results, do results reporting, results notification, and multimedia support—images, waveforms, pictures, sounds.

3. **Order Entry/Management:** Computerized provider order entry (CPOE) for such areas as electronic prescribing, laboratory, microbiology, pathology, XR, ancillary, nursing, supplies, consults. Even with little or no decision support, they can still improve workflow processes by eliminating lost orders and ambiguities caused by illegible handwriting, generating related orders automatically.

4. **Decision Support:** A computerized decision support system, which enhances clinical performance by providing drug alerts, other rule-based alerts, reminders, clinical guidelines, and pathways. It also helps in improving drug dosing, and drug selection. It can be used for chronic disease management, clinician work lists, diagnostic decision support, and automated real-time surveillance.

5. **Electronic Communication and Connectivity:** Electronic communication can be between provider-provider, patient-provider, trading partners such as pharmacies, insurers, laboratory, radiology, and among team members for coordination. Electronic connectivity includes integrated medical record within facility, within different facilities of the same healthcare system, and among different healthcare systems.

6. **Patient Support:** Patient support includes patient education, family and informal caregiver education, data entered by patient, family, and/or information caregiver such as home monitoring

7. **Administrative Processes:** Administrative processes include electronic scheduling systems for hospital admissions, inpatient and outpatient procedures, and identifying eligible or potential eligible patients for clinical trials.

8. **Reporting and Population Health Management:** This feature would report patient safety and quality data, public health data, and disease registries. It makes the reporting process less labor-intensive and time-consuming (Institute of Medicine, 2003).

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 1, Issue 4, edited by J. Tan, pp. 55-67, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 4.3
Virtual Reality Simulation in Human Applied Kinetics and Ergo Physiology

Bill Ag. Drougas
ATEI Education Institute of Epirus, Greece

ABSTRACT

Virtual reality is today an excellent tool for a full simulated experience in a modern environment where any researcher or any individual scientist may work with vital experimental environments or use parameters that sometimes does not really exist. It is already a vital step for the future of science and for the modern experiment. Ergo physiology today has many applications for research. We can find new unknown parameters for the human body searching biokinetics and ergo physiology, and it is time to use modern technologies and applications. The vital issues discussed in this chapter may offer many applications for human kinetics and movement and may also discuss biokinetics research using the physical laws and parameters in various biokinetics and physiology fields.

INTRODUCTION

Virtual reality is today an important part of modern scientific methodology and research, using modern high-speed computers of lately designed technologies for research and simulation in various scientific fields such as ergo physiology and biokinetics. This is a new field in contemporary science, methodology, and experimentation, and we can recognize that during the past years, it has continued with great success. An important field using this new way of research is the simulation of human movement in ergo physiology and applied biokinetics science. Virtual reality is very useful for researchers in these fields because they can have simulations of the physical human body at any time they want for study or experimentation.

BACKGROUND

One of the first authors who wrote about virtual reality was Howard Rheingold (1991), who wrote about data visualization and 3-D CAD (computer-aided design) in which someone may use his or her hands and fingers. Many applications can be found from the middle war period by the U.S. Air Force to create flying simulations.

Myron Krueger, during the '60s, worked on the affiliation of human and computer with special research in computer-controlled responsive environments, which were named by him artificial reality (Krueger, 1993). He also designed the video place, a system that contains a projection screen and a video camera that is controlled by a computer. By this method, human movements in each activity are transferred to computer graphics in software (Boudouridis, 1994).

So, there can be a connection between human and technical things in space with computer graphics. This was one of the first methodologies in human-kinetics research and applications.

Tom Furness was another scientist who designed the Super Cockpit for the U.S. Air Force after many years of research. In a small place, a human could use computers and a HMD (head-mounted display) to understand vital secrets of the flight without any danger (Furness, 1991).

But the man who is the father of the terms virtual reality and reality engine is Jaron Lanier, an informatics scientist who, with another young man named Tom Zimmerman, established the Visual Programming Language Research Inc in 1980 (Boudouridis, 1994). This company was the first to make important tools for virtual reality programs and applications, such as data gloves and HMDs.

ISSUES

In the beginning, many other scientists worked with computer data for virtual reality applications in various fields with very big success.

Today, there are many different fields and applications of virtual reality technology. Table 1 summarizes some of the virtual reality applications similar to those of ergo physiology.

Especially in the fields of ergo physiology and biokinetics, virtual reality is used in many applications. Some of the characteristics of human movement, the human body, and parameters such as space, geometry, color, and sound may help virtual reality programs become more effective in various methodologies of research and virtual applications.

The importance of this is to find a methodology for using virtual reality and a way to recognize the results, such as some of the official physiology results that can give to researchers many new discoveries in existing science and theory, or future science research in finding new signals from the human body during simulations.

Today we have all the modern technology to make better simulations for the human body and to see new fields that had not previously existed. This is, of course, the future of research.

For example, it is not possible for anyone to fly at high speeds without any danger so that scientists can see the behavior of the human body or parameters such as blood pressure, the behavior of the heart, muscle energy, signals or other problems in the eyes, and so forth. But with virtual reality simulations, we can today register

Table 1. Summary of the virtual reality applications similar to those in ergo physiology fields

- Human behavior in flying simulations
- Human behavior in space simulations
- Learning and human-movement programs in kinetics
- Neurodisease science
- Rehabilitation
- Physical behavior in space
- Step-correction learning and research
- Adapted methodologies in kinetics
- Special gymnastics programs
- Study of the human senses and their characteristics
- Pain confrontment
- Continuing education

Table 2. Important parameters in virtual reality simulations for ergo physiology and human movement

- Length
- Height
- Time
- Sound frequency
- Color frequency
- Wave length
- Velocity
- Pressure
- Area
- Energy
- Power

many of these parameters and standards, see how they change after the experimentation, and see it is used in our theories.

The very best aspect of this method is that we can stop the experiment when we must or when we want to begin the experiment again from the last step, or, if we want, we can design another model with new parameters. So, we can always put something new into our experimentation and theory, and this is important because every person is different from every other. By this methodology, we can register statistical effects or make a very good mathematical analysis of the problem to continue with advanced research and measurements or to make other project experiments.

So many past methodologies or past technologies will be changed after these experiments. Ergo physiology and applied human kinetics today use many important laws of physics, physiology, mathematics, statistics, and, of course, medical data, and it is important to continue to grow using new technologies like virtual reality. Table 2 summarizes some of the parameters that are important in virtual reality simulation for the human body and movement.

In some experiments, all of the parameters in Table 2 can be used, and in some others, only a few of them can be used. However, the parameters in the simulations are very near to those of physics, and the philosophy is about the same in similar experiments.

Table 3 summarizes some of the fields in which virtual reality can help to recognize and study the physical parameters of Table 2.

In every subject of the study of human abilities, there are various techniques and methodologies. However, the philosophies of all methodologies are about the same and have similar directions.

So it is important to understand that with virtual reality simulations in ergo physiology, we may simulate human models in an environment that does not really exist physically. In this way we can find new and unknown parameters of the basics of human biokinetics and ergo physiology. In the same way, a researcher can establish new models of experiments and fields for research using many parameters of the modern population and all the modern kinetic problems that appear in people. Researchers can include these problems in the data of the simulation and study the methodology or the effects within the simulation. For example, the influence of a brain problem on human kinetics ability can be registered from various kinetics parameters and/or kinetics behavior and movement in various daily skills. Many years before, this was not so clear, and many of these problems seemed like they didn't exist. In simulations, we have two bodies: the virtual and the real.

It is important to understand that two human bodies in the same experiment are the same but

Table 3. Summary of the fields of the physical parameters in virtual reality

- Kinetics conception
- Space-creation parameters
- Relative-movement creation
- Color-conception studies
- Sound-conception studies
- Kinetics-parameters registration
- Study of velocity
- Study of frequency
- Study of space parameters
- Study and registration of time

also different.

We have two human bodies during the experiment that we can use, or many more if it is possible with the program. Virtual reality biokinetics research programs must have both human bodies: the real and the virtual one. The parameters and simulations must be very close between these two categories. These two bodies are different persons, but closely exhibit the same movements. It is very important to recognize the abilities of the bodies. The information about kinetics behavior or any other information from any data structure that defines the behaviors of the movement and kinetic parameters of one virtual human can be passed directly to all the other virtual humans if there is more than one human in the program of the experiment. In this way, we can have parameters and measurements from various kinetic models at the same time and study many different behaviors. This is the philosophy of a multiple experiment.

Two or more students or researchers can work with two or more models at the same time and give different conclusions and different theories. Virtual humans are real models that can be used many times for the same work and for the same or a different research. This is very good for the physiology studies and medicine experiments for human biokinetics and behavior, and, of course, for education programs for researchers or students at the graduate or postgraduate levels.

Table 4 summarizes some of the simulated and virtual applications to human abilities.

Many of the methodologies that exist today began first in front of a computer screen and, with the help of modern programs of simulation in physiology, biology, and mathematics, they continue to establish some of the new theories.

Today, electronic methods give us the opportunity to continue our research in a safer work environment; in the past, various experimental problems in the physical environment were impossible to avoid. Virtual reality and its development are currently underway in a number of organizations such as NASA, IBM, Boeing, and so forth, and also in many official research programs of famous university laboratories and technological educational institutes. In many of these institutes, there exist many new international programs of human study and ergo physiology research.

Table 4. Summary of some of the simulated and virtual applications

- Simulation of ergonomic environments
- Simulation-based training
- Rehabilitation and modern orthopedics
- Virtual patients and human organs for surgery
- Psychotherapy and applied physiology
- Sports simulations
- Physiological behavior with various phobias
- Biomechanical walking
- Virtual training for balance

CONCLUSION

Virtual reality is today an excellent tool for a fully simulated experience in a modern environment where the researcher or simple scientist may work with vital experimental environments or use any parameter that sometimes does not really exist. This is important for people who are researching the abilities of the human body. The vital issues discussed in this article may offer many applications for human kinetics and movement, and biokinetics research using physical laws and parameters. In the same philosophy, any researcher may propose or design new models for informatics scientists. Ergo physiology today has many applications for research, and the modern computer systems for simulations and virtual reality give us many applications and the opportunity to work in easier experimental environments to make and design experiments using real and unreal parameters without any danger for us or any other human. This is already a step toward the future of science and the experiments.

REFERENCES

Boudouridis, M. (1994). *The technology of virtual reality*. Presentation to the International Summer School of Communications, Athens, Greece.

Computers: Artificial intelligence. (n.d.). *Time Life*, 73-77.

Deisinger, J., Breining, R., & Robler, A. (n.d.). ERGONAUT: A tool for ergonomic analysis in virtual environments. In J. D. Mulder & R. van Liere (Eds.), *Virtual Environments 2000, Proceedings of the 6th Eurographics Workshop on Virtual Environments*, Amsterdam, The Netherlands. New York: Springer Wien.

Drougas, B. Ag. (n.d.). Rehabilitation by new technologies. *Research and Theory Journal*, 12-15.

Drougas, B. Ag. (2003). *Theory and laboratory applications for the lesson of teleinformatics and applications in medicine*. ATEI of Epirus, Department of Teleinformatics.

Drougas, B. Ag. (2004a). *Home care and the modern technologies* (1st ed.). D-publications Arta Greece.

Drougas, B. Ag. (2004b). *Telemedicine applications in a contemporary environment*. D-publications Arta Greece.

Drougas, B. Ag. (2004c). *Virtual reality and health: Manual of the Laboratory of the Teleinformatics and Applications in Medicine*. ATEI of Epirus, Department of Teleinformatics.

Furness. (1991). Harnessing virtual space. *Society for Information Display Digest of Technical Papers* (1st ed.).

Hoffman, H. G. (2004). Therapy applications of virtual reality. *Scientific American* (Greek ed.), 28-36.

Koutsuris, D., Nikita, K., & Pavlopoulos, S., (2004). *Medical represented systems*. Thesaloniki, Greece: Tziolas Publications.

Koutsuris, D., Pavlopoulos, S., & Prentza, A. (2003). *Introduction into biomedical technology and analysis of medical signals*. Thesaloniki, Greece: Tziolas Publications.

Krueger, M. (1993). *Artificial reality*. New York: Addison-Wesley.

Rheingold, H. (1991). *Virtual reality*. New York: Simon & Schuster.

Thalmann, D. (n.d.). *Animating autonomous virtual humans in virtual reality*. Swiss Federal Institute of Technology, Computer Graphics Lab.

Zeimbekakis, G., (2003). Telematics applications. Athens: Modern Season Publications.

KEY TERMS

Biokinetics: A scientific division of ergo physiology for the research and recognition of physical laws, standards, and existing parameters of human movement and kinetics, or the changes from any external or internal effect that can change any physiological movement. The creation of a physical model helps researchers to understand the problems of movement and kinetics and their relation with various pathological problems.

Ergo Physiology: A division of physiology that searches physiological and physical parameters, constants, and standards for the creation of energy, power, balance, velocity, and other physical changes in the human body. This division includes a vital search for the physical and physiological laws related to various kinetics problems for humans.

Kinetics Behavior: A number of characteristics of human movement including the physical behavior of the human body in movement and kinetics. These characteristics draw a behavior model for the physical body and offer the opportunity to researchers in the area of physiology and biokinetics to search the behavior of humans and to create scientific programs for rehabilitation.

Physical Biokinetics Parameters: A number of different parameters, laws, and measurements that are recognized from a physical model from physiology and human movement. We can use this model in various applications recognizing human behavior and movement.

Rehabilitation: A scientific methodology using physical methods, especially applied movements, kinetics, and modern psychology applications, to change the human body and mind behavior, helping people to recreate their energy levels and their physical ability, or to use the near environment in the best way to live with their problems for a better life.

Simulation: A number of presentations of different parameters for things, models, or environments designed from physical standards and laws. In this virtual presentation, one can use his or her experience to have an interactive participation and use this virtual environment for personal training, education, rehabilitation, and so forth.

Virtual Human: A virtual human body in a 3-D immersive, interactive simulation, created by computer software, for the parallel presentation of reality in the virtual using any influence from the real environment.

Virtual Reality: An immersive, interactive simulation for real and unreal environments. We can also use this term for simple interactive models that cannot offer to a user any immersive presentation in a virtual environment. Virtual reality is the most useful tool in medicine today, with many applications in various fields, especially in rehabilitation, education, and so forth.

Visualization: A medicine application using a 3-D human body model that has been made from a number of different 3-D scans from a medical computed tomography.

This work was previously published in Handbook of Research on Informatics in Healthcare and Biomedicine A.A.Lazakidou, edited by G. Ganiatsas, pp. 246-254, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 4.4
Implementation of an Error-Coding Scheme for Teleradiology System

Shobha Rekh
PSG College of Technology, India

Subha Rani
PSG College of Technology, India

Hepzibah Christinal
Karunya Institute of Technology & Sciences, India

Easter Selvan
Université de la Méditerranée, France

ABSTRACT

Teleradiology, sending of x-rays, is the most common application of telemedicine in use today. Mobile telemedicine is the latest area of research where the patient can be monitored remotely by a doctor. In this paper, we propose a teleradiology system whereby the communication is established via cellular networks and cell phones. Since the data is sent over a wireless channel, it is more prone to data loss due to the mobile environment. To prevent data from getting lost due to noise in the wireless channel, an error-coding scheme is applied. Turbo code is a powerful error coding technique employed nowadays in communication systems. In this paper we apply turbo codes to an x-ray image and an ECG image and simulate the transmission system by adding Gaussian noise to the image. The performance of Turbo codes, in terms of bit error rate is better than other error coding schemes even in the region where the signal strength is very low. The Quality of the image is retained in the receiving end by proper design of the error-coding scheme. An analysis has been done on various parameters considered in the design of turbo codes. The images obtained

after decoding are found to be suitable for recognition and diagnosis by the doctors in their mobile phones. This novel technology will enhance the health care in rural area where the opinion of a specialized doctor is not available.

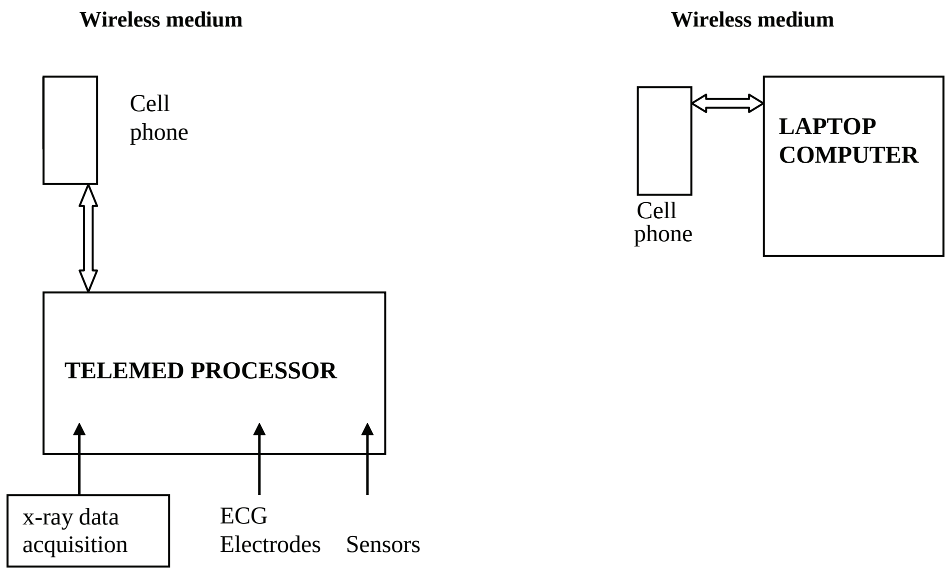
INTRODUCTION

Mobile telemedicine is a new research area that takes advantage of the cellular technology to provide highly flexible medical services that are not possible with standard telephony. Conventional telemedicine systems using public switched telephone network (PSTN) land lines are already available to enable a doctor to monitor a patient remotely for home care or emergency applications as mentioned by the authors Rezazadeh and Evans (1990), Patel and Babbs (1992), Clarke, Fragos, Jones, and Lioupis (2001), Coyle, Brown and Boydell (1995). Systems using satellite communications are dealt by the authors Pierucci and Del Re (2000); Murakami, Shimizu, & Yamamoto (1994). Teleradiology is the term used when any digital image is moved outside the local environment. Teleradiology involves the transmission of digitized radiological signals to a remote site for medical diagnosis. The transmission site is a hospital. The remote site is the location of the specialized doctor. Since the doctor may not be available at all times in a hospital, it becomes difficult to attend to trauma patients. With teleradiology, the territory of a doctor is expanded by being able to receive the images over his cellular phone. The doctor will make an initial diagnosis and communicate to the hospital. For further diagnosis, the doctor will attend to the patient at his own convenience. The global system for mobile communications (GSM) has been applied along with signal processing for transmission of ECG signals by Woodward, Istepanian, and Richards (2001). In GSM, the error-coding scheme used is Convolutional coding. The limited bandwidth of GSM (9.6kbps) restricts potential mobile telemedical services. Turbo coding is the error-coding scheme used nowadays in wireless transmission since it almost reaches Shannon limit. In this article, the images are coded using Turbo encoder and decoded using the Turbo decoder. The noise signal is simulated and added to the image after encoding the data. While decoding, the algorithm efficiently removes the noise and the image quality is analysed. These images will be made available in the cellular phones. The doctor can diagnose based on that image or can download the image into a computer to do a detailed analysis. The QoS (Quality of Service) parameters of Turbo coding are modified to make the transmission of images more efficient with better quality. The paper is organized as follows. In section 1, the existing Teleradiology system is discussed and the proposed system is explained with a diagram. The second section explains the operation of the Error coding scheme .The optimal design of the QoS parameters used in Turbo coding for this application is discussed in detail in the third section. The fourth section gives the simulation results. Conclusion and future work is given in the fifth section.

TELERADIOLOGY SYSTEM

Teleradiology system is a system used for sending raw data from a complete patient study to a remote location for a radiologist to make a final decision. A teleradiology system consists of an image acquisition system, an image server to compress, and a telecommunication network to transmit the images. The network could be a Local Area Network (LAN) or a broadband internet connection. LAN cannot be applied for far distances. Broadband internet connections are not easily accessible in all areas. In this paper, we propose a system by means of which the images could be sent over the easily available cellular network, which is very economical. The system consists of two segments, one for sending, and

Figure 1. Teleradiology system



the other for receiving. The functional block diagram is shown in Figure 1. The function of the processor is to store, compress, frame, and multiplex the data. The input signals are analog data from electrocardiogram (ECG) acquired from conventional chest electrodes, x-ray data from x-ray data acquisition systems, or digital data from other types of physiological sensors. The processor is a computer that can have any amount of memory. Since the signals are from many channels, multiplexing is done to send the desired signal based on a priority scheme in a particular time instant. Since the data size is large, it is necessary to compress the data before transmission. To have a secure transmission, encryption algorithms are applied. The processed signals are then fed to a mobile telephone via a standard infrared (IrDA) port or a bluetooth enabled device. Signals from the telephone are transmitted at the standard Universal Mobile Telecommunications system (UMTS) rate to the base station. The rate of transmission supported by UMTS is 384 kbps and it is targeted at reaching 2 Mbps. The signals would be received by another mobile cellular phone carried by a medical specialist equipped with a laptop computer, without being at a hospital. The receiving segment demodulates the data and then decodes the data to get back the original data that was transmitted. There are lot of systems to view picture efficiently. Some researchers have begun to develop PDA (Personal Digital Assistant) applications that drive larger picture archiving and communication system (PACS) displays or act as digital image capture stations. Laptop-based systems have been used by Reponen, Illkko and Jyrkinen (1998), but are not as portable as a PDA-based system and have much longer start-up times. Laptops and touch-screen input–based Tablet PCs (personal computers) are also limited by their relatively short battery life, whereas PDAs can function for days between charges. If image quality were sufficient, a handheld teleradiology system would be very useful for practicing radiologists, especially for those who provide emergency and other time-critical services.

The communication between the two segments is via the wireless media. Since the data is more prone to error, it is necessary to implement a suitable error correction scheme to reduce the noise in transmission. In this paper an error correction scheme has been applied to a chest x-ray and ECG data. The simulation model is shown in Figure 2. The input image gets corrupted when transmitted. Hence, the data is encoded by means of Turbo coder. The noise signal is simulated for any Signal to noise ratio level (SNR) and is added to the encoded signal. The decoder algorithm is applied to get back the original image without much loss in the information.

Having the ability to view radiologic images any time, any place, whether in a hospital or off-site could reduce time to diagnosis and increase the efficiency of the daily workflow, possibly reducing costs and improving patient care.

ERROR-CODING SCHEME

Telemedicine takes advantage of telecommunications to offer medical care at any time and at any place. The 2G mobile technology, GSM (global system for mobile communications), has been used by Woodward, Istepanian, and Richards (2001), Kyriacou, Pavlopoulos, Koutsouris, Andreou, Pattichis, and Schizas, (Oct.2001), and Elena, Quero, Tarrida, and Franquelo (2002), which has enabled remote diagnosis in mobile environments and to areas, which cannot be reached by fixed networks. GSM is used to carry circuit switched data at a maximum bit rate of 9.6 kbps. For multimedia applications where images have to be transmitted, GSM cannot be applied due to low bit rate. Universal Mobile Telecommunications system (UMTS) overcomes these limitations as it can support up to 2 Mbps with different quality of service (QoS). A complete performance analysis of transmission of medical data over UMTS channel has been done by Jose, Hernandez, Maria, Lafuente,Valdovinos, and Fernandez (2005). In this paper, we adopt the UMTS standards and make the error-coding scheme adapt to the desired QoS. Turbo codes emerged in 1993 and been implemented by Berrou, Glavieux and Thitimajshima (1993), Berrou, and Glavieux (1996), Benedetto and Montorsi (1996a), Branka and Jinhong (2000) and have since become a popular error-coding scheme in the area of wireless transmission. The important characteristics of turbo codes are the small BER achieved even at low

Figure 2. Simulation model

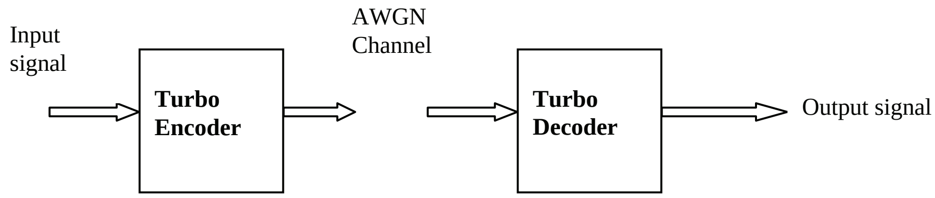
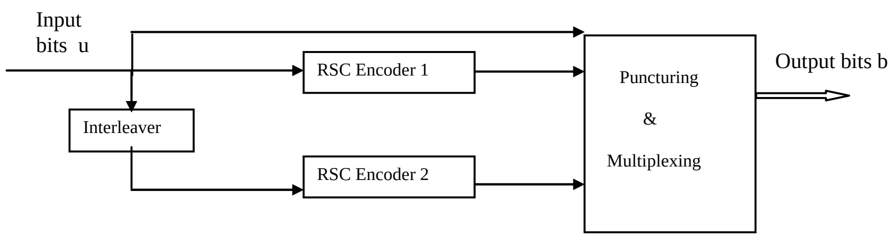


Figure 3. Structure of turbo encoder



Signal to noise ratio (SNR) and the flattening of the error rate curve (i.e., the error floor at moderate and high values of SNR). The performance of turbo codes is due to the randomness created by the interleaver and the iterative decoding. Turbo codes consist of two binary recursive systematic convolutional (RSC) encoders with small constraint lengths usually set between 3 and 5 which are concatenated in a parallel fashion by using a Turbo code interleaver and a puncturing and multiplexing device as shown in Figure 3. The basic role of an interleaver is to construct a long block code from small memory convolutional codes, as long codes can approach the Shannon capacity limit. There are two main algorithms to decode the data. One of the algorithms is called Soft Output Viterbi Algorithm (SOVA) and the other one is Maximum A Posteriori Algorithm (MAP) implemented by Benedetto, Divsalar, Montorsi, and Pollara (1996b).

OPTIMAL CHOICE OF PARAMETERS

In mobile telemedicine, the transmission is more prone to error due to the high speeds of the mobile doctor. Users moving at high speeds have worse channel conditions and thus require higher signal to noise ratios for the same QoS constraints. Turbo codes, the latest error-coding scheme is adaptable to the needs of different QoS. The different parameters of turbo codes are designed to achieve the required QoS.

Choice of Constituent Code in Encoder

The code design criteria depends on the region of SNR's where the code is used. A turbo code performing well in the region of high SNR's may not perform well in the region of low SNR's, or vice versa. Hence, the design of constituent codes is made, based on the range of SNR's where the code is used. One of the parameter involved in the encoder design is the effective free distance defined as:

$$d_{free,eff} = 2 + 2z_{\min} \quad (1)$$

where $z_{\min}$ is the lowest weight of the parity check sequence in error paths of RSC component encoders generated by an information sequence with weight 2. The binary rate r=1/2 RSC code is shown in Figure 4. The generator matrix of this coder is represented by:

$$G(D) = \left[1 \frac{g_2(D)}{g_1(D)} \right] \quad (2)$$

where $g_1(D)$ and $g_2(D)$ are the feedback and feedforward polynomial. The design criteria for constructing good turbo codes is to maximize the effective free distance which has been done by Benedetto and Montorsi (1996b). Based on this design objective, turbo encoder has been designed by choosing a primitive feedback polynomial to be used in the RSC component encoder. Based on the design, the generator matrix is [111,101] where g_1(D) is 111 and g_2(D) is 101. The input is d_k and the outputs are u_k and v_k which is the parity bit. The states of the encoder are represented by a_k, a_{k-1}, and a_{k-2}. Since the encoder is a recursive systematic coder, ak is recursively calculated as:

$$a_k = d_k + \sum_{i=1}^{K-1} g_{1i} a_{k-i} \quad (3)$$

The output pair of this RSC code is v_k, d_k where v_k is given by:

$$v_k = \sum_{i=0}^{K-1} g_{2i} a_{k-i} \quad (4)$$

The interleaver length and the memory of component codes have to be trade off according to the system requirements. In fact, increasing the

Figure 4. Structure of RSC encoder

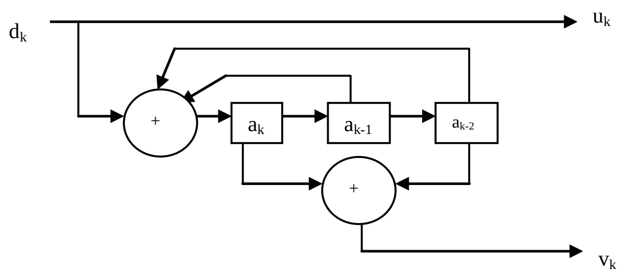
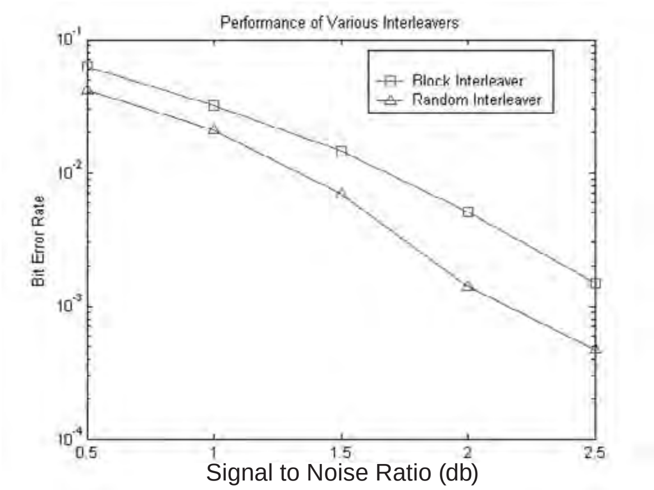


Figure 5. Performance analysis for different types of interleavers



interleaver size increases the decoding delay at almost no expense in complexity, whereas increasing the memory increases the complexity with only slightly increasing the decoding delay. Hence, the memory size chosen for the encoder is 2.

Choice of Interleaver

The proper design of interleaver plays a major role in the performance of turbo codes. The design criterion at high SNR or low BER is to maximize the code effective free distance. At low SNRs or high BER, the performance of the code is determined by its distance spectrum. At low SNRs the interleaver size is the only important factor as it can increase the interleaver gain and enhance the code performance. At high SNRs both the structure and size play an important role in maximizing the code effective free distance, thus enhancing the performance of the turbo code. Random and block interleavers are the most commonly used interleavers. Block interleaver formats the input sequence in a matrix of m rows and n columns, such that $N=m\times n$. The input sequence is written into the matrix row-wise and read out column-wise. In the random interleaver a block of N input bits is read into the interleaver and read out randomly. Based on simulations shown in Figure 5, random interleaver is found to be the best as the error rate is less compared to the block interleaver.

At low SNR's, the turbo code performance is dominated by the interleaver size. Turbo codes have been shown to exhibit remarkable performance when the frame size is large. The size of the interleaver is same as the size of frame. However, the latency in a turbo-coded system is directly proportional to the frame size. Thus, there is a direct tradeoff between latency and bit error rate (BER) based on the choice of interleaver size. Thus, the frame size has to be chosen based on the required QoS.

Latency is given by:

$$t_d = \frac{K_f}{R_b} N_i \qquad (5)$$

where R_b is the bit rate, K_f is the frame size and N_i is the number of decoding stages. From the formula we are able to conclude that latency is directly proportional to the frame size. Larger the frame size, larger is the latency or the delay. For applications where delay is a major constraint, the frame size should be reduced to get less delay or fast transmission.

The bit rate considered is 384 kbps. The number of decoding stages is 5. We are able to see that for larger frame size of 1000 bits, the latency is 13ms and for smaller frame size of 500 bits, the latency is 6.5ms. But on the performance analysis given in Figure 6, we are able to see that the error rate curve is very high for smaller size frame. To achieve the desired bit error rate of 10^{-3}, there is a gain in .3db for a frame size of 1000 compared to the frame size of 500. Hence we optimally choose 1000 as the frame size or the interleaver size.

Design of Decoder

There are two main algorithms used in the decoding of turbo codes. One is the MAP algorithm and the other is SOVA. Both the algorithms can be used as an iterative algorithm. SOVA is less complex compared to MAP but the performance of MAP is better compared to SOVA in terms of the bit error rate. In SOVA, a branch metric is assigned to each state transition for every state and for each decoding cycle. The path with the best path metric is selected. The path metrics are accumulated for the different time instants. Based on the history of the best path, the values of the data are taken from a memory available in the decoder.

In the MAP decoder, for each transmitted symbol, an estimate is made on the received bit by means of a posteriori probability on the basis of the received sequence r. The decoder computes a log-likelihood ratio given by

$$\Lambda\left(c_t\right) = \log \frac{P_r\left\{c_t = 1 \middle| r\right\}}{P_r\left\{c_t = 0 \middle| r\right\}} \qquad (6)$$

for $1 \le t \le n$ where n is the received sequence length and compares this value to a zero threshold to determine the hard estimate The estimate of c_t is taken as 0 if the log-likelihood ratio is lesser than 0, otherwise c_t is taken as 1.

Based on the performance analysis of both the decoders, MAP algorithm proves to be better as shown in Figure 7. In this article, iterative MAP algorithm is applied as performance is better when the number of iterations are increased. The iterative MAP decoder is shown in Figure 8. It consists of two component decoders serially concatenated via an interleaver. The first MAP decoder takes as input the received information sequence r_0 and the received parity sequence generated by the first encoder r_1. The decoder then produces a soft output, which is interleaved and used to produce an improved estimate of the a priori probabilities of the information sequence for the second decoder. The other two inputs to the second MAP decoder are the interleaved received information sequence r_3 and the received parity sequence produced by the second encoder r_2. The second MAP decoder also produces a soft output, which is used to improve the estimate of the apriori probabilities for the information sequence at the input of the first MAP decoder. The feedback loop is a distinguishing feature of this decoder. After certain number of iterations, the soft outputs of both MAP decoders stop to produce further performance improvements. Then the last stage of decoding makes a hard decision after deinterleaving.

The number of iterations lead to more latency but increased performance.

SIMULATION

The images are divided into frames and are sent to the encoder. The coded sequence is the information and the parity sequences. This coded data is added with Gaussian noise. Adding noise involves generating Gaussian random numbers, scaling the numbers according to the desired energy per symbol to noise density ratio E_s/N_0, and adding the scaled Gaussian random numbers to the coded sequence. The energy per symbol is given by

$$E_s/N_0 = E_b/N_0 + 10\log_{10}(k) \qquad (7)$$

where k is the code rate and E_b/N_0 is the signal to noise ratio in db. The variance of the noise which is added is given by σ^2 which is equivalent to $N_0/2$ and

$$\sigma = \sqrt{\frac{1}{2(E_s/N_0)}} \qquad (8)$$

Gaussian random numbers of the above variance which relates the SNR is generated and added to the coded sequence. The data, which reaches the decoder is the corrupted data (i.e., the coded sequence added with the Gaussian variable). The aim of the decoder is to get back the original information. But practically, it is not possible to achieve transmission without any loss. The MAP decoder is run for five iterations inspite of the extra latency since the image quality is better after five iterations.

The following parameters have been chosen based on the analysis done:

Figure 6. Performance analysis for different frame sizes

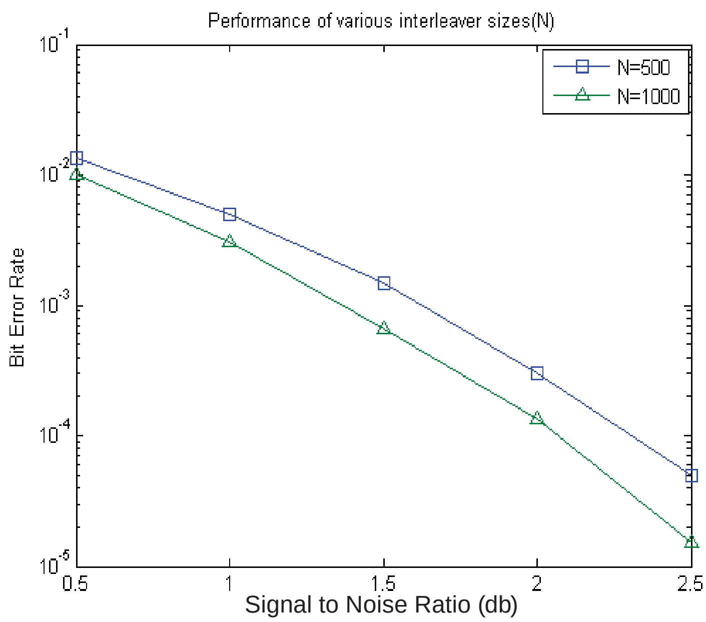
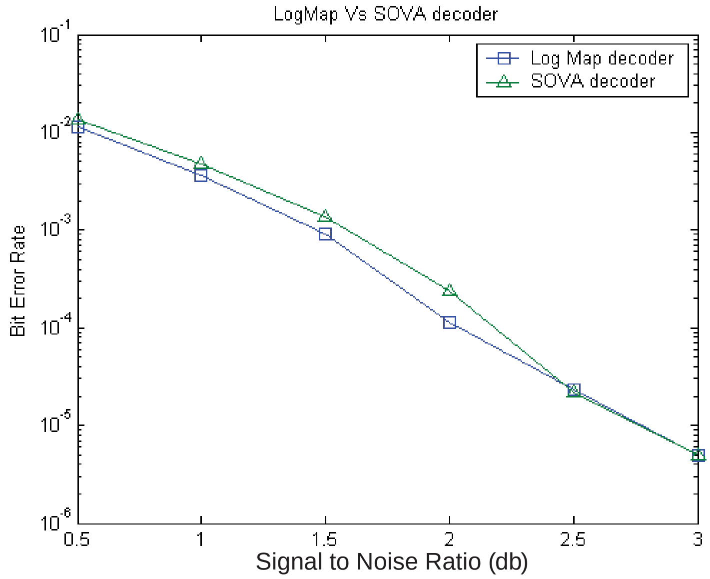


Figure 7. Performance analysis of Turbo codes with different types of decoders



- Frame size: 1000 bits.
- Interleaver choice: Random interleaver.
- Generator Polynomial: g_1= 111 and g_2=101.
- Decoding algorithm: MAP algorithm.
- Noise level added: 1.0db AWGN noise.

A chest x-ray image and an ECG has been simulated for 1.0db noise level. The images after decoding are shown in Figure 9 and Figure 10. The error metrics used to analyse the quality of image is the Peak Signal to Noise Ratio (PSNR) and Mean Squared Error (MSE). The MSE is the cumulative squared error between the reconstructed and the original image, whereas PSNR is a measure of the peak error. The mathematical formulae are:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left[\text{Image}_{original}(i) - \text{Image}_{reconstructed}(i)\right]^2 \quad (9)$$

where N is the total number of pixels in the original image.

$$PSNR = 20\log_{10}\left[\frac{255}{\sqrt{MSE}}\right] \quad (10)$$

A lower value for MSE means less error. A higher value of PSNR is desirable because it means that the ratio of Signal to Noise is higher. Here the signal is the original image. The PSNR values are shown in Figure 11 and Figure 12. The decoding algorithm is run for five iterations. In each iteration, the PSNR value increases. For images with noise, a PSNR value of 30 is acceptable. For ECG, the PSNR is higher than 30 and for chest x-ray, it will approach 30 with another iteration. The images are sent to a bluetooth enabled cellular phone to visualize the image quality. The photographs of the images are shown Figure 13 and Figure 14.

Figure 8. Structure of iterative MAP decoder

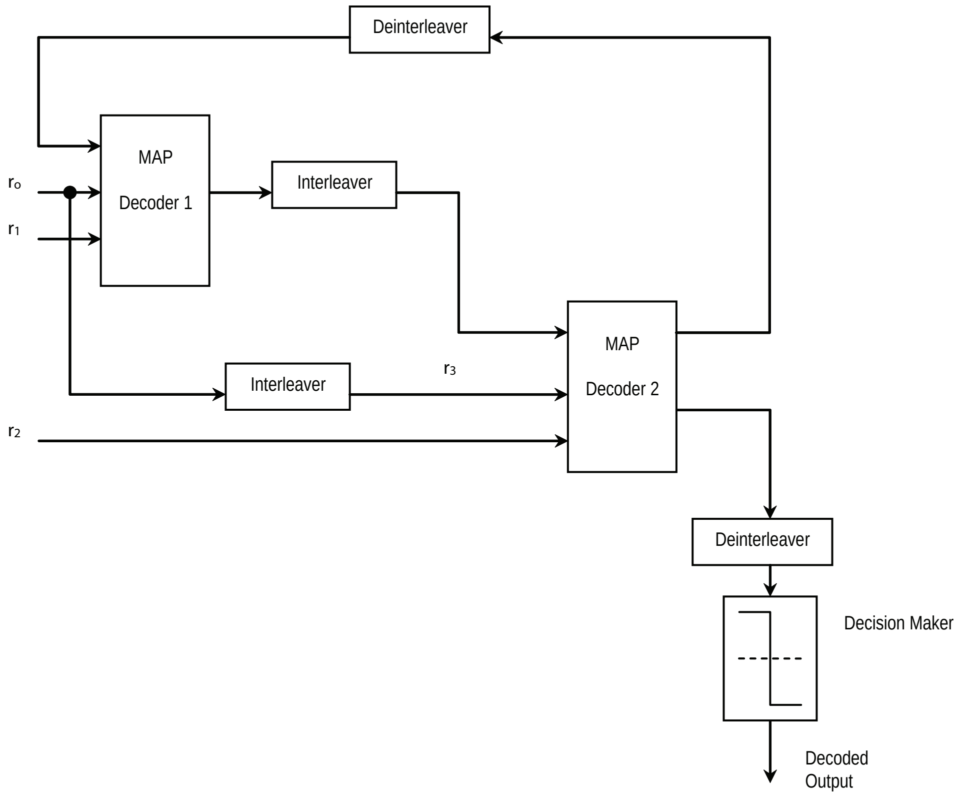


Figure 9(a) Original image, Figure 9(b) , 9(c) , 9(d) , 9(e) and 9(f) images after successive iterations

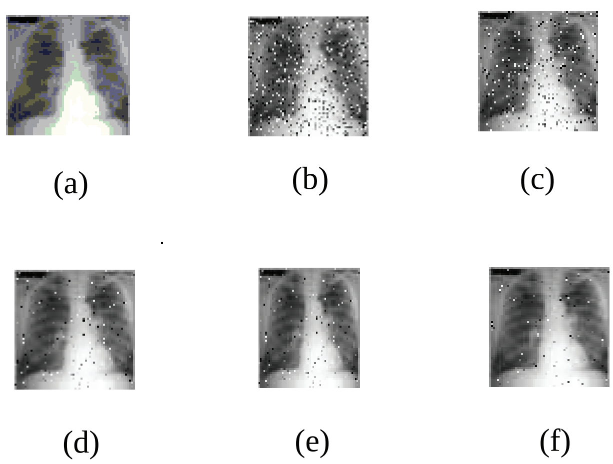


Figure 10(a) Original image Figure 10(b), 10(c) , 10(d) , 10(e) and 10(f) Images after successive iterations

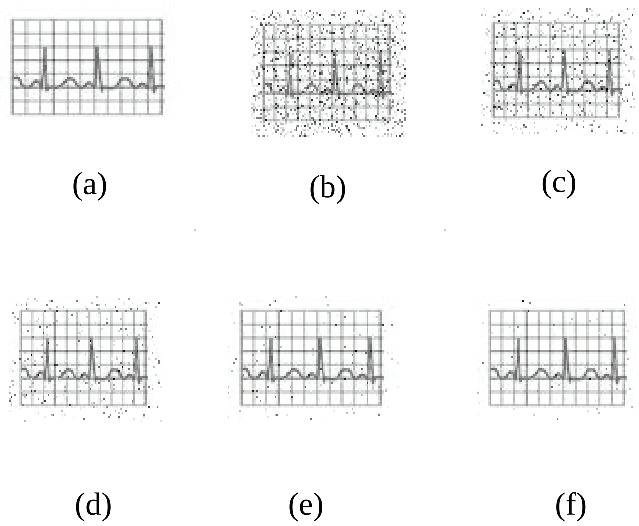


Figure 11. PSNR values for chest X-ray after decoding

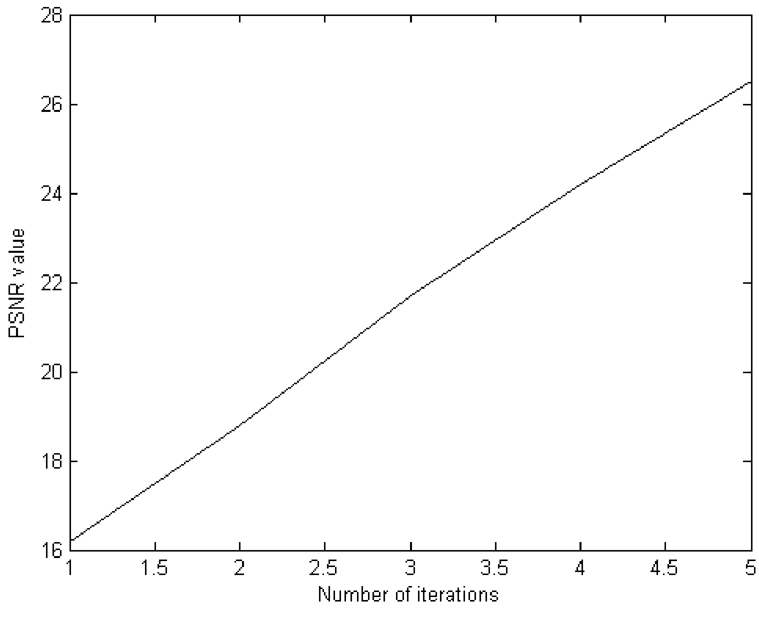


Figure 13. X-ray Image after being captured in cell phone

Figure 14. ECG Image after being captured in cell phone

CONCLUSION AND FUTURE WORK

Teleradiology is an upcoming technology in the field of telemedicine. Most of teleradiology is done using dedicated links, which is costly. Cellular networks are the cheapest and easily available wireless networks. It has also reached all parts of the world. The UMTS networks offer high rate of transmission and hence this scheme is reliable and cheaper. The image quality is an important factor in wireless links. In this paper, we have implemented an error coding technique with optimal parameters to offer the necessary Quality of Service. Future work can be to incorporate streaming technology for EEG and transmit EEG signals.

The complete teleradiology system can be implemented in real time and the performance analysis can be done. Incorporation of compression algorithms and encryption techniques for secure transmission could be another interesting area of future work

REFERENCES

Benedetto, S., & Montorsi, G. (1996a). Unveiling turbo codes: Some results on parallel concatenated coding schemes. IEEE Transaction Information Theory, 42, 409-428.

Benedetto, S., & Montorsi, G. (1996b). Design of parallel concatenated convolutional codes. IEEE Transaction on Communication, 44(5).

Benedetto, S., Divsalar, D., Montorsi, G., & Pollara, F.(1996b). A soft-input soft-output Maximum A Posteriori (MAP) module to decode parallel and serial concatenated codes. TDA Progress Report 42-100.

Berrou, C., Glavieux, & Thitimajshima, P. (1993). Near Shannon limit error-correcting coding and decoding: Turbo-codes. Proceedings of the IEEE International Conference Communication (pp. 1064-1070).

Berrou, C., & Glavieux, A. (1996). Near optimum error correcting coding and decoding: Turbo codes. IEEE Transaction Communications, 44(10), 1261-1271.

Branka, V., & Jinhong, Y.. (2000). Turbo codes--Principles and Applications. Boston: Kluwer Academic Publishers.

Clarke, M., Fragos, A., Jones, R., & Lioupis, D.(2001). Optimum delivery of telemedicine over low bandwidth satellite links. Proceedings of the IEEE Engineering Medicine and Biology Society 2001 (Vol. 4, pp. 3606-3609).

Coyle, G., Brown, L., & Boydell, L. (1995). Home telecare for the elderly. Journal of Telemedicine Telecare, 1, 183-185.

Divsalar, D., & McEliece, R. J. (1996, February). The effective free distance of turbo codes. Electronic Letter, 32(5), 445-446.

Elena, M. Quero, J., Tarrida, & Franquelo.(2002). Design of a mobile telecardiology system using GPRS/GSM technology. Proceedings of the IEEE Engineering Medicine and Biology Society 2002 (Vol. 3, pp. 1859-1860).

Jose, R., Hernandez, A., Maria,C., Lafuente, J., Valdovinos, A., & Fernandez, J.(2005). Performance analysis of multiplexed medical data transmission for mobile emergency care over the UMTS channel. IEEE Transactions on Information Technology in Biomedicine, 9(1).

Kyriacou, E., Pavlopoulos, S., Koutsouris, D., Andreou, A., Pattichis, C., & Schizas, C., (Oct.2001). Multipurpose healthcare telemedicine system. Proceedings of the IEEE Engineering Medicine and Biology Society 2001 (Vol. 4, pp. 3544-3547).

Murakami, N., Shimizu, K., & Yamamoto, K. (1994). Telemedicine using mobile satellite communications. IEEE Trans. Biomed. Eng., 41, 488-497.

Patel, U., & Babbs, C. (1992). A computer based, automated telephonic system to monitor patient progress in home setting. J. Med. Syst., 16, 101-112.

Pierucci, L., & Del Re, E. (2000). An interactive multimedia satellite telemedicine service. IEEE Multimedia, 7, 76-83.

Reponen, J., Illkko, E., & Jyrkinen, L.(1998). Digital wireless radiology consultations with a portable computer. J Telemed Telecare, 4, 201-205.

Rezazadeh, M., & Evans, N. (1990). Multichannel physiological monitor plus simultaneous full-duplex speech channel using a dial-up telephone line. IEEE Trans. Biomed. Eng., 37, 428-432.

Woodward, B., Istepanian, R., & Richards, C. (2001). Design of a telemedicine system using a mobile telephone. IEEE Transactions on Information Technology in Biomedicine, 5(1).

Woodward, B., Istepanian, R., & Richards, C. (2001). Advances in telemedicine using mobile communications. Proceedings of the IEEE Engineering Medicine and Biology Society 2001 (Vol. 4, pp. 3556-3558).

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 1, Issue 4, edited by J. Tan, pp. 15-28, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 4.5
Methods and Applications for Segmenting 3D Medical Image Data

Hong Shen
Siemens Corporate Research, USA

ABSTRACT

In this chapter, we will give an intuitive introduction to the general problem of 3D medical image segmentation. We will give an overview of the popular and relevant methods that may be applicable, with a discussion about their advantages and limits. Specifically, we will discuss the issue of incorporating prior knowledge into the segmentation of anatomic structures and describe in detail the concept and issues of knowledge-based segmentation. Typical sample applications will accompany the discussions throughout this chapter. We hope this will help an application developer to improve insights in the understanding and application of various computer vision approaches to solve real-world problems of medical image segmentation.

INTRODUCTION

The advances in medical imaging equipment have brought efficiency and high capability to the screening, diagnosis and surgery of various diseases. The 3D imaging modalities, such as multi-slice computer tomography (CT), magnetic resonance imaging (MRI) and ultrasound scanners, produce large amounts of digital data that are difficult and tedious to interpret merely by physicians. Computer aided diagnosis (CAD) systems will therefore play a critical role, especially in the visualization, segmentation, detection, registration and reporting of medical pathologies. Among these functions, the segmentation of objects, mainly anatomies and pathologies from large 3D volume data, is more fundamental, since the results often become the basis of all other quantitative analysis tasks.

The segmentation of medical data poses a challenging problem. One difficulty lies in the large volume of the data involved and the on-time requirement of medical applications. The time constraints vary among applications, ranging from several tens of milliseconds for online surgical monitoring, to seconds for interactive volumetric measures, to minutes or hours for off-line processing on a PACS server. Depending on the application, this puts a limit on the types of methods that may be used. Another major hurdle is the high variation of image properties in the data, making it hard to construct a general model. The variations come from several aspects. First, the complexity of various anatomies maps to the large variation of their images in the medical data. Second, the age, gender, pose and other conditions of the patient lead to high inter-patient variability. Last, but not the least, are the almost infinite variations in an anatomy due to pathology or in the pathological structures. On the other hand, medical applications usually have a strong requirement of robustness over all variations. Beside the above challenges, system issues exist for the major modalities, such as noise, partial volume effects, non-isotropic voxel, variation in scanning protocols, and so forth. These all lead to more difficulties for the medical segmentation problem.

Knowledge-Based Segmentation

Medical image segmentation has the advantage of knowing beforehand what is contained in the image. We also know about the range of size, shape, and so forth, which is extracted from expert statements. In other fields of computer vision, such as satellite image analysis, the task of segmentation sometimes contains a recognition step. Bottom-up strategy is usually used, which starts with the low-level detection of the primitives that form the object boundaries, followed by merging. One sophisticated development is Perceptual Organization (Sarkar & Boyer, 1994; Guy & Medioni, 1996; Mohan & Nevatia, 1989), which attempts to organize detected primitives into structures. It is regarded as the "middle ground" between low-level and high-level processing. In 3D medical data, grouping is much more difficult due to the complexity of medical shapes. Because of this, top-down strategies prevail in 3D medical image analysis.

Knowledge-based segmentation makes strong assumptions about the content of the image. We use prior knowledge to find and tailor a general segmentation method to the specific application. Global priors are applied when the local information is incomplete or of low quality. It is a top-down strategy that starts with knowledge or models about high-level object features and attentively searches for their image counterparts.

The past several decades witnessed dramatic advances in the fields of computer vision and image analysis, from which the area of medical image analysis is derived. Various methods and frameworks for segmentation have been proposed, and many are driven by the needs of medical image analysis. These provide valuable theoretical thoughts as the basis of knowledge-based segmentation, but only at a high level. Typically, such a method is shown to be generic as it works on a number of cases from various applications with reasonable successes. This is quite different from the requirement of medical image segmentation in the real world, which depend heavily on the specific application—the workflow of the medical procedure, the anatomy and pathology of interest, the performance and accuracy requirements and the user inputs. Given all the priors and conditions of a medical application, we need to design the algorithm that will be the best compromise between accuracy, speed, robustness and user inputs. The resulting system will be specific to the given application; not only the algorithm, but also the parameter settings. A medical image analysis algorithm will be put to tests on thousands of data sets before it can be made into a clinical product. Hence, even if it is application specific, such an algorithm is general in that it has to cover all possible variations of the application.

A developer in this field not only needs to master the various methods in computer vision, but also understand the situations each of them may be applied to, as well as their limitations. In the next sections, we will focus on real world issues of medical image segmentation. We will first give an overview of the relevant low-level methods with examples. We will then discuss the popular frameworks for incorporating global priors and their limitations. We will present the principles of a novel strategy that was developed recently (Shen, Shi, & Peng, 2005) for an sample application. Finally, we will conclude the chapter with an outlook on the future trends in this field.

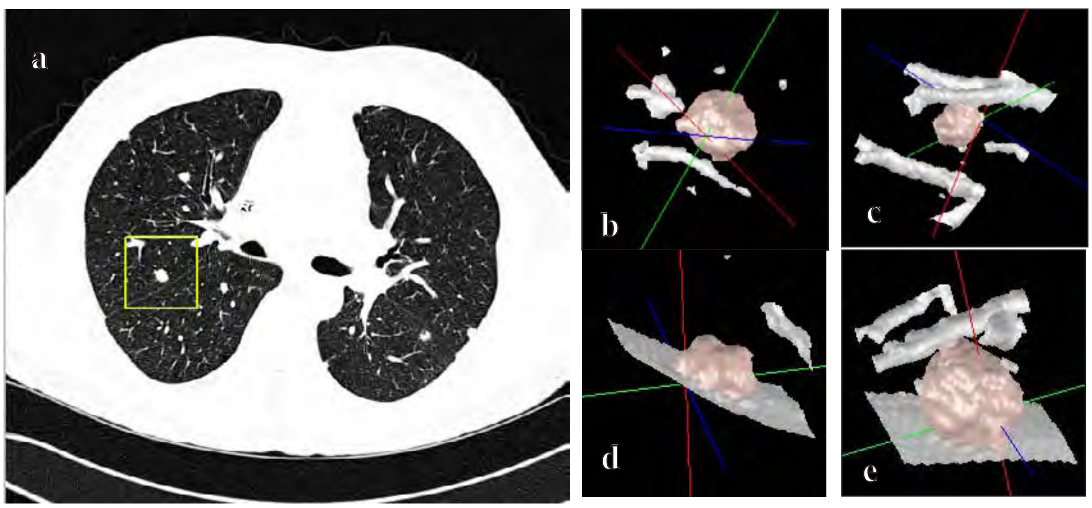
APPLICATIONS USING LOW-LEVEL SEGMENTATION METHODS

In this section we will give 3D medical application examples that use low-level segmentation methods. We define low-level methods as those that do not incorporate global shape priors of the object to be segmented, but mostly rely on local image properties. Many surveys and textbooks (Sonka, Hlavac, & Boyle, 1998) give detailed descriptions of all the segmentation methods, and this short chapter will not serve as a duplication of these efforts. Instead, we will give examples to show where each category of low-level method is adequate, and also review the recently developed methods.

Segmentation Using Regional and Morphological Methods

Region-based methods group together voxels that have some common properties. They are applicable when there is high contrast between object and background and intra-object variation is minimal. As shown in Figure 1, in a chest CT data, the lung area maps to dark regions, while the chest wall, the vessels, the airways and nodules map to bright regions of the same intensity range. A typical application is the segmentation of a lung nodule given a click point in the nodule. Since there is high contrast between object and background, a simple 3D region grow (Sonka et al., 1998; Adams & Bischof, 1984) is the best method to obtain the foreground voxels that are above a clear-cut threshold. The challenges are as follows: First, the nodule can be connected to a vessel and/or the chest wall, which has the same intensity range as the nodule. Second, it is required that the segmentation result be com-

Figure 1. Lung nodule segmentation from multi-slice chest CT data



pletely independent of the click point as long as it is within the nodule. Last, but not the least, is the high expectation from the user of success rate on a large number of cases due to the fact this looks like an easy problem.

These challenges are met by incorporating prior knowledge about the application. First, the click point is guaranteed to be in the nodule. Second, the nodule size range is known. Third, the nodule is known to be compact, while a vessel is assumed to be of elongated shape, and the chest wall is mostly a smooth surface. Given the click point and size range, we define a cubic volume of interest (VOI) centered at the click point, and hence isolate out the object and its immediate neighbors. In this VOI, methods such as mathematical morphology are applied to consistently find the nodule center and separate the nodule from the connected structures.

Mathematical morphology utilizes local shape information (Sonka et al., 1998; Haralick, Sternberg, & Zhuang, 1987). Although grayscale operations are also defined, this type of method is mostly used in binary images that are generated from an earlier process such as region grow. As illustrated in Figure 1(c), the foreground object contains a nodule connected to a vessel. The basic operators such as dilation, erosion, opening and closing can be used for boundary modifications and to separate two connected objects. However, the more effective separation techniques are based on the important concept of geodesic distance. A geodesic distance of a point inside the foreground object is the shortest travel distance to the background by any path that is completely inside the object.

Using a geodesic distance map, the vessel voxels can be removed from the nodule since their distance values are relatively small. Again, the knowledge about relative sizes between the nodule and the vessel are used to determine the distance thresholds. We could use more sophisticated watershed algorithms (Vincent & Soille, 1991) to find and separating distance basins, if the over-segmentation issue can be overcome (Meyer & Beucher, 1990).

The region-based methods are based on the similarity or connectivity of voxels, therefore they can be extended to more general situations by redefinition of the similarity or connectivity. Recently, the concept of fuzzy connectivity has been introduced (Udupa, Saha, & Lotufo, 2002; Herman & Carvalho, 2001), which is defined between every possible pair of voxels. All objects are initialized with a reference point and then grow by competition to acquire voxels according to the strengths of fuzzy connectivity.

Overall, region-based segmentation is robust, simple to implement and fast when the object is small. In some cases, region merging and splitting is necessary to the grown region according to other criterions (Sonka et al., 1998; Chang & Li, 1994). However, the more heuristics that have to be applied in order to get a decent result, the less reliable the algorithm will become. The bottom line is that region-based, low-level methods should be applied to regions with relatively homogenous intensities.

Segmentation Based On Primitive Detection

Image primitives are local properties such as edges, ridges and corners. Instead of looking for properties of all voxels inside the region, primitive-based methods look for the region boundaries. There are many recent developments in primitive detection, especially edge detection, which is the basis of most segmentation schemes. The detected primitives are usually fragmented, and the challenging problem of linking edges and lines into boundaries was the focus of many works in the early and middle 90s (Basak, Chanda, & Majumder, 1994; Wu & Leahy, 1992).

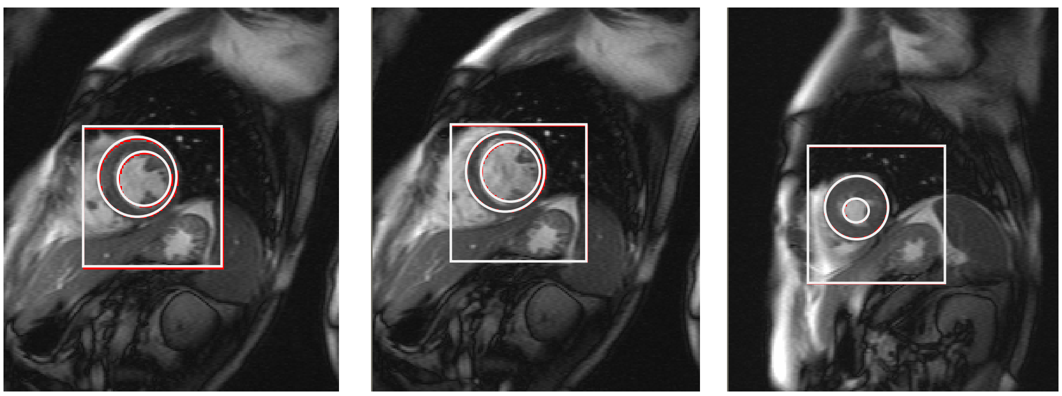
Basic edge detection is achieved by applying an edge operator (Sonka et al., 1998) followed by non-maxima suppression. A systematic treatment of edge detection was provided by Canny (1986), which is an optimal detector for step edges, and can be adapted to detect other features when different filters are selected. Recently, Brejl and Sonka (2000) invented a 3D edge detector that is particularly fit for anisotropic situations, which are typical in 3D medical data. Freeman and Edward (1991) proposed a method of systematically detecting all linear primitives, including step edges, ridges and valleys, as well as their orientations. They designed a basis of filters from which an arbitrarily-oriented filter can be generated to selectively detect the primitives of a specified orientation. This is particularly useful for high-level feature detection. For instance, if we know beforehand there should be a long ridge at a certain orientation and location, a ridge filter of that direction can be applied.

As a special type of primitive-based method, Hough Transform (Sonka et al., 1998; Shankar, Murthy, & Pal, 1998) utilizes a very powerful scheme for noisy medical images—voting. It is applicable when the object roughly fits a simple parametric model, such as a rectangle. Although in medical applications precisely regular shape is rare, some are roughly spherical or with a circular cross-section. For instance, shown in Figure 2 are the segmentation results of the left ventricle from cardiac MR data. The outer and inner boundaries approximate a circle, which has three free parameters in its parametric equation. Edge detection followed by non-maxima suppression is performed within the region of interest, and each edge pixel casts one vote for every circle equation it satisfies. The circle with the highest vote is taken as the object boundary. To achieve good results, the parameter space needs to be carefully divided into cells. Sometimes the vote can be weighted by edge strengths.

The power of this algorithm comes from its strong assumption. It works in spite of noise, weak boundaries and even occlusions, as long as the good edges can form a majority over the outliers. Certainly, the parametric model has to be simple before the algorithm becomes too complex. However, in some cases even when the object shape is not regular, we may still use it to find at least the correct centroid of the object.

The Hough Transform is a type of robust estimation method. Its principle of utlizing the majority of pixels can be widely applied. In the presence of abnormality, noise and high complexity, the robustness of a segmentation algorithm is usually determined by the number of voxels the decision is made upon.

Figure 2. Segmentation of the left ventricle wall from cardiac MRI data. The white circles mark the inner and outer boundaries of the left ventricle.



Statistical Segmentation Methods

Statistical methods treat the intensity of each voxel as an event, and model the local intensity variations by the estimation of probabilistic distributions. The goal is to use classification frameworks to determine whether a voxel belongs to a certain object. The set of voxels that are labeled as object points forms the segmented object. A simple scheme is to estimate the probability density functions (pdf) of different objects or background by computing the frequency of histograms from sample data. The estimated pdfs can then be applied to the test data. Each voxel can be classified using conventional classification methods, such as Bayesian classification or maximum likelihood decision. A more sophisticated algorithm would estimate the joint pdfs of the voxels in the neighborhood to capture the spatial local interaction properties between image labels. The most popular method is the Markov Random Field (MRF) framework, in which the pdfs are estimated with advanced algorithms such as EM algorithm. The segmentation is often obtained using a MAP estimator (Sun & Gu, 2004).

These type of methods can be applied when there is little knowledge about object shape and the intensities are nonhomogenous. A typical case, as shown in Figure 3, is the segmentation of ground glass nodules (GGN) from chest CT volume data, which has strong clinical significance due to the high malignance rate of GGNs (Zhang, Zhang, Novak, Naidich, & Moses, 2005). Due to large inter-nodule variations, an iterative approach is applied to gradually capture the local image properties. A rough intensity model is generated from training data and applied to the first round of classification. The resultant segmentation is then used to estimate a more specific distribution for the current nodule, which is in turn applied to the next round of classification. Several iterations combined with other heuristics give consistent segmentation results with respect to different user clickpoints. The same iterative strategy was used by Marroquin, Vemuri, Botello, and Calderon (2002) in the segmentation of brain MRIs. The establishment of pdfs is a learning process that requires large sample sets for each object, which may not be available for medical image segmentation. Further, large variations in

Figure 3. Segmentation of ground glass nodules (GGN) from chest CT data using Markov Random Field. Top row: Slices of volumes of interest (VOI). Bottom row: Slices of segmented GGN from the VOI above it.

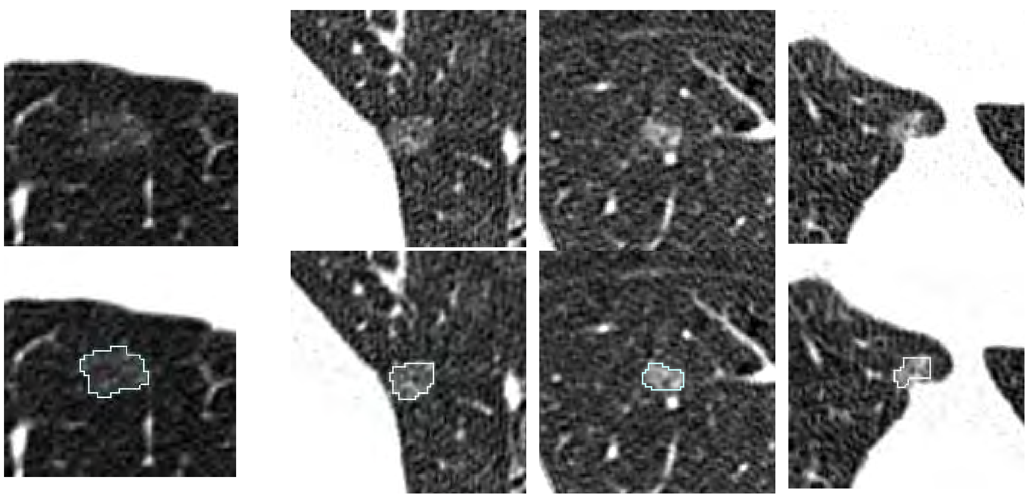


image properties often lead to overlapped pdfs. Usually, heuristics have to be combined to achieve a decent result.

SEGMENTATION INCORPORATING GLOBAL PRIORS

For more complex situations, low-level methods will not be successful. This is because the objects to be segmented in a medical application usually have high internal nonhomogeneity and strong internal edges. In these cases, a global prior is needed.

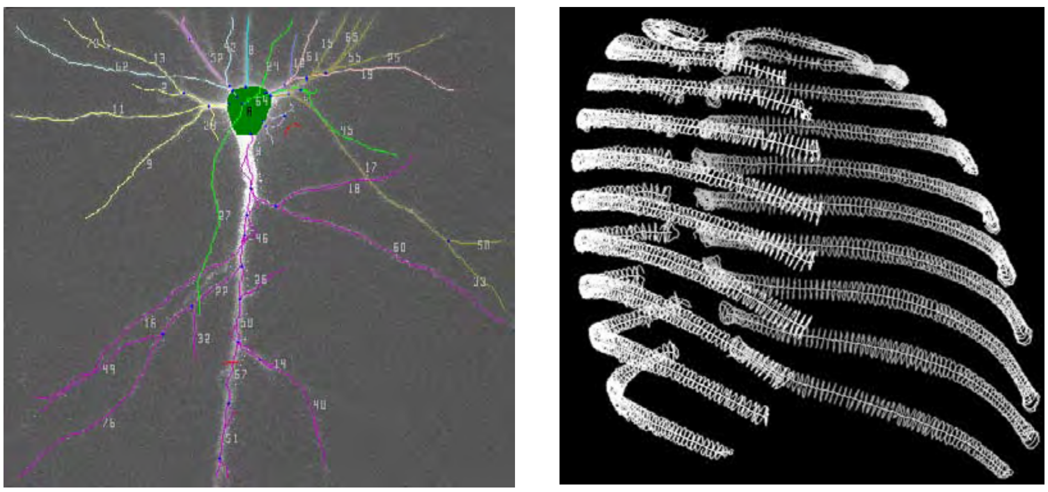
We first give as intuitive examples two special applications that utilize the global priors of object shapes—the tracing-based segmentation of tube-like structures. As shown in Figure 4, 3D tracing algorithms were developed for the segmentation of neurons from laser-scanning confocal image stacks (Al-Kofahi et al., 2002), and of rib structures from chest CT data (Shen, Liang, Shao, & Qing, 2004). A tracing algorithm uses the global prior knowledge that the object has piecewise linear shapes with smooth surfaces. First, seed points are automatically detected for all objects by techniques such as sparse grid search or 2D cross-section analysis. From each seed, tracing proceeds iteratively along each object, using edge information to determine the tracing directions. Carefully designed stopping criteria are applied to terminate a trace.

This approach has the advantage of obtaining object centerlines and boundaries simultaneously, and all tube structures are naturally separated. Another significant advantage is the low computational cost. This type of algorithm is exploratory. Only the voxels on or near the objects are processed, which is a small portion of the image data.

Deformable Models

The deformable models provide a framework that allows the incorporation of global priors. It is widely applied to medical data, and the potential is far from fully explored. Before global prior is introduced, a deformable model is a unified ap-

Figure 4. 3D exploratory tracing-based segmentation



(a) Segmentation result of neuron from confocal image stacks, the centerlines are overlaid on original image (2D projection) (Courtesy of Khalid Al-kofahi from Thomson Legal & Regulatory, Eagan, MN); (b) Segmentation result of rib structures from chest CT image shown in Figure 1. Shown are the rib contours and centerlines. They are naturally separated. A graphical surface reconstruction algorithm can be applied to obtain the rib surfaces.

proach to the capture of local image properties through front evolution, without the need of local primitive grouping. Further, by the constraints that a smooth shape is to be maintained during evolution, the frontier points move with interaction rather than independently. Figure 5a and 5b show the two example segmentations using the level set method, which is one popular deformable model.

Kass, Witkin, and Terzopoulos (1988) proposed one of the first deformable models—snakes—in which they introduced the energy minimizing mechanism. By using gradient decent and variational approach, the 2D contour is attracted to the local energy minimum. Through contour or surface evolution, a deformable model integrates three factors to jointly determine the segmentation. First, the smoothness requirement of the shape constrains the relative motion of frontier points. These points are only allowed to move such that a smooth front is maintained. Second, the surface actively adapts to the local image primitives, such as edges, ridges, etc. Finally, it is possible to impose global geometrical constraints on the shape. In the initial treatment of snakes, these three factors are integrated as energy terms in the partial differential equation (PDE).

Accordingly, three issues are to be addressed when designing a deformable model. First of all, we need to decide how to represent a shape or surface. The other two major issues are the design of local image forces and the incorporation of global priors to constrain the front evolution.

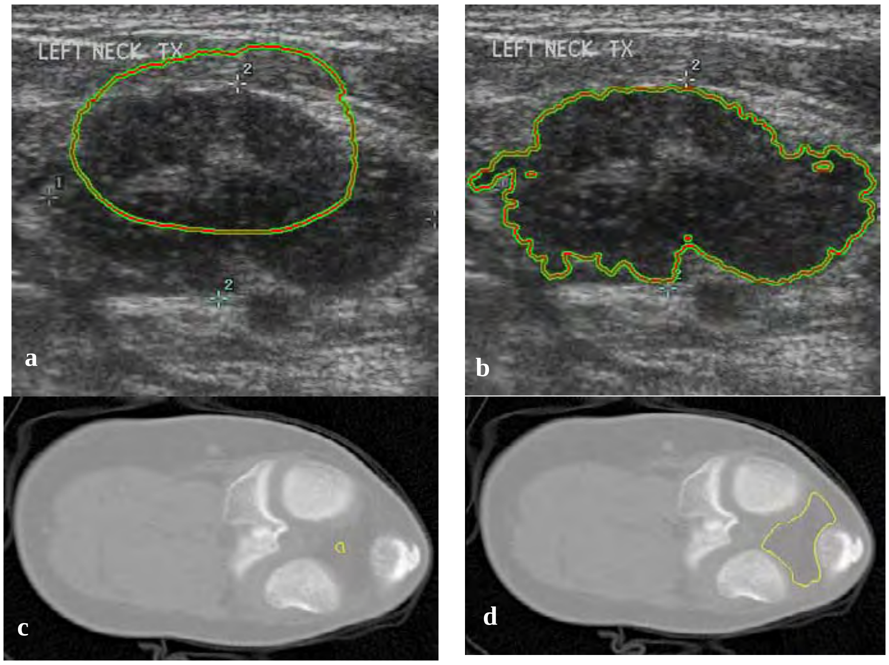
The issue of shape representation becomes more challenging in a 3D situation. Pentland and Sclaroff (1991) used an approach based on finite element method (FEM) to establish a parametric model of shapes. Lorenz and Krahnstover (1999) used a 3D mesh to represent the vertebra shape. Recently, Slabaugh and Unal (2005) developed a more general deformable mesh for surface evolving.

The Level Set Method for Shape Representation and Evolution

In the past decades, the level set methods (Malladi, Sethian, & Vemuri, 1995; Adalsteinsson & Sethian, 1995) received much attention. The surface is represented implicitly. Typically, for a close surface, a signed distance function is defined in a volume that contains that surface. The value of each voxel is the geodesic distance of that voxel to the closest point on the surface. Since this value on the surface is zero, the set composed of all surface points is named the zero level set of the signed distance function. The distance values of voxels in and outside the surface are set to be negative and positive, respectively. This implicit representation of the surface by the zero level set of the signed distance function provides great simplicity at the cost of an increase in dimension. The implementation is relatively simple, and provides a flexible platform to design various algorithms. Further, level set is transparent to dimension, which makes it much more desirable than some other methods that are hard to adapt to 3D.

Until recently, most level set researchers focused their efforts on the problem of surface evolving under local image properties, such as defining local image force or speeds, either edge-based or region-based (Chan & Vese, 2001; Chakraborty, Staib, & Duncan, 1996). A recent survey by Suri, Liu, and Singh (2002) gave a systematic summary of all the important efforts that address in depth the local surface evolvement. Without a prior shape model, however, a deformable model such as level set only reacts to local properties. For edge-based force, the surface will converge at local edges. For region-based force, the surface will converge when the average intensity of voxels in the object and background differs the most. A medical structure usually has complex local image properties, such as incomplete boundaries, noise, nonuniform internal intensities, etc. Therefore the most common problems for low-level segmentation methods also bother a deformable model—the

Figure 5. Surface evolving in the level set segmentation. The left column shows initialization, and the right column shows converged boundary.



(a) and (b) Larynx tumor in ultrasound data (c) and (d) Hoffa pad in CT knee data.
Courtesy of Gozde Unal and Greg Slabaugh from Siemens Corporate Research

leaking at weak boundaries or being caught by internal edges. The surface will truthfully conform to whatever local information it encounters. Even with the most advanced local constraining schemes, these problems persist.

Statistical Model as Global Prior Representation

Cootes, Taylor, Cooper, and Granham (1995) proposed the popular active shape models (ASM). Cootes, Edwards, and Taylor (2001) later proposed the active appearance models (AAM). Both of the two methods used statistical models to represent prior knowledge. The surface was represented by a point distribution model (PDM), which was a set of landmark points that were distributed on the object boundaries. For a given shape, the locations of all the landmarks were concatenated to form long vectors as the representation. The success of a PDM depends on correspondence of landmarks, which is challenging in 3D. Overall, the PDM models are difficult to implement in 3D. It is a very active field that attracts many contributions to solve these difficulties (Hill & Taylor, 1994; Walker, Cootes, & Taylor, 2002; Horkaew & Yang, 2003).

On the other hand, this was among the first systematic efforts to introduce statistical representation of global shape priors into a deformable model. Sample vectors were extracted from a set of registered training shapes and principal component analysis (PCA) is performed on the set of sample vectors. The dimension of the covariance matrix equals that of the sample vectors, but only the eigenvectors with the largest eigenvalues are taken to form the basis of the linear model space. The segmentation process first transforms the mean shape to the data, and then finds the correspondence of the mapped model landmarks by searching for strong edges in the contour's normal direction. Subsequently the shape vector computed from the correspondent points is projected onto the model basis to find the best matched model. This means a large number of residual components are discarded. The matched model is again matched onto the image data for next iteration. In a PDM, the representation of surface and the incorporation of global priors are combined and a shape is only allowed to vary within the model space. The local surface evolution is limited and for the only purpose of finding the matched model.

For the level set framework, recently Leventon, Grimson, and Faugeras (2000) introduced the similar PCA model. PCA analysis is performed on a set of registered training samples, each represented by a signed distance function defined in a volume. Compared to the PDM, the linear PCA model on signed distance function is only valid within the limited space centered at the mean shapes, which is clearly a disadvantage.

Limitations of the PCA Statistical Model

The PCA statistical models have limitations both in the representation and application of global priors. The representation is incomplete, especially when the structure is complex and with high variations due to age, gender, race and almost infinite types of pathological abnormalities. Coverage of these variations requires a large number of modes, in addition to the need for high-valued coefficients. Setting up such a large and high-dimension model space is impractical due to the limited size and coverage of the training set, in which every sample needs to be obtained by arduous manual segmentation. Further, searching in such a complex model space is also impractical if any reasonable performance is desired. Therefore, any practical PCA model space will be restricted to that of a relatively low dimensional, and the model shape can only vary in the neighborhood of the mean shape.

The established PCA model is applied to segmentation as follows. In an iteration of surface evolution, a model is identified from the PCA model space such that it has the minimum shape difference from the current surface. The difference is then used as a penalty term and added to the evolving force. In such a scheme, how to determine the relative strengths of model and local force is a critical issue. In a PDM, the local image force is only used to evolve the surface to its best model, and the shape residues not covered in the limited model space will be ignored. A level set scheme better understands the limitations of a model, and allows the surface to converge to a shape not covered by a model under local image force. The high-level model force and the local image force are added together, with a multiplier on one of the two terms to adjust the relative strength.

For a PDM, precise segmentation of a complex structure is obviously not possible, since the object is completely restricted to an inaccurate model space. In the literature, PDM showed its success mostly in 2D cases. Mitchell et al. (2002) extended it to 3D segmentation of a left ventricle from cardiac MR and Ultrasound images, in which the object of interest has a relatively simple shape and hence the issue of inaccurate model space is not severe. In some applications, such as ultrasound, the shape is not well defined. Even humans cannot consistently delineate the object boundaries. The goal is then to find a shape that best fits the visual clues of boundaries. This is where PDM is useful but, strictly speaking, it is geometric modeling rather than precise segmentation. On the other hand, geometric modeling sometimes is the best we can do since validation of results is not possible. Depending on application, geometric modeling may be a satisfactory solution even if it does not represent the actual object boundaries. When there is consistent ground truth from human, precise segmentation should be achieved.

In comparison, the level set scheme is fit for precise segmentation, since surface evolution is allowed to converge to local image primitives. However, the PCA model (Tsai et al., 2003; Yang, Staib, & Duncan, 2004) introduced the same-level competition scheme between global prior and local image forces. Because of the inaccurate model, the global prior term will often disagree with the local image term. In a situation where the shape is not well defined, the model term can be made very strong by adjusting the multiplier, and the resulting surface boundaries will be very close to that of a model shape. A example is shown in Figure 5c and 5d, in which a prior shape model was applied. In such a case, the user has only vague visual clues to identify the actual object boundary, therefore the boundaries suggested by a model shape are well accepted. Much work in the literature follows these types of strategies.

In a situation where an object shape is well defined, the user would expect a precise segmentation. A good example is the human vertebra in high-contrast data, such as multi-slice CT images, as shown in Figure 6.

This 3D complex structure has a relatively well-defined shape, at least for a human observer, who can perform consistent delineation of the object boundaries on 2D slice images. However, there are many places where local boundaries are weak, diffused and have gaps. Further, such a structure is adjacent or even connected to a neighboring structure that has similar image properties. This is the typical case where we cannot afford an accurate model, and the inaccurate model will compete with the local image forces. If we make the model strong, the surface will not converge to some of the strong local features and abnormal shapes that are out of the model space. On the other hand, if we reduce the strengths of the model, then level set will leak out at places where local primitives are weak and hence need guidance. Obviously, this is a dilemma that cannot be completely addressed by adjusting the multiplier. In the work in which Leventon et al. (2000) defined this competition scheme, an attempt was made to segment the human vertebrae. The training shapes are from the vertebrae of the same data set, leaving the middle vertebrae out as the test data. The problem manifested itself. The not-well-fit model shape competes with local primitives to pull the evolving surface away from these strong features and the abnormal shapes. From the published result, the convergence was quite far from the object boundaries that can be easily delineated by a human observer.

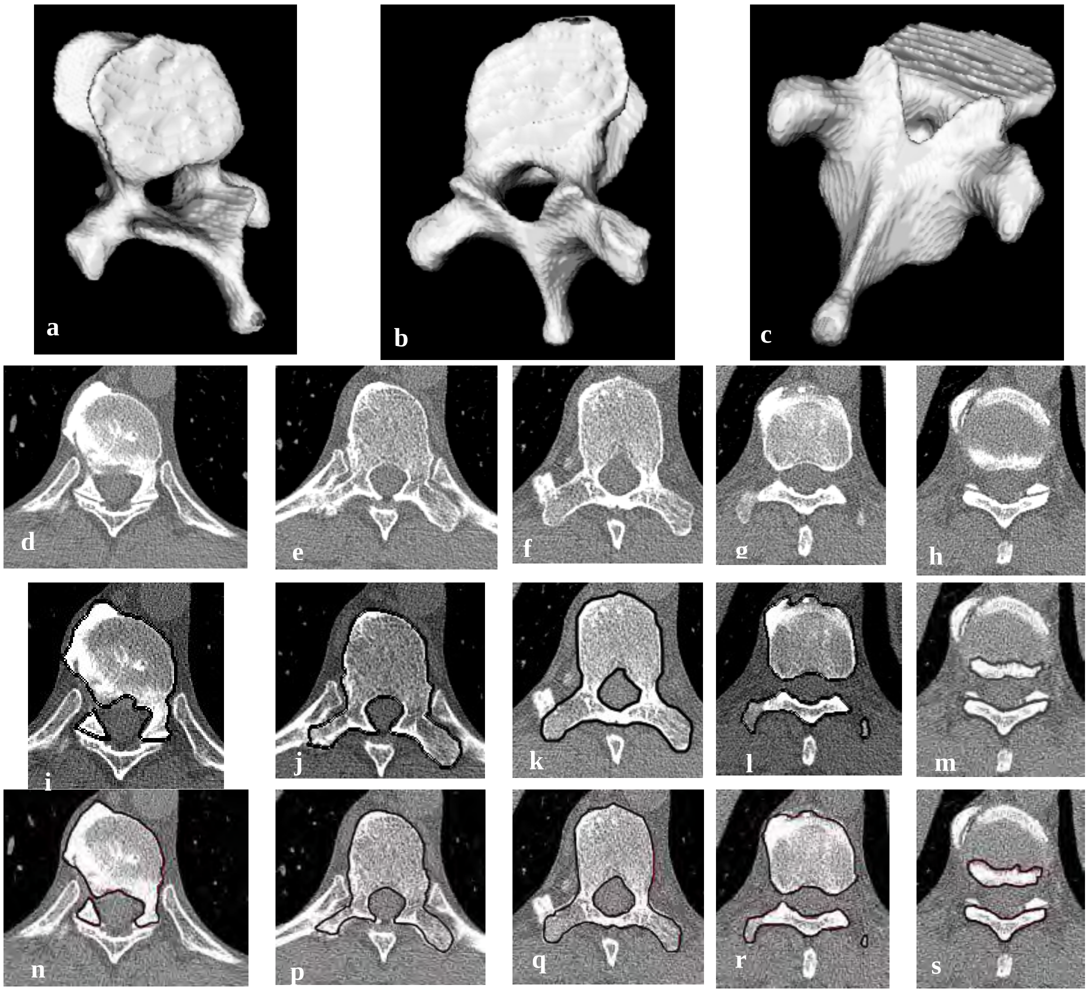
A Better Strategy

The limitations of the PCA models for representing global prior knowledge can be understood from an alternative point of view. Mathematically, the PCA models are based on the moment average of the training samples, and hence high frequency information will not be preserved. On the other hand, human observers to identify a complex structure most effectively use high frequency information, such as abrupt ridges, corners, etc. These are singular points, lines or planes where derivatives do not exist. A surface model represents smooth surfaces well, but does not incorporate these singular features. In other words, this important portion of the prior knowledge has to be represented more explicitly rather than implicitly embedded in the shape model.

Further, global prior knowledge, such as surface models, should not be put in competition with local image properties at the same level. When simply added to the local image force, a global prior loses it “global” status and degrades into a local factor. A global prior should act as a high level guidance to local evolution, so that the surface will converge to the “correct” local primitives, hence avoiding the problems of local minima and leak-out.

To determine what global priors are needed, we want to find out what causes the failure of the surface evolution when no global prior is applied. As shown in Figure 6, the internal intensity of the vertebra is very nonhomogenous, and there are

Figure 6. 3D level set segmentation of the vertebra



(a)-(c): 3D view of the complex structure of a vertebra shape. (d)-(h): Vertebra on selected axial slices, with its neighboring vertebra and rib structures. While the shape looks well defined for a human observer, local boundaries are far from clear. Also notice the pathology induced abnormal shape on the top-left of (d). (i)-(m): The 2D boundaries delineated by a human observer. Regions outside the contours belong to the neighboring vertebras and rib structures. (n)-(r): Converged 3D level set surface of segmented object projected onto selected axial slices.

many internal edges. A level set front when initialized inside the vertebrae will ensure converge to these internal edges. The only way to prevent this from happening is to keep these internal edges away from the evolution path. Since the convergence of level set depends on the initial front, a major effort is to design the initial front such that it is close to the boundaries but outside of the vertebrae. The surface will then converge to the outside boundaries before it even encounters the internal edges. Another problem is the weak boundary locations, from which the surface will leak out. We can carefully design the speed map according to the various local properties to avoid this problem. For instance, we can use region based speed at locations where edges are weak. Next, a more challenging issue is the adjacent, or even connected, ribs and vertebrae that have the same image properties. To separate them from the vertebrae of interest, we need to explicitly detect the high-level features formed between the vertebrae and its neighboring structures.

From the above, global priors should then include a surface model as a general description of shape, augmented by a set of known high-level features. The surface model is constructed as follows: First, a small number of vertebrae, roughly covering the spectrum of shape variations within the normal range, are selected and their surfaces are manually segmented. These surfaces are transformed into a common coordinate system using similarity transformation such that the spinal channel and inter-vertebrae planes are registered. We do not intend to match the surfaces themselves, since that is unnecessary and will require nonlinear warping of the surfaces. Afterwards, the registered surfaces are represented as a set of signed distance functions. We then compute the mean signed distance function, whose shape is shown in Figure 7a. It has the basic look and structure that belong to and characterize a vertebra. The mean shape model is used as a rough guide or template and hence does not need to cover the large variations. This low requirement allows a small sample size.

Using the mean shape as a reference system, we define planes that divide the volume into regions. For each region, the proper speed design is applied, which varies in the type of speed and the constants. The region division is shown in Figure 7b and 7c.

The mean shape representation is augmented with a set of high-level features, which are formed by linear primitives including intensity edges, ridges and valleys. If the locations of the strong and same-type primitives in a local neighborhood roughly fit to a simple parametric model, such as a plane, that plane is recorded as a high-level feature. There are two types of high-level plane features, and both are recorded with reference to the common coordinate system. If fitted by the ridge or edge primitives from the object boundary, it is named a high-level boundary feature. A high-level context feature, on the other hand, is fitted by valley primitives located at the interface between the object and a neighboring structure. Plane is the simplest parametric model, which can be extended to piecewise linear models. Examples of high level boundary and context features are shown in Figure 7d and 7e.

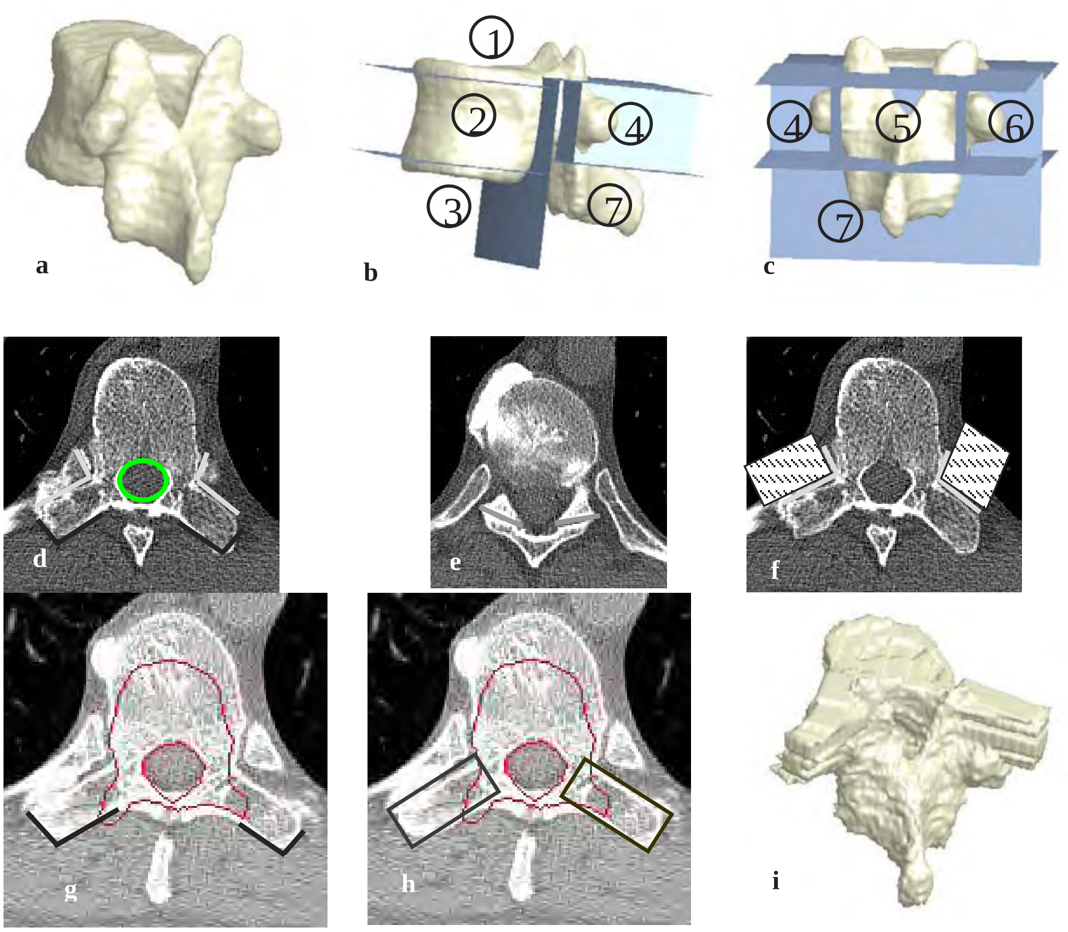
Application of Global Priors in Surface Evolution

Similar to the registration process of training samples, the mean shape is mapped to the image data using the estimated similarity transformation, as shown in Figure 7g. This is, of course, specific to this application. However, a transformation can be estimated with other more general methods to map the initial shape, and we have seen examples in the literature (Baillard, Hellier, & Barillot, 2001). The boundaries of the mapped mean shape are usually close to the boundaries of the object, except that the surfaces at the two transverse processes are too far inside. Such an initial front will be captured by local minima and not converge to the outside true boundaries. It therefore needs to be modified, and we use the high level features to achieve this.

The high-level feature models are also mapped to the neighborhood of the features in the image data. The detection of high-level boundary and context features uses mainly parametric model-fitting techniques, such as Hough Transform. This is a typical knowledge-based strategy, which fits known models to local image data, rather than estimate models from detected image primitives. The detected high-level features have impacts on the correct convergence of the surface, as described below.

As shown in Figure 7g, the high-level boundary features at the two transverse processes are detected, and we use them to modify the mean shape into a good initial front. We place two rectangular parallelepipeds in connection with

Figure 7. Representation and application of prior knowledge



(a) The mean shape. (b)(c) The plane division of the volumes into regions. Each region has its specific speed design. (d)-(e) High-level boundary features and context features which are marked with black and gray, respectively. A context feature is not on object boundary, but is on the object interfaces. (f) Blocker speed regions. (g) Mean shape mapped to the image. Two high-level boundary features are detected. (h) The mean shape is augmented to make it close to the object boundaries. (i) 3D structure of the augmented mean shape to be used as the initial front for level set surface evolving.

the mean shape to form the initial front, as shown in Figure 7h and 7i. This is a general idea that is applicable to many applications: If we are given a true knowledge of a partial boundary, we can use it as part of an arbitrary shape to augment the initial front. After the modification, the initial front is pulled to the actual boundary, which helps in avoiding the internal local minimums.

The high-level context features, as shown in Figure 7d and 7e, are used for the introduction of a new type of speed to affect level set evolution—namely the blocker speed. The blocker speed serves as an energy potential that prevents the surface from evolving beyond the detected context feature. With this we can prevent the leakage to neighboring objects. As shown in Figure 7f, the two plane context features divide the space into two regions. Using the center of the spinal channel as the reference, we can determine the region which the surface should not evolve into. For that region, the speed terms are made negative, which prevents the surface from evolving beyond the context features. Note that the high-level context features do not belong to object boundary; rather, they belong to background. The surface will not converge to the context feature, but will converge to the object boundary that is on the inner side of the context feature.

From the above, the leakage and local minimum issues that frequently bother a level set surface evolution are principally addressed. Finally, the separation planes associated with the mean shape are also mapped to the image data. Local primitive detection is performed to generate the speed terms specially designed for each region. Careful design of the speed is also an important part of this knowledge-based segmentation scheme.

As shown in Figure 6n-6s, after iterations of surface evolution, the level set surface converges to the correct object boundaries despite the internal inhomogeneity, the weak boundaries and the presence of neighboring structures. From this example we showed the power of prior knowledge when it is well represented, organized and applied to medical image segmentation. We only presented high level concepts in this example, and the reader is referred to the work of Shen et al. (2005) for a detailed treatment including mathematical deductions.

CONCLUSIONS, DISCUSSION, AND OUTLOOK

We have made an attempt in this chapter to give an intuitive introduction to the various methods and frameworks that are useful in 3D medical image segmentation. We also discussed the concepts of our approach of knowledge-based segmentation, in which the effective representation and application of prior knowledge is most important. From the discussion, the most prominent high-level features are of most value and need to be intensively explored before other image properties come into play.

The state-of-the-art medical image segmentation methods still leave much to be desired. Other than the difficulties of the segmentation problem itself, the relatively short time in which the 3D medical data become available is also one factor. The earlier scanners produced data in which the axial resolution is much lower than the other two directions. Therefore 3D correlation between two adjacent axial slices is weak, and a 3D segmentation method is not particularly more beneficial than 2D segmentation on each slice image. In the past two years, multi-slice CT scanners have achieved true isotropic voxels. With the development of medical instrument technology, 3D medical image segmentation will become increasingly popular. 3D methods are inherently better for the segmentation of complex structures, in that they implicitly involve the 3D structural information that is difficult to capture in 2D methods. On the other hand, it will still take time for people to have full awareness of the properties of these data and develop insights.

Medical imaging applications have now become an important driving force for the advance of computer vision. On the other hand, medical image analysis needs to address real-word issues that have been outside the realm of computer vision. These issues come largely from the fact that the end systems are mostly used by the physician. The human factor is essential, since any successful solution will have to be accepted by a physician and integrated into one's medical procedural workflow. This put strong constraints on the type of applicable methods. Because of this, there has been a discrepancy between the advanced frameworks presented in computer vision and the low-level and ad-hoc methods used by researchers working on real medical application solutions. Recently, we have seen the hope of practically applicable frameworks which are relatively simple in implementation, and reasonable in performance and robustness. In this respect, we foresee several trends. First, existing and coming frameworks from computer vision will improve themselves to be more practical and compete for their prevalence in medical image analysis. Second, more medical imaging researchers will make use of these frameworks, with their efforts focused on the incorporation of application-specific priors. Third, more efforts from the computer vision community will be attracted to solving real-world problems, and, hence, lead to the invention of more general methods for prior knowledge incorporation. As a byproduct, general frameworks will also be invented for the incorporation of user interaction into difficult segmentation problems.

ACKNOWLEDGMENTS

I would like to thank my colleagues at Siemens Corporate Research: Dr. Gozde Unal, Dr. Greg Slabaugh, Dr. Leo Grady, Dr. Li Zhang, Dr. Zhizhou Wang, Zhigang Peng, Dr. Marie-Pierre Jolly, Ludek Hasa and Julien Nahed for their support in writing this chapter. Gratitude also goes to Dr. Khalid Al-Kofahi from Thomson Legal & Regulatory and Yonggang Shi from Boston University. They have provided useful images and references, and most importantly, insightful discussions, without which this chapter would not be possible.

REFERENCES

Adalsteinsson, D., & Sethian, J. (1995). A fast level set method for propagating interfaces. *Journal of Computational Physics*, *118*, 269-277.

Adams, R., & Bischof, L. (1984). Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *16*(6), 641-647.

Al-Kofahi, K., Lasek, S., Szarowski, D., Pace, C., Nagy, G., Turner, J.N., et al. (2002). Rapid automated three-dimensional tracing of neurons from confocal image stacks. *IEEE Transactions on Information Technology in Biomedicine*, *6*(2), 171-187.

Baillard, C., Hellier, P., & Barillot, C. (2001). Segmentation of brain images using level sets and dense registration. *Medical Image Analysis*, *5*, 185-194.

Basak, J., Chanda, B., & Majumder, D. D. (1994). On edge and line linking with connectionist models. *IEEE Transactions on Systems, Man, and Cybernetics*, *24*(3), 413-428.

Brejl, M., & Sonka, M. (2000). Directional 3D edge detection in anisotropic data: Detector design and performance assessment. *Computer Vision and Image Understanding, 77,* 84-110.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 8*(6), 679-698.

Chakraborty, A., Staib, L. H., & Duncan, J. S. (1996). Deformable boundary finding in medical images by integrating gradient and region information. *IEEE Transactions on Medical Imaging, 15*(6), 859-870.

Chan, T., & Vese, L. (2001). An active contour model without edges. *IEEE Transactions on Image Processing*, *10*(2), 266-277.

Chang, Y. L., & Li, X. (1994). Adaptive region growing. *IEEE Transactions on Image Processing*, *3*(6), 868-72.

Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(6), 681-685.

Cootes, T. F., Taylor, C. J., Cooper, D. H., & Granham, J. (1995). Active shape models—their training and application. *Computer Vision and Image Understanding*, *61*, 38-59.

Freeman, W. T., & Adelson, E. H. (1991). The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*(9), 891-906.

Guy, G., & Medioni, G. (1996). Inferring global perceptual from local features. *International Journal of Computer Vision*, *20*(1/2), 113-133.

Haralick, R. M., Sternberg, S. R., & Zhuang, X. (1987). Image analysis using mathematical morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *9*(4), 532-50.

Herman, G. T., & Carvalho, B. M., (2001). Multiseeded segmentation using fuzzy connectedness. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(5), 460-474.

Hill, A., & Taylor, C. J. (1994). Automatic landmark generation for point distribution models. *Proceedings of the 5th British Machine Vision Conference*, *2*(2), 429-438.

Horkaew, P., & Yang, G. Z. (2003). Optimal deformable surface models for 3D medical image analysis. *IPMI*, 13-24.

Kass, M., Witkin, A., & Terzopoulos, D. (1988). Active contour models. *International Journal of Computer Vision*, *1*(4), 321-331.

Leventon, M. E., Grimson, W. E., & Faugeras, O. (2000). Statistical shape influence in geodesic active contours. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, *1*(1), 316-323.

Lorenz, C., & Krahnstover, N. (1999). 3D statistical shape models for medical image segmentation. *International Conference on 3-D Digital Imaging and Modeling*, (pp. 414-423).

Malladi, R., Sethian, J. A., & Vemuri, B. C. (1995). Shape modeling with front propagation: A level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *17*(2), 158-75.

Marroquin, J. L., Vemuri, B. C., Botello, S., & Calderon, F. (2002). An accurate and efficient bayesian method for automatic segmentation of brain MRI. *Proceedings of ECCV* (pp. 560-574).

Meyer, F. & Beucher, S. (1990). Morphological segmentation. *Journal of Visual Communication and Image Representation*, *1*(1), 21-46.

Mitchell, S. C., Bosch, J. G., Lelieveldt, B. P. F., van de Geest, R. J., Reiber, J. H. C., & Sonka, M. (2002). 3-D active appearance models: segmentation of cardiac MR and ultrasound images. *IEEE Transactions on Medical Imaging*, *21*(9), 1167-1178.

Mohan, R., & Nevatia, R. (1989). Using perceptual organization to extract 3-D structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*(11), 1121-1139.

Pentland, A. P., & Sclaroff, S. (1991). Closed-form solutions for physically based modeling and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*(7), 715-729.

Sarkar, S., & Boyer, K. L. (1994). Computing perceptual organization in computer vision, *World Scientific Series on Machine Perception and Artificial Intelligence.* Singapore: World Scientific.

Shankar, B. U., Murthy, C. A., & Pal, S. K. (1998). A new gray level based Hough transform for region extraction, an application to IRS images, *Pattern Recognition Letters*, 19, 197-204.

Shen, H., Liang, L., Shao, M., & Qing, S. (2004). Tracing based segmentation for labeling of individual rib structures in chest CT volume data. *Proceedings of the 7th International Conference on Medical Image Computing and Computer Assisted Intervention* (pp. 967-974).

Shen, H., Shi, Y., & Peng, Z. (2005). Applying prior knowledge in the segmentation of 3D complex anatomic structures. *Computer Vision for Biomedical Image Applications: Current Techniques and Future Trends*, *An International Conference on Computer Vision Workshop*, *Lecture Notes of Computer Science*, *3765* (pp.189-199). Beijing, China.

Slabaugh, G., & Unal, G. (2005). Active polyhedron: Surface evolution theory applied to deformable meshes. *Conference on Computer Vision and Pattern Recognition*.

Sonka, M., Hlavac, V., & Boyle, R. (1998). *Image processing, analysis, and machine vision* (2nd ed.). PWS Publishing.

Sun, J., & Gu, D. (2004). Bayesian image segmentation based on an inhomogeneous hidden Markov random field. *Proceedings of the 17th International Conference on Pattern Recognition* (pp. 596-599).

Suri, J. S., Liu, K., Singh, S., Laxminarayan, S.N., Zeng, X., & Reden, L. (2002). Shape recovery algorithm using level sets in 2-D/3-D medical imagery, a state-of-the-art review. *IEEE Transactions on Information Technology in Biomedicine*, *6*(1), 8-28.

Tsai, A., Yezzi, A., Wells, W., Tempany, C., Tucker, D., Fan, A., et al. (2003). A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Transactions on Medical Imaging*, *22*(2), 137-154.

Udupa, J. K., Saha, P. K., & Lotufo, R. A. (2002). Relative fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(11), 1485-1500.

Vincent, L., & Soille, P. (1991). Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*(6), 583-598.

Walker, K. N., Cootes, T. F., & Taylor, C. J. (2002). Automatically building appearance models from image sequences using salient features. *Image and Vision Computing*, *20*(5-6), 435-440.

Wu, Z., & Leahy, R. (1992). Image segmentation via edge contour finding. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 613-619).

Yang, J., Staib, H.S., & Duncan, J.S. (2004). Neighbor-constrained segmentation with level set based 3-D deformable models. *IEEE Transactions on Medical Imaging*, *23*(8), 940-948.

Zhang, L., Zhang, T., Novak, C. L., Naidich, D. P., & Moses, D. A. (2005). A computer-based method of segmenting ground glass nodules in pulmonary CT images: Comparison to expert radiologists' interpretations. *Proceedings of SPIE Medical Imaging*.

This work was previously published in Advances in Image and Video Segmentation, edited by Y. Zhang, pp. 250-269, copyright 2006 by IRM Press (an imprint of IGI Global).

Chapter 4.6
Implementation and Performance Evaluation of WWW Conference System for Supporting Remote Mental Health Care Education

Kaoru Sugita
Fukuoka Institute of Technology (FIT), Japan

Giuseppe De Marco
Fukuoka Institute of Technology (FIT), Japan

Leonard Barolli
Fukuoka Institute of Technology (FIT), Japan

Noriki Uchida
Global Software Corporation, Japan

Akihiro Miyakawa
Nanao City, Ishikawa Prefecture, Japan

ABSTRACT

Information technology (IT) has changed our lives and many applications are based on IT. IT can be helpful for remote mental health care education. Because there are very few mental health care specialists, it is very important to decrease their moving time. But it is not easy to use the conventional TV conference systems for ordinary people, mental health care specialists, and their students because they are not computer specialists. For this reason, we have developed a WWW conference

system. Our system can communicate between the mental health care specialists and their students by using the live video on WWW browser. In this paper, we show the implementation and the evaluation of proposed system. The experimental results over the Internet show that our system can be used for real time communication between Fukuoka, Ishikawa, and Iwate prefectures.

INTRODUCTION

Recently in Japan, mental health care has become a very important issue because there are many people suffering from mental problems. Also, there are only a few specialists and researchers to deal with these problems. For example, bullying at school is one of the big problems worrying teachers in Japan today. Also, there are other problems such as refusal to go to school and school violence. However, there are few counselors in Japanese schools. It should be noted that many of these specialists are also doing other works. In general, one mental health care specialist should take care for many patients and they want to see their face. They should take care of not only the counseling but also the individual aftercare.

Information technology (IT) has changed our lives. People can communicate between themselves and students can study various courses at anywhere and anytime using the Internet. IT can be helpful for mental health care, education, aftercare, and counseling for patients and their families. Because there are very few mental health care specialists, it is very important to decrease their moving time. Also, it is very important to see the facial expression and to talk to people for mental health care education, aftercare, and counseling. For this reason, the video and voice are needed.

Recently, because of the use of ADSL and FTTH (NTT West, 2004) many people can use several Mbps on the Internet. Therefore, many people can use the streaming live video and TV conference. For example, the streaming live video is used for the tourist attractions broadcasting (Fuji-shi, 2005; Kumamoto-shi, 2001; Sapporo-shi, 2004), assembly broadcasting (The House of Councilors, 2005; The House of Representatives, 1999;), and public information broadcasting (The ministry of public management, 2004). In these live streaming video methods, the image refreshed at regular intervals over WWW (Fuji-shi, 2005; Kumamoto-shi, 2001; Sapporo-shi, 2004) and RealPlayer or Windows Media Player (The House of Councilors, 2005; The House of Representatives, 1999; The Ministry of Public Management, 2004) are used.

Some TV conference systems are used as special hardware (NEC Engineering, 1996; Polycom, 2004; SONY, 2004; VTEL, 2004) and in other systems are used as application software or PC connected to the existing cameras (Advanced Solutions, 2001, 2002; Hitachi Hybrid Network, 2004; Microsoft, 2004; Visual Nexus, 2004). When using the special hardware, the users need to connect to the Internet and set up the hardware in each spot before using the system. But, it is difficult to use the special hardware in counseling because the users are usual people. While, when using the application software and PC is connected to the existing camera, the users need to connect the PC and camera, setup the PC, and install the application software in each spot before using the system. Also, it is difficult to introduce these systems for mental health care because the mental health care specialists and their student are not computer specialists.

In order to realize a remote mental health care education, we have developed a WWW conference system. Our system is able to support the communication between the mental health care specialists and their students. Also, our system can provide the communication between the mental health care specialists, patients, and their families by using the live video on WWW browser, point to point communication, point to multi-point communication, and multi-point to multi-point communication.

The organization of this article is as follows. In the next section, we will introduce the related works. Next, we describe the WWW conference system and explain system architecture. The conference types are treated in the following section. After that, we show the connection management for each conference type and the flow of WWW conference system. Then, we show implementation of our proposed system by using Macromedia Flash and Macromedia Flash Communication Server and provide the experimental results. Finally, we give some conclusions and future works.

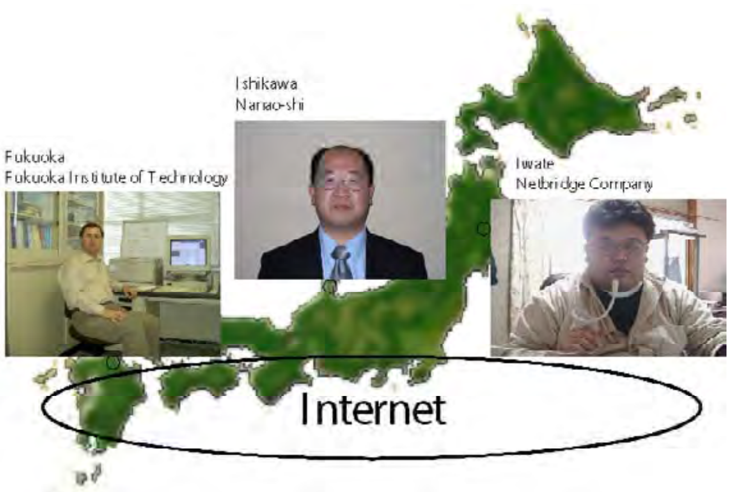
RELATED WORK

Existing communication systems are usually realized on the LAN environment and leased lines. Many remote education systems are proposed in Inoue, Okada, and Matsushita (1997), Mori, Oyabu, Nomura, and Oshita (1992), Sakiyama, Ohono, Mukunoki, and Ikeda (2001), and Wakahara (1998). In Inoue et al. (1997) and Wakahara (1998), the multi-point to multi-point communication is used. While in Sakiyama et al. (2001), a video stream selection according to lecture context is proposed. In these systems, the remote users can communicate between them using the live video.

Recently, the live video communication systems have become very popular and the network speed is increased very fast. Therefore, the video conference systems can be applied in the medical field. There are many video conference systems such as the medical tele-consultation support system using super high definition imaging system (Yamaguchi et al., 2001), remote medical support system using the transcoding function (Kawamura, Maita, Hashimoto, & Shibata, 2003), and care communications service between hospitalized patients and their families (Abiko, Iijima, Koyama, Kamibayashi, & Narita, 2003). In these studies, the system was implemented only for intranet or leased lines.

Some other communication systems using live video for remote learning systems (Maeda et al., 1997; Miyoshi, Okanaga, Kou, & Kondo, 2000) and streaming video (Kato, Jiang, & Hakozaki, 2003) has been proposed. In the multimedia communication environment for distance learning proposed in Maeda et al. (1997) are given

Figure 1. WWW conference



the experimental results for remote learning using intranet and Internet environments. In this study is shown that an efficient remote learning can be achieved by using 600 Kbps bandwidth. Now, Internet users can have more bandwidth by using ADSL and FTTH. And the live video over the Internet is possible. In Kato et al. (2003), the authors propose a streaming video system for best-effort networks using adaptive QoS control rules to improve the satisfaction of streaming video services.

However, these systems can't be used by usual people and they need special hardware and software. These systems use point to point communication, so they can not be used for multi-point to multi-point communication. In order to overcome these problems, we have developed a WWW conference system, which uses multi-point to multi-point communication and can transmit live video over the Internet.

PROPOSED WWW CONFERENCE SYSTEM

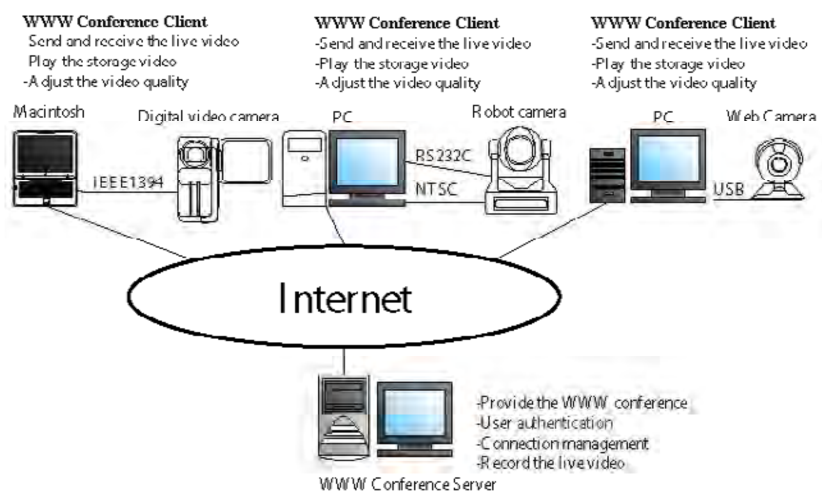
The proposed WWW conference system is shown in Figure 1. If the user has a video camera and PC connected to the Internet, the user can communicate to the remote users by using live video on WWW. Using our system, the user does not need to install and setup the application software. He just only needs to access the WWW conference server by using the WWW browser.

The concept of WWW conference is shown in Figure 2. The user can use various cameras such as the WWW camera, digital video camera and robot camera. But, the camera should be identified and connected to the PC. Our system is composed of the WWW conference client and the WWW conference server. When the user accesses the WWW conference server by using a WWW browser, the WWW conference client is downloaded to the PC. After downloading the WWW conference client, the system provides the following functions:

1. User authentication
2. Live video function
3. Video record function
4. Quality setting function

The user authentication function is provided by user ID and password. When the user enters

Figure 2. WWW conference system



into the WWW conference server, the user inputs its user ID and password. The user ID and password are compared with the registered ones in the WWW conference server. When the user ID and password do not match, the user login fails and the user needs to input again its user ID and password. It should be noted that the user login may fail when the WWW conference server has already logged on the maximum number of users.

The live video function provides the video data and voice data. The video data and voice data are sent and received by the video stream. They are encoded by H.263. Our system provides three types of conference using this function. In order to realize the live video, the WWW conference client has the connection management function, the video encode function, and the video decode function. The WWW conference server has the connection management function.

The video record function records the live video stream and plays the recorded video stream on the WWW conference server. The user can record the live video stream and play the recorded video stream by using the WWW conference client. The recorded video stream is stored in the WWW conference server. The recorded video stream is used for learning the existing cases of counseling and replaying the instructions for mental health care.

The quality setting function modifies the video quality for local WWW conference client and remote WWW conference client. The local WWW conference client can modify the video size and frame rate. But, the remote WWW conference client can modify only the frame rate. When the PC does not have enough CPU power or network bandwidth, the voice data are delayed and video data experiences frame loss and frame delay. However, our system can guarantee the quality of service such as "Frame rate is high and video size is small" and "Frame rate is low and video size is large" when the user modifies the video quality using this function.

SYSTEM ARCHITECTURE

In order to implement our system, we propose a system architecture, which is organized by the WWW conference client and the WWW conference server as shown in Figure 3.

WWW Conference Client

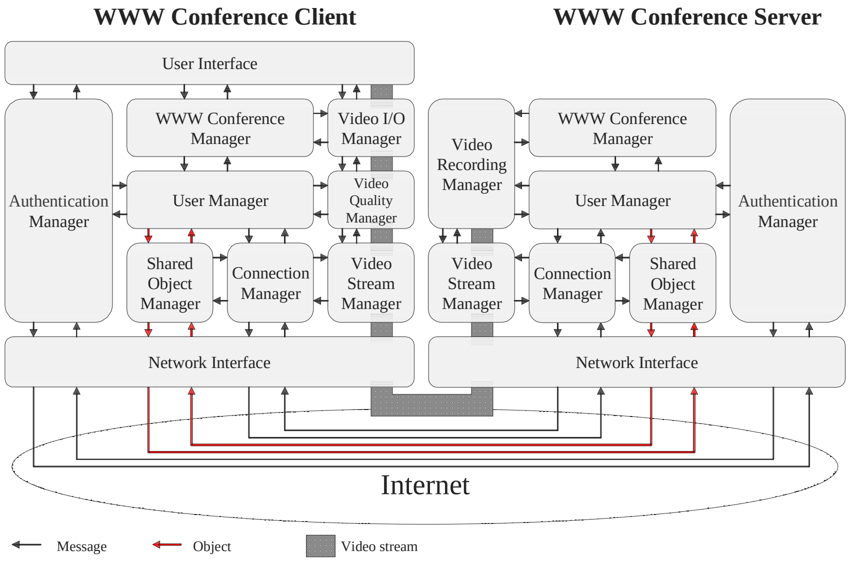
Presently, various network environments exist, such as broadband network and narrowband network. The bandwidth varies from 10 Kbps to 100 Mbps. In order to realize the communication between various network environments, the WWW conference client is organized by User Interface (UI), Authentication Manager (AM), User Manager (UM), WWW Conference Manager (WCM), Connection Manager (CM), Shared Object Manager (SOM), Video I/O Manager (VI/OM), Video Quality Manager (VQM), and Video Stream Manager (VSM).

The UI provides the following functions:

1. Insertion of user ID and password
2. Setup of video input and output device
3. Display of live video
4. Modification of video quality

The AM authenticates the user ID and password by using the registered user ID and password on the WWW conference server. When the AM fails for the authentication, the AM requests the user ID and password to the user again. When the AM authentication is a success for the authentication, the AM sends the user ID to the UM. The UM manages the system users, the video quality for each video stream and the name of the recorded video stream. When maximum number of users has been logged in to the WWW conference server, the UM rejects other users. The WCM sets up the number of maximum users and provides the user interface. The maximum number of users and the user interface are decided by the conference type. The CM manages the connections of

Figure 3. System architecture



the other WWW conference clients participating in the WWW conference. The CM converts the number of maximum users to the number of maximum connections. The SOM gets the shared object from the WWW conference server. The recorded video stream is played by using the video name. We use the shared object to record the name of video stream on the WWW conference server. When the user wants to record the live video stream through the UI, the SOM adds the name of video stream as the shared object to the WWW conference server. The VI/OM inputs the video data and sound data from the camera and the video stream is selected by the WCM. The VI/OM displays the video data and sound data from the video stream and the camera. Also, the VI/OM sends the video data and sound data to the VQM. The VQM decides the frame rate and the size of video by using the identified video quality. The VSM sends and receives the video data and voice data as the video stream to the WWW conference server. The other WWW conference clients operate in the same way.

WWW Conference Server

In order to realize the WWW conference and record the live video stream, it is needed to manage the video stream including live video stream and recorded video stream. So, the WWW conference server is organized by the AM, WCM, UM, CM, SOM, Video Record Manager (VRM), and VSM. The AM manages the user ID and password. It compares the user ID and password with the registered user ID and password when the AM requests the authentication from the WWW conference client. The UM manages the users inside the system, the video quality for each video stream and the name of the recorded video stream. The WCM manages the user and video stream based on type of conference. The CM manages the connection of the WWW conference clients. Then, it decides the number of maximum connections by using the number of maximum users. When the WWW conference server has already the maximum number of users, the CM rejects the new connection. The SOM stores the shared object, which is the name of video stream on the WWW conference server. The shared object remains even after disconnecting the WWW conference client. The VRM records and manages the video stream on the WWW conference server. The VSM manages the video stream between the WWW conference clients and the WWW conference server.

TYPE OF CONFERENCES

For the remote mental health care education are needed various conference types. We introduce three conference types because our system is used for the counseling, aftercare, and education.

Point to Point Conference

The point to point conference is used for counseling the patients. The point to point conference provides functions for point to point communication as shown in Figure 4. This type displays the local live video stream and the remote live video stream. The user can record the local live video stream and play the recorded video stream. Using the point to point conference, the patient can take the counseling at home.

Multi-User Conference

The multi-user conference is used for aftercare. The multi-user conference provides functions for the multi-point to multi-point communication as shown in Figure 5. In the multi-users conference, the users can communicate in a distributed way at the same time. This is why the local video stream and the multiple remote live video streams can be display at the same time. By using the multi-user conference, the patient and their family members can talk together with psychology specialists even if they are not at the same place.

Figure 4. Point to point conference

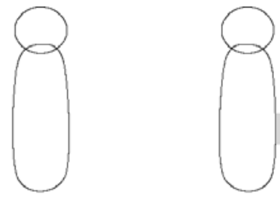


Figure 5. Multi-user conference

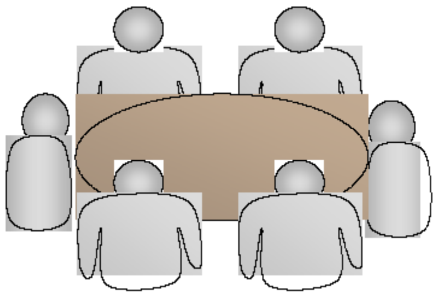
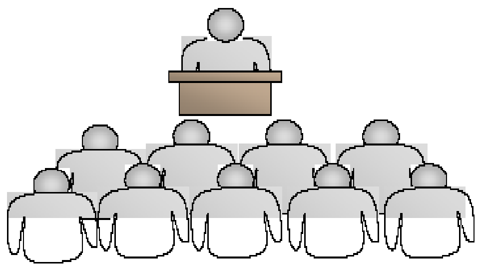


Figure 6. Broadcast conference



Broadcast Conference

The broadcast conference is used for health care education. It provides broadcast functions to broadcast the live video to many users as shown in Figure 6. This method is used for live broadcast (The House of Councilors, 2005; The House of Representatives, 1999). The broadcast conference uses two types of WWW conference clients: the sender and receiver clients. The sender client displays the local live video, selects the sending channel, sends the live video stream to receiver client, and records the local live video stream. The receiver client selects the receiver channel, receives and displays the video stream and plays the recorded video stream. Using the broadcast conference, the students can learn the clinical psychology, the family and patient can learn the mental health care at home.

CONNECTION MANAGEMENT

In order to provide the live video for distributed users, the video data and voice data are transmitted in the video streams. The number of video streams is decided based on the type of conference and the number of users. The WWW conference client can be connected to the remote WWW conference client via WWW conference server to create the video stream. Our system uses the channel to create and manage the video stream. Each channel has a channel number. The channel number is used for the clients connection and selection of the video stream. In the point to point conference and multi-user conference, the channel numbers are assigned based on the connection order. While, in the broadcast conference, the channel numbers are selected by the user.

An example of the relation between the video stream and channel number in the point to point conference is shown in Figure 7. This conference uses two video streams in each WWW conference client. One is used for the sender client and another for the receiver client. The sender video stream and receiver video stream have different channel numbers. After allocating the channel number, the WWW conference client creates the video streams by using the channel number.

In this example, the user 1 is connected to the WWW conference server before the user 2. For this reason, the channel number starts from 1. The channels 1 and 2 are shown in Figure 7. In the channel 1, the user 1 is the sender and the user 2 is the receiver of the video stream. While, in the channel 2, the user 2 is the sender and the user 1 is receiver.

The relation between the video stream and channel number in the multi-user conference is shown in Figure 8. In this type, each user has its own video stream. The channel number is assigned the same as type one. After allocating the channel number, the WWW conference client creates the video stream for each receiver.

In this example, four users are connected to WWW conference server. In the channel 1, the user 1 is the sender and the users 2, 3 and 4 are receivers. Each WWW conference client creates the video streams by using these channel numbers.

The relation between the video stream and channel number in the broadcast conference is shown in Figure 9. The user can select the sender channel number using the sender client. Then, the video stream is created based on the selected channel number. Also, the user can select the receiver channel number using the receiver client. It should be noted that a user can not send the video

Figure 7. Connection management for point to point conference

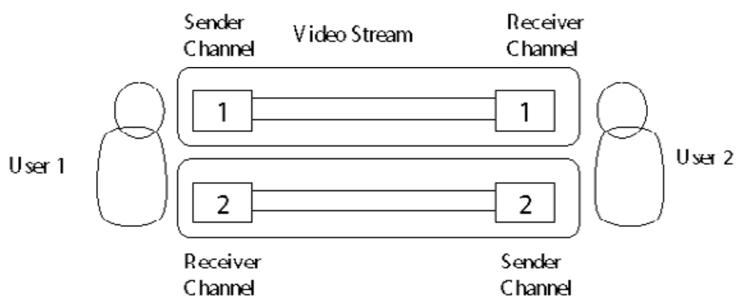


Figure 8. Connection management for multi-user conference

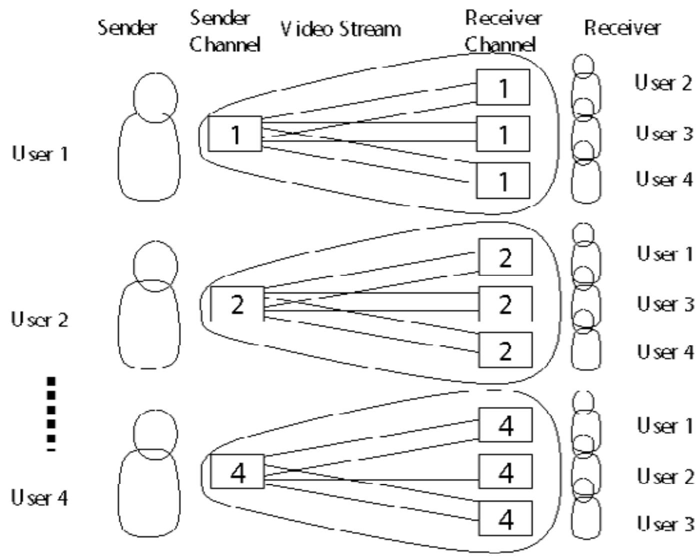
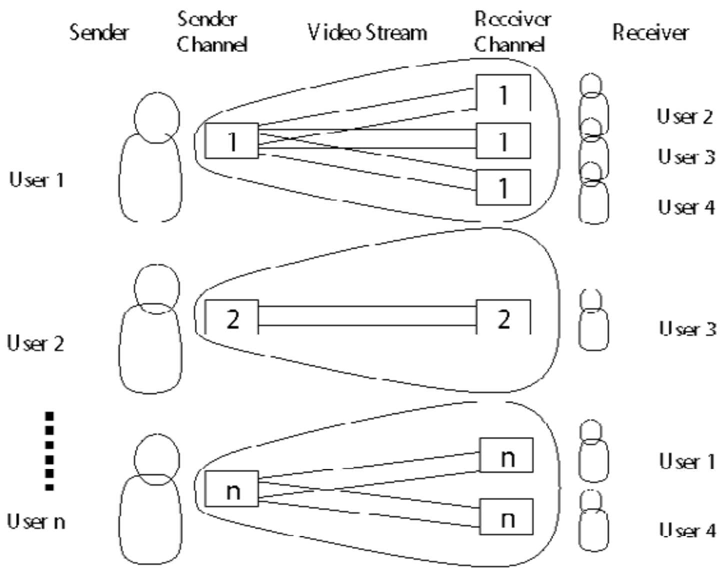


Figure 9. Connection management for broadcast conference

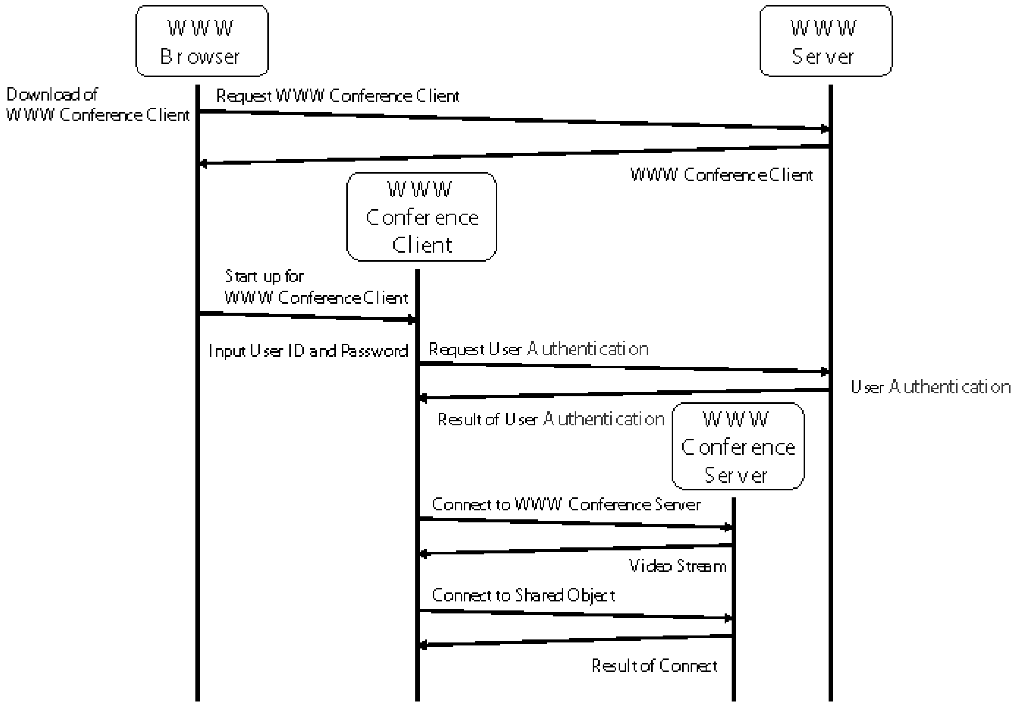


stream in a channel when this channel is used by another user. But, the receivers can get the same channel at the same time. When the sender channel number has already been used, the WWW conference client displays this message: "This channel number has already been used."

WWW CONFERENCE OPERATION

The WWW conference operation has two phases: the start up phase and live video communication phase. In the start up phase, the WWW conference client is downloaded and the user enters into the WWW conference server using the WWW

Figure 10. Startup phase



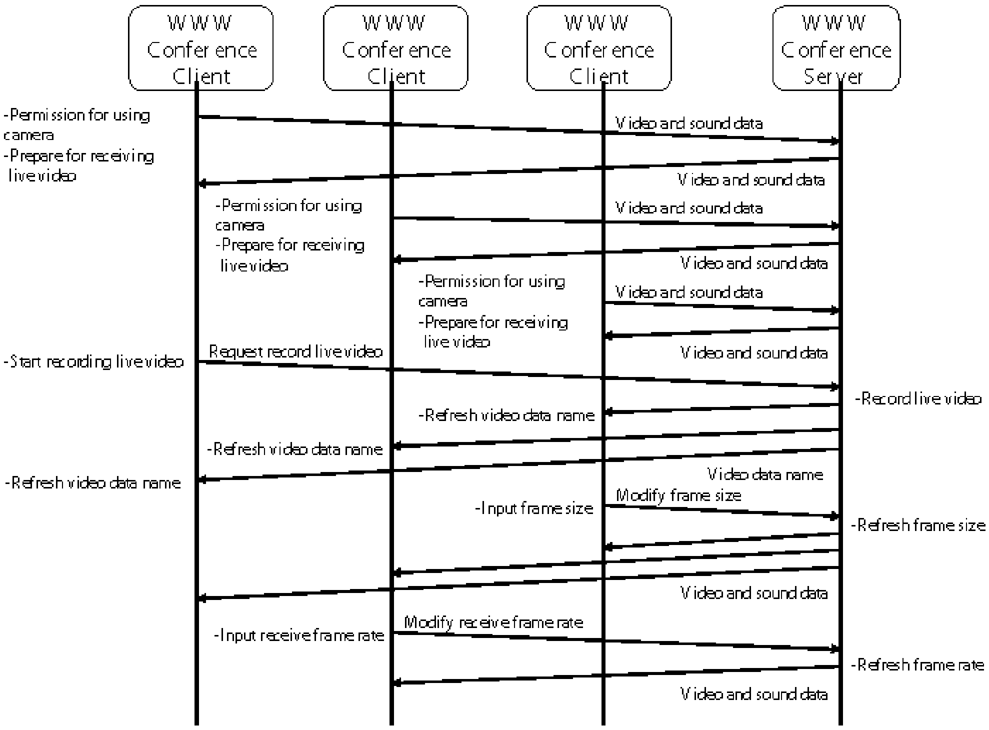
browser. The live video communication phase consists of camera setup, video quality modification, and sending and receiving the live video.

The start up phase is shown in Figure 10. First, the user downloads the desired type of the WWW conference client from the WWW conference server. After that, the WWW conference client starts up automatically and requires the user ID and password. When the user inputs the user ID and password, the WWW conference client sends this information to the WWW conference server to authenticate the user. When the authentication fails, the WWW conference client requests the user ID and password again. When the authentication is a success, the WWW conference client creates the video streams. Then, the WWW conference client connects to the shared object to share the name of recorded video stream with other WWW conference clients.

The live video communication phase is shown in Figure 11. First, the WWW conference client asks the user for the permission to use the camera. Next, the WWW conference client allocates the channel number to send and receive the live video stream. When the WWW conference client finishes the channel allocation, it starts to send/receive the live video streams to/from other clients via WWW conference server. When the user requests to record the live video stream, the WWW conference client sends this request to the WWW conference server.

The WWW conference server adds the new name of live video stream to the shared object. Then, the name of live video stream is added to the connected WWW conference clients. When the user inputs the frame size, the WWW conference client updates the frame size, modifies the captured frame size and sends the modified frame size to the WWW conference server. When the WWW conference server receives the modified frame size, it updates the new frame size and sends the live video stream to the WWW conference

Figure 11. Live video communication phase



clients using the new frame size. Also, the frame rate is updated the same as the frame size.

IMPLEMENTATION

The implemented system structure is shown in Figure 12. The WWW conference client is implemented by using the Macromedia Flash Studio MX1.5. The WWW conference server is realized by using the Apache1.3.29 and the Macromedia Flash Communication Server MX1.5. The Apache1.3.29 is used to provide the WWW conference client to the users. The Macromedia Flash Communication Server MX1.5 is used for sending and receiving the video data and voice data. The live video streams are encoded by H.263. We implemented three types of applications for each type of conference. When the user downloads the desired type of application, the WWW conference client displays the authentication interface. The user can take part in the WWW conference after authentication.

The point to point conference application is shown in Figure 13. The bottom side of interface shows the remote live video, recorded live video, and local live video from left to the right side. The list box and play button are shown below the recorded video. The list box is used to select the name of video stream and play button is used to play the recorded live video stream. The stop button is shown in the left side below the local live video stream. This button is used to stop the recorded live video stream. The record button is shown in the right side below the local live video. The user can record and play the live video stream by using these buttons. The main video display can be selected by clicking one of the videos shown in the bottom side. The video quality setup is shown in the upper side of the interface. The interface also has in the bottom side some list boxes, which are

Figure 12. Implemented system structure

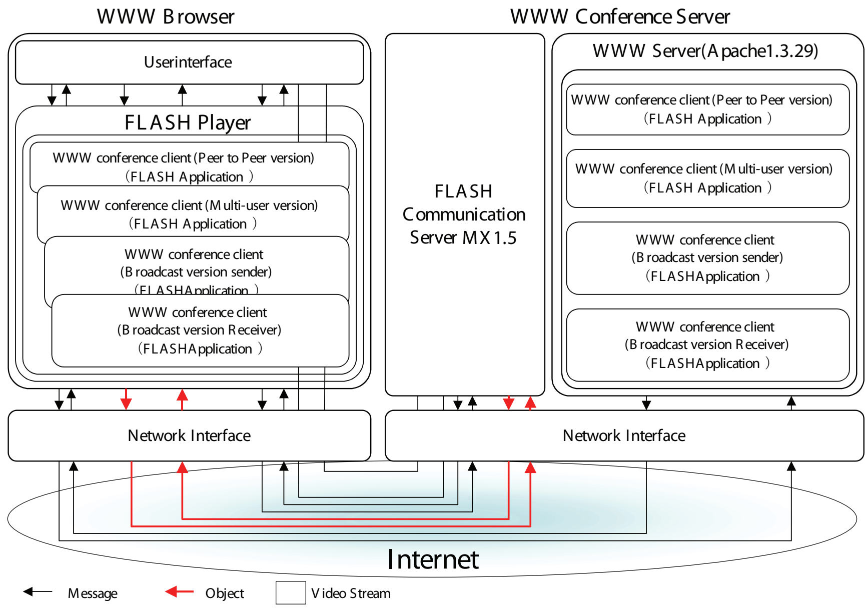


Figure 13. Point to point conference application

Figure 14. Multi-user conference application

used to change the frame size and frame rate. The frame size can be changed only at the local live video. But, the frame rate can be changed in both: the local live video and remote live video.

The multi-user conference application is shown in Figure 14. In the bottom side of interface are shown the remote live videos and local live video. The right side displays the local live video. The user can modify the frame rate for each remote

Figure 15(a). Broadcast conference sender application

Figure 15(b). Broadcast conference receiver application

live video using the list box. Also, the user can modify the frame rate and frame size for the local live video using the list box.

The broadcast conference applications are shown in Figures 15(a) and (b). These applications include the channel select interface, video quality set up interface and video display interface.

The sender client interface is shown in Figure 15(a) and the receiver client interface is shown in Figure 15(b). The main video in the center of the display shows the snap shot of the live video captured from the video camera. In the bottom side of interface are shown the text box, stop button and play button. The text box displays the name of recorded video stream. On the right side of the interface are shown the sender channel select buttons. The user can select the channel number from 1 to 12. Also, the user can input the name of live video stream and record the live video on the WWW conference server. When the user pushes the channel select button, the sender client starts to send the live video stream and the name of live video stream using the selected channel number. The video quality set up interface has the list box shown below the video. The user can modify the frame size and frame rate by selecting the values in the list box. The video display interface displays the live video and interface select buttons. The receiver client interface is almost the same with sender client interface, but the sender client interface can record the live video while the receiver client interface can play the recoded video.

EXPERIMENTS

Experimental Results

In order to evaluate our system for remote mental health care education, we used our system in a real Internet environment between Fukuoka, Ishikawa, and Iwate prefectures. Our evaluation environment is shown in Figure 16. The LANs are connected to the Internet environment by leased line (1 Gbps), FTTH (100 Mbps) and ADSL (12 Mbps). The experiment showed that the live video quality is good enough for communication. But it should be noted that sometimes the live video quality deteriorates because the bandwidth is not enough to send and receive the live video. In such cases, the frame rate is decrease and voice data is delayed about 1 second. Also, we evaluated the performance of our system on the LAN environment (100 Mbps) to analyze the possibility of realization of remote mental health care. We analyzed the throughput and load average for the multi-user conference. In this experiment, we use low performance computers because usual people use in general low performance computers. We used five PCs in this experiment. One PC is used as WWW conference server and four other ones are used as the WWW conference clients. The PC characteristics are shown in Table 1.

From these results, we conclude that the bandwidth of the LAN environment is enough for our system. But, the PCs should have more power to send, receive and display the live video.

The relation between the frame rate and load average at client side is shown in Table 2 and the relation between the frame rate and throughput is shown in Figure 17. We see that the load average and throughput are affected by the frame rate at client side. But, we found that the throughput is not changed more than 12 fps. This is because the PCs have not enough CPU power and the load average is close to 100%. After that, the throughput remains the same. In this case, the frame is delayed and the load average becomes high.

The relation between the frame size and load average at client side is shown in Table 3 and the relation between the frame size and throughput is shown in Figure 18. We see that the load average is affected by the frame size at client side. But, we found that the throughput is not affected by frame size.

The relation between the number of clients and load average at client side is shown in Table 4 and the relation between the number of clients and throughput is shown in Figure 19. We see that the load average and throughput are affected by the number of clients. Especially, the data sent from the server are affected by the number of clients because each client needs to receive all of live videos except for the own live video. But, the data sent from the client remain the same.

Comparison of Proposed and Conventional Systems

In Table 5 we compare our proposed system with conventional systems considering network environment, implemented platform, hardware/software, installation and setup, and communication service.

Some of conventional systems can be used not only in LAN environment/leased line, but also in Internet environment. However, these systems don't run in the multi-platform OSs. Using these systems, each user needs to install and setup the special hardware or software. In some of them is needed only the setup, but each user needs to make installation of special hardware or software. Furthermore, there are few systems, which can be used for point to point, multi-point to multi-point, and broadcast communication.

Table 1. PCs characteristics for performance evaluation

	CPU	Memory	OS
Server	Pentium4 (1.6GHz)	1G	WindowsXP
Client 1	Pentium3 (450MHz)	128M	Windows2000
Client 2	Pentium3 (667MHz)	128M	WindowsXP
Client 3	Pentium3 (664MHz)	128M	Windows2000
Client 4	Celeron (700MHz)	192M	WindowsXP

Figure 16. Evaluation environment

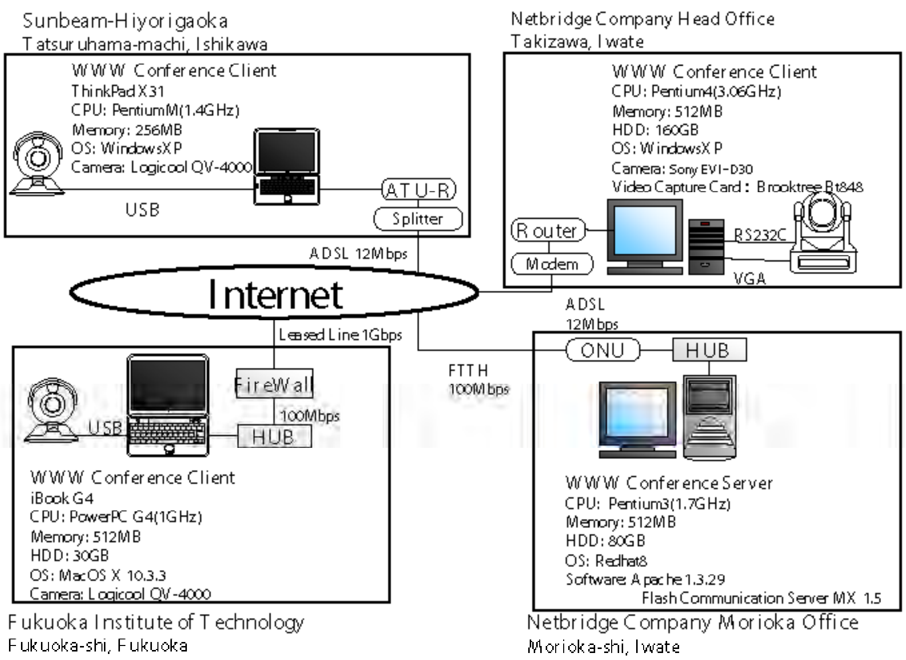
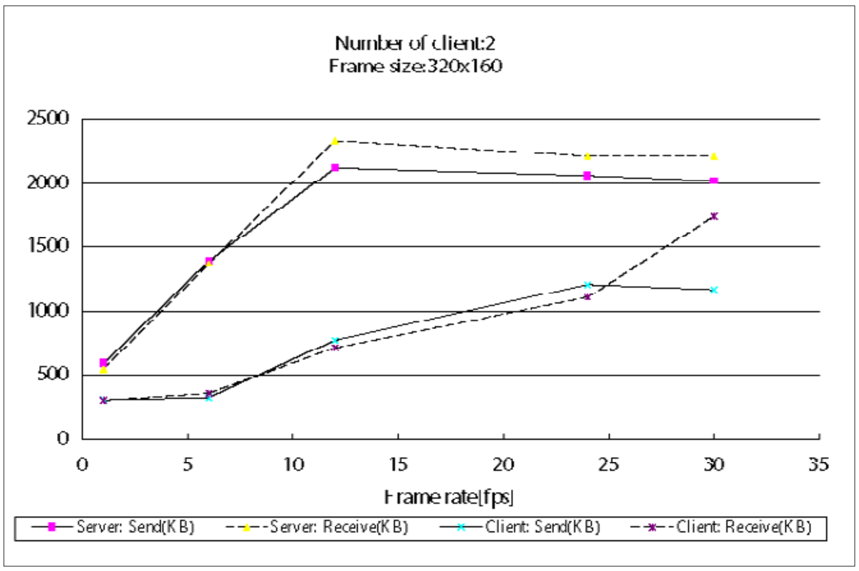


Figure 17. Relation between frame rate and throughput



While, our proposed system can be used in real Internet environment, and Windows, Macintosh and Linux OSs. In order to use our system, the user needs only to access the WWW conference server by using the WWW browser. After that, the system is installed and set up automatically. Also, the proposed system supports point to point, multi-point to multi-point and broadcast communication.

Table 2. Relation between frame rate and load average

Frame Rate	1	6	12	24	30
Load Average	40	69	87	93	94

Table 3. Relation between frame size and load average

Frame Size	160x120	320x240	640x480
Load Average	30	69	92

Figure 18. Relation between frame size and throughput

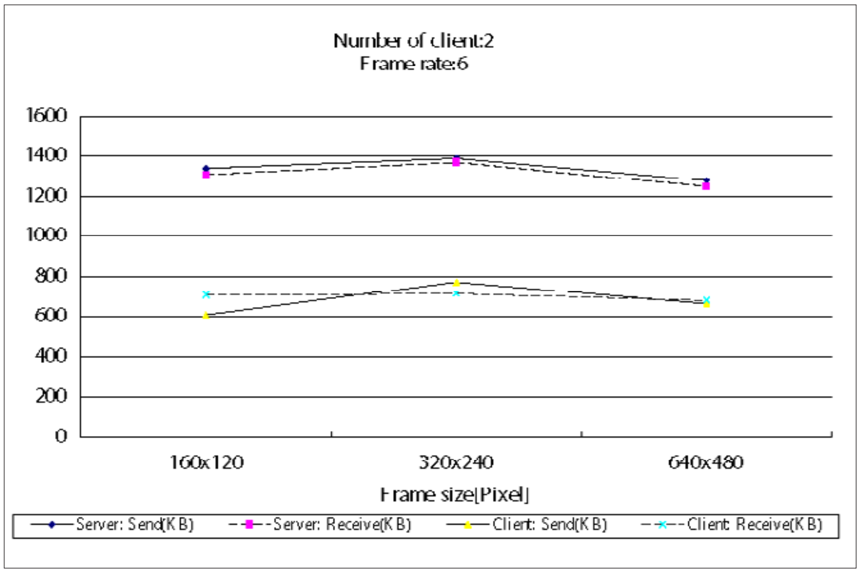


Table 4. Relation between number of clients and load average

Number of Clients	2	3	4
Load Average	69	76	87

CONCLUSION

In this article, we proposed a WWW conference system for supporting remote mental health care education. We showed the conference types and connection management methods for realizing each type of conference. We presented the WWW conference operation. Finally, we discussed the implementation and evaluation of our proposed system. From experimental results, we conclude as the follows:

1. The proposed system can provide the same communication between Fukuoka, Ishikawa, Iwate prefectures over the Internet.
2. The WWW conference client can send and receive the live videos for four users in the LAN environment.
3. The WWW conference client needs a high CPU power to send, receive, and display the live video.

Now, we are evaluating the proposed system for different kind of networks and number of hops. In the future, we would like to use the proposed system for mental health care education and counseling.

REFERENCES

Abiko, T., Iijima, H., Koyama, A., Kamibayashi, N., & Narita, N. (2003). *Implementation and verification of care communications service between hospital-*

Figure 19. Relation between number of clients and throughput

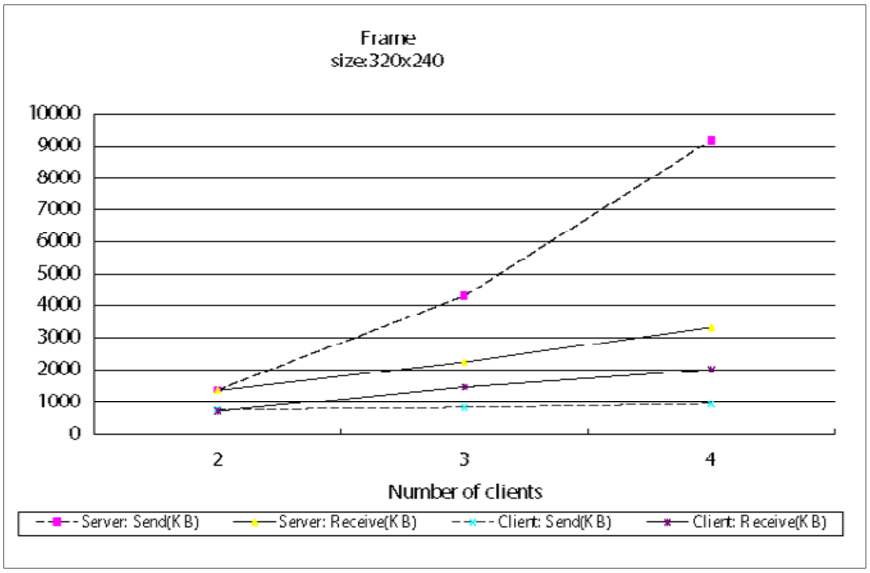


Table 5. Comparison of proposed and conventional systems

Feature / Systems	Network Environment	Implementation Platform	Hardware Software	Installation and Setup	Communication Service
T.Inoue et al., 1997	LAN Environment	Sun Workstation	Special Software Special Hardware	Needed	Mult-point to Multi-point
K.Kawamura et al., 2003	LAN Environment Leased Line	Windows	Special Software Special Hardware	Needed	Point to Point
Y.Kato et.al., 2003	Internet	Windows	Special Software	Needed	Broadcast
Sony, 2004	Internet	Special Hardware	Special Hardware	Need only Setup	Point to Point Mult-point to Multi-point Broadcast
Netmeeting	Internet	Windows	Special Software	Need Only Setup	Point to Point
Proposed WWW Conference System	Internet	Windows Macintosh Linux	WWW Browser	Automatically	Point to Point Mult-point to Multi-point Broadcast

ized patients and their families (IPSJ SIG Technical Report DPS114-6), pp. 37-43. (in Japanese)

Advanced Solutions. (2001). *moNet*. Retrieved from http://www. asi.co.jp/monet/index.html (in Japanese)

Advanced Solutions. (2002). *Impression live*. Retrieved from http://www.asi.co.jp/imlive/index.html (in Japanese)

Fuji-shi. (2005). *Fuji-shi live camera*. Retrieved from http://www.city.fuji.shizuoka.jp/live/

Hitachi Hybrid Network. (2004). *IP visual communication system NetCS*. Retrieved from http://www. hitachi- hybrid.co.jp/business/net/netcs01.html

Inoue, T., Okada, K., & Matsushita, Y. (1997). Spatial design for integration of face-to-face and video meetings: HERMES video conferencing system. *IEICE, J80-D-II*(9), 2482-2492. (in Japanese)

Kato, V., Jiang, D., & Hakozaki, K. (2003). A proposal of a streaming video system in best-effort networks using adaptive QoS control rules. In *Proceedings of IPSJ DPSWS'2003. IPSJ Symposium Series, 2003*(19) (pp. 7-12). (in Japanese)

Kawamura, N., Maita, Y., Hashimoto, K., & Shibata, Y. (2003). *Remote medical support system using the* *transcoding function* (IPSJ SIG Technical Report DPS 113-5), pp. 69-74. (in Japanese)

Kumamoto-shi (2001). *Today's Kumamoto Castle*. Retrieved from http://www.city.kumamoto.kumamoto.jp/ castleToday.html

Macromedia. (2003). *Flash Communication Server MX 1.5*. Retrieved from http://www.macromedia.com/software/ flashcom/

Macromedia. (2004). *Flash MX 2004*. Retrieved from http://www.macromedia.com/software/flash/

Maeda, K., Aibara, R., Kawamoto, K., Terauchi, M., Kohno, E., & Nishimura, K. (1997). Multimedia communication environment for distance learning. *IEICE*, *J80-B-I*(6), 348-354. (in Japanese)

Microsoft. (2004). *Netmeeting*. Retrieved from http://www.microsoft.com/windows/netmeeting/

Miyoshi, K., Okanaga, Y., Kou, S., & Kondo, S. (2000). Design, development, and experiments of distance learning systems by satellite Internet services. *IEICE, J83-D-I*(6), 644-650. (in Japanese)

Mori, K., Oyabu, Y., Nomura, A., & Oshita, S. (1992). Inter-campus tele-education and its evaluation. *IEICE, J75-A*(2), 244-255. (in Japanese)

NEC Engineering. (1996). *Remote conference system*. Retrieved from http://www.nec-eng.com/kaigi/index.html (in Japanese)

NTT West. (2004). Technical Report 12. Retrieved from http://www.ntt-west.co.jp/ipnet/ip/siryou.pdf (in Japanese)

Polycom. (2004). Retrieved from http://www.polycom.com/home

Sapporo-shi. (2004). *Tokeidai at the moment*. Retrieved from http://web.city.sapporo.jp/livecamera/

SONY. (2004). *Video conference system PCS-1*. Retrieved from http://bssc.sel.sony.com/Professional/markets/videoconferencing/overview.html

Sakiyama, T., Ohono, N., Mukunoki, M., & Ikeda, K. (2001). Video stream selection according to lecture context in remote lecture. *IEICE, 84-D-II*(2), 248-257. (in Japanese)

The House of Councilors. (2005). *The House of Councilors Internet TV*. Retrieved from http://www.webtv.sangiin.go.jp/webtv/index.php (in Japanese)

The House of Representatives. (1999). *The House of Representatives Internet TV*. Retrieved from http://www.shugiintv.go.jp/top.cfm

The Ministry of Public Management. (2004). *Home affairs, posts, and telecommunication*. Retrieved from http://www.soumu.go.jp/media/index.html

Visual Nexus. (2004). *Visual online conference system*. Retrieved from http://www.visualnexus.com/en/index.html

VTEL. (2004). Retrieved from http://www.vtel.com/

Wakahara, T. (1998). Configuration and characteristics of distance learning system over ATM-PVC Network. *IEICE, J81-B-I*(8), 494-506. (in Japanese).

Yamaguchi, T., Sakano, T., Fujii, T., Ando, Y., & Kitamura, M. (2001). Design of medical tele-consultation support system using super high definition imaging system. *IEICE, J84-D-II*(6), 1203-1212. (in Japanese)

This work was previously published in International Journal of Distance Education Technologies, Vol. 4, Issue 3, edited by S. Chang; T. K. Shih, pp. 77-96, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 4.7
ICT in Medical Education in Trinidad and Tobago

Marilyn Lewis
The University of the West Indies, Trinidad and Tobago

INTRODUCTION AND BACKGROUND

Information and communication technology (ICT) allows users to access information without taking geographic position into account. These users are also unconstrained by time, volume, or format of the information. ICT applications have enormous potential as a tool for aiding development in countries such as Trinidad and Tobago. Telemedicine, which can provide medical services to persons in isolated places, in emergencies, to the homebound, or the physically challenged, is but one example. Mansell and Wehn de Montalvo (1998) noted that "ICT applications facilitate telemedicine" (p. 85), and that "economic development can be fostered by tele-working and tele-services in some developing countries" (p. 83).

The twin-island nation of Trinidad and Tobago lies at the southern end of the Caribbean chain of islands, approximately seven miles off the northeast coast of Venezuela. The area covers 1,864 square miles (5,128 sq. km.), with a population of approximately 1.5 million. The economy of this small nation state is based mainly on petroleum and gas-based industries, but there is a growing service sector. PAHO figures (2002a, b) show a highly literate population with an overall adult literacy rate of 98.5% (males at 99.1% and females at 97.9%).

Transshipment and telecommunications facilities contribute to this country's position as the most industrialized in the Caribbean. The country's technical capacity and access to information have grown enormously in recent years. Telecommunication tools extend to the vast majority of the population. Per capita GDP stands at US$8,500.

There is a shortage of medical staff in general, with the ratio of doctors to inhabitants at 7.5 per 10,000. Shortages in primary health care are more acute than in other areas and have resulted in the employment of retired nurses and the recruitment of professional staff from other countries, particularly from Nigeria, India, and more recently, from Cuba. Trinidad and Tobago therefore stands poised to benefit from further development by fully embracing ICT, especially in the areas of education and medicine.

MEDICAL EDUCATION IN TRINIDAD AND TOBAGO

Medical education in Trinidad and Tobago engenders self-directed, lifelong learning through the use of the problem-based learning (PBL) method of teaching. The Faculty of Medical Sciences (FMS) of the University of the West Indies opened in St. Augustine, Trinidad and Tobago, in 1989, and has utilized PBL from its inception. Students' relative independence has been noted (Donner & Bickley, 1993) in students following PBL programs. Donner and Bickley noted that "they differ markedly from those following traditional medical programmes… [becoming] more skilled at an eclectic style of learning" (p. 297). These students show particular personal characteristics that encourage them to take a proactive role in their own learning, making them lifelong learners.

Research has also shown that PBL students make maximum use of library resources and that librarians taught the use of technology as a means of accessing, organizing, and managing information (Marshall, 1993). Library instruction is therefore a required part of the curriculum. Librarians become not just providers of books and other materials but also instructors in the use of modern technology. The library, therefore, prepares medical students for wider use of other applications and technologies to support their future information needs. This has implications for how these students will operate when faced with adverse conditions such as rural health offices and hospitals with limited resources, and for development in the community generally; these students in their homes, in their practices, and in the wider community will generate a multiplier effect.

INSTRUCTIONS IN THE USE OF MODERN TECHNOLOGIES IN THE MSL

From its inception, the Medical Sciences Library (MSL) has embarked on a program of information literacy for undergraduates and other categories of users. From as early as 1993, topics such as "MEDLINE: basic and advanced"; "International Pharmaceutical Abstracts (IPA)"; "MedCarib—health literature for the Caribbean"; "ProCite"; "Introduction to Computers"; and later, "EPI Info"; "Introduction to the Internet" and "PubMed" have been taught. In facilitating this training, the library equips its clientele with survival skills for the 21st century.

The Trinidad and Tobago Ministry of Health also recognized a need for retraining, because new demands were being placed on practitioners by health care transition, health care reforms, increased public and patient expectations, and advances in medical sciences and technology. The Ministry found that medical practitioners required additional skills. This was part of the rationale for the introduction in 2000 of a new postgraduate diploma in Primary Care and Family Medicine being offered by dual mode, face-to face initially, and thereafter, through distance education.

The library component of this course focused on skills such as "Locating and evaluating health information"; "Skills base for managing health information resources"; "Innovations in health information practice"; "Effective search and retrieval principles"; "MEDLINE on the Internet"; "Finding biomedical information on the Internet"; "Evaluating information resources"; and "Managing bibliographic references". Assessment tasks included:

- Joining and leaving an electronic discussion group

- Subscribing and unsubscribing to a mailing list
- Posting to a discussion group
- Locating an electronic serial and printing an article or abstract
- Executing a search on MEDLINE or PubMed and printing the results
- Creating a small database and generating a bibliography

Each session represented distinct skills requirements and supported the utilization of applications to manage the efficient exchange of information among health professionals.

ICT IN EDUCATION: PRIMARY, SECONDARY, AND TERTIARY

Primary and secondary schools in Trinidad and Tobago are also embracing the technology. Many secondary schools have computer science as a subject on the curriculum and typically have computer laboratories. More than 35% of the 78 primary and 120 secondary schools listed in the telephone directory for 2003–2004 have computers with Internet access facilitated by Telecommunication Services of Trinidad and Tobago (TSTT), the only telephone company on the twin islands. Additionally, there are 22 Internet cafes listed in the yellow pages of this directory. Some of these Internet cafes are located in rural areas such as Enterprise, in Central Trinidad, and Penal in the south of the island. Eighteen Internet Service Providers are listed as well.

Other initiatives to produce a computer literate society in Trinidad and Tobago include the government making computer loans available to all public servants. In 2002, the government also launched an initiative, the National Information and Communications Technology Plan (2004), that aims "to connect people, communities, business, government and educational institutions through an integrated technology network. It will also examine the policy, financial and skills development requirements that will be necessary to ensure sustainability and to ensure that the benefits of connectivity continue to grow, and accelerate, as new technologies, innovation and thinking emerge."

A survey (NIHERST, 2002) designed to provide empirical data on the penetration of ICT in private households reflects the varying penetration of ICT in private homes of varying socioeconomic status. Data were collected from a representative sample of 2,812 households throughout Trinidad and Tobago. Thirteen percent (13%) of the households in Trinidad and Tobago (approximately 44,600 households based on national statistics for 2000) had a home computer as of June 2001. By comparison, more than 30% of the households in a number of Organisation for Economic Co-operation and Development (OECD) countries were equipped with computers in 1997, and more than half (54%) of the households in Australia had computers in May 2000. Other important findings of the survey were that affordability was the major constraint in 56% of all households without computers; ranging from 43.9% in the City of Port of Spain to 78% in the Borough of Point Fortin. Also, 53% of households purchased computers from private savings; 13% accessed government loans. Households (20%) with gross monthly incomes of $6,000–$7,999 had the largest proportion of home computers, followed by 15% of households with incomes of $4,000–$5,999. Only 5% of households with monthly incomes of less than $2,000 had computers. In 2000, 27% of the computers were acquired compared with 6.7% in 1997. Almost three out of four persons (73%) in each household used the computer. The proportion of male (51%) to female (49%) computer users was generally similar. Approximately 16.6% of computer users were between 15 and 19 years of age, 16.3% between 30 and 39 years and 14.5% between 40 and 49 years. Of computer users, 50% had acquired secondary level education, and only 3.8% had a university level education in computer studies. Approximately 50% of computer users were employed and self-employed, and 39% were students. In private enterprises, 59.8% of employees used the computer compared with 29.7% in government. Windows 98 and 95 were the main operating systems in 74.4% of households. Most households (70.8%) used the computer daily between two and five or more hours. Only 11.8% of households were engaged in software development, and 20.2% accessed distance learning/education compared with other activities such as games (78.4%), Microsoft Office (66%), e-mail (62.4%), and Web searches (61.5%).

Apart from the Faculty of Medical Sciences, the university as a whole is also involved in programs to implement ICT applications. The UWI, St. Augustine, in its latest strategic plan (2003–2007) identified student-centeredness as one of its strategic objectives. The campus libraries, of which the MSL is a branch, determined that promoting the use of ICT in delivery of service and products was one way of meeting this objective. Additionally, an increasingly complex print and digital environment was emerging, and people began to expect certain services without coming to the library. Facilitated by the technology, the libraries were able to accomplish one of their main missions by launching a valuable new campus-wide online course, "Foundations of Information Literacy" on March 2, 2004. It is a modular course covering the following seven topics:

1. Basic computer literacy
2. Basic research skills
3. Using the OPAC (online public access catalogue)
4. The Internet as a research tool
5. West Indiana and Special Collections
6. Online databases
7. Managing references

There were tutorials on how to use each of the databases to which the libraries subscribe, as well as a quiz and online feedback, discussion board, and on-campus WebCT server.

In its quest to keep abreast of technology, the St. Augustine campus libraries is looking at further ways to implement applications in ICT in its ongoing program of work. Initiatives such as a digital reference service are also being considered. This project aims to answer students' reference queries in real time and will enable persons needing live human assistance while using the Web to immediately get the help they need from librarians who can quickly provide the answers. Speed and responsiveness are critical to this initiative, which will radically change the way we serve and support our clientele.

A TOOL FOR COMMUNITY DEVELOPMENT

Based on our epidemiological circumstances, where the Caribbean is second only to Southern Africa in the proliferation of HIV/AIDS and other STDs, ICT can indeed be seen as an effective tool for development. The toll of HIV/AIDS is heaviest on young persons between the ages of 18 and 45, who form the majority of the workforce. CAREC figures indicate that in Trinidad and Tobago, 5,000 persons have died of AIDS since 1985. The nature of the response to these diseases has to do with preventive education, capacity building, and treatment, care, and support.

ICT can play a role in the way people connect with information. Through ICT, people in remote areas can have access to the same information as many of the people in developed countries. With an infusion of ICT learning, people in underdeveloped countries such as Trinidad and Tobago can access and share critical health information that can help in the fight against HIV and AIDS, and its related morbidity and mortality. Ideally, health education should begin in primary schools. When this occurs, the transformative nature of ICT on the epidemiological status of affected countries can become visible.

CONCLUSION

The type of medical education available in Trinidad and Tobago, the campus libraries' program of instruction in the use of ICT, the government's initiatives to make computers available to public servants, the provision of computers in primary and secondary schools, the burgeoning economy, and the developing infrastructure all signal that this country is capable of embracing fully ICT applications that can enhance community development. Additionally, the postgraduate course has been viewed as an investment in the health care of Trinidad and Tobago nationals, in terms of an enhanced quality of health care that will be rendered by the practitioners involved. Healthy individuals contribute to a more productive economy. The course is ongoing, therefore, there will need to be an increment to those receiving training, as well as those receiving enhanced care. Three further cycles of the postgraduate course have been held, resulting in at least 60 primary care physicians who have already been exposed to and benefited from the postgraduate training provided by the MSL.

By teaching the core competencies of information literacy, identifying the information needs, accessing information, and understanding the legal and other issues in the use of information, the St. Augustine campus libraries aim first to increase the competitiveness of the student. However, not only have the libraries supported the academic enterprise, but they have also made available resources for knowledge creation and capacity building. At the Medical Sciences Library, we assisted in the retraining of primary care physicians to make them comfortable with the new technologies. This has resulted in more effective, efficient high-quality health care for the national community and has contributed to development.

Today's information consumer wants seamless access whenever and wherever they want it. They are comfortable with Web-based information. In this complex and rapidly changing, increasingly interconnected environment, ICT can be an effective tool for community development.

REFERENCES

Donner, R. S., & Bickley, H. (1993). Problem-based learning in American medical education: An overview. *Bulletin of the Medical Library Association, 81*(3), 294-298.

FOLDOC. (2004). Free online dictionary of computing. Retrieved September 27, 2004, from *http://foldoc.doc.ic. ac.uk/foldoc/index.html*

Mansell, R., & Wehn de Montalvo, U. (1998). *Knowledge societies information technology for sustainable development*. Oxford: United Nations Commission on Science and Technology for Development/Oxford UP.

Marshall, J. G., Fitzgerald, D., Busby, L., & Heaton, G. (1993). A study of library use in problem-based and traditional medical curricula. *Bulletin of the Medical Library Association, 81*(3), 299-305.

National Institute of Higher Education Research Science and Technology. (2002). *Utilisation of information technology by households, 2001*. NIHERST: Port of Spain.

Pan American Health Organization. (2002a). Country health profiles. Retrieved September 25, 2004, from *http://www.google.tt/search?q=cache:YxVT NV_AD20J: www.paho.org/english/HIA1998/Trinidad.pdf+ trinidad+ and+tobago+doctor+p atient+ratio&hl=en*

Pan American Health Organization. (2002b). Promoting health in the Americas. Retrieved September 22, 2004, from *http://www.paho.org/english/dd/ais/cp_780.htm*

[Trinidad and Tobago] Ministry of Public Administration and Information [home page].

(2004). *National Information and Communication Technology Plan*. Retrieved September 27, 2004, from *http://www.nict.gov.tt/plan/documents/summary.asp*

KEY TERMS

Digital Reference Service: A human-mediated, Internet-based service in which users' queries are answered in real time.

Discussion Group: Any system that supports group messaging, e.g., a shared mailbox, Usenet, bulletin board system, or possibly a mailing list, used to publish messages on some particular topic (FOLDOC, 2004).

Information Communications Technology (ICT): The study of the technology used to handle information and aid communication. The phrase was coined by Stevenson in his 1997 report to the UK government and promoted by the new National Curriculum documents for the UK in 2000 (FOLDOC, 2004).

Listserv: An automatic mailing list server, initially written to run under IBM's VM operating system by Eric Thomas. Listserv is a user name on some computers on BITNET/EARN which processes e-mail requests for addition to or deletion from mailing lists. Examples are listserv@ucsd.edu, listserver@nysernet.org. Some listservs provide other facilities such as retrieving files from archives and searching databases. Full details of available services can usually be obtained by sending a message with the word HELP in the subject and body to the listserv address. Eric Thomas has recently formed an international corporation, L-Soft, and has ported Listserv to a number of other platforms including Unix. Listserv has simultaneously been enhanced to use both the Internet and BITNET. Two other major mailing list processors, both of which run under Unix, are Majordomo, a freeware system, and Listproc, currently owned and developed by BITNET (FOLDOC, 2004).

MedCarib: A database of the health literature of the Caribbean.

MEDLINE: A CD-ROM database of medical literature in journals produced by the National Library of Medicine, United States.

Problem-Based Learning (PBL): A concept in which students focus from the beginning of their course on a series of real professional issues, where the knowledge of the various academic disciplines that relate to these issues is integrated.

PubMed: An online database of over 14 million citations of biomedical articles from 1950 to the present time. This database is available free over the Internet.

WebCT: Software that provides electronic learning in a flexible integrated environment.

This work was previously published in Encyclopedia of Developing Regional Communities with Information and Communication Technology, edited by S. Marshall, W. Taylor, X. Yu, pp. 382-386, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 4.8
Telehealth Organizational Implementation Guideline Issues:
A Canadian Perspective

Maryann Yeo
University of Calgary, Canada

Penny A. Jennett
University of Calgary, Canada

ABSTRACT

The current status of policies, guidelines and standards related to the organizational context of clinical telehealth practice were investigated. The directions these should take to meet the healthcare needs of Canadians also were outlined. An environmental scan approach was employed, consisting of a literature review, stakeholder survey questionnaire, and 12 key informant interviews. The literature review resulted in 260 sources related to organizational leadership issues, of which 176 were abstracted. The stakeholder survey questionnaire response rate was 64% (156/245), with 55% (84/154) completing the organizational context section. All (100%) key informants who were selected for interviews participated. Findings were categorized into four key organizational themes: organizational readiness, quality assurance, accountability, and continuity. Organizations need to review existing policies, standards, and guidelines in order to determine whether telehealth is covered and, if not, revise them or develop new telehealth-specific policies. Telehealth policies and procedures should be integrated with those in existence for face-to-face services.

INTRODUCTION

In Canada, there are 14 health jurisdictions; all are engaged in deploying telehealth applications. Telehealth, as defined in this project, is "the use of information and communications technology to deliver health and healthcare services and information over large and small distances" (Picot, 1998). Clinical telehealth applications are now operational in tertiary and community healthcare settings. As the number of telehealth projects, programs, and services has increased steadily, greater attention is being focused on policy and quality issues related to the delivery of telehealth services. There is increasing interest among healthcare professionals and administrators, healthcare institutions, organizations, businesses, government agencies, and regulatory bodies to develop and adopt policies, procedures, guidelines, and standards for use within provinces across Canada.

The National Initiative for Telehealth Guidelines (NIFTE) was established in order to develop consensus on a national, interdisciplinary framework of guidelines for use by health-sector organizations (National Initiative for Telehealth Guidelines, 2003). The guidelines were designed by telehealth providers for use by health professionals in developing their specific standards and as benchmarks for standards of service and by accrediting agencies in developing accreditation criteria. A major activity of the project was an Environmental Scan designed to examine four content areas related to telehealth: Organizational Context, Technology and Equipment, Clinical Standards and Outcomes, and Human Resources (National Initiative for Telehealth Guidelines Research Consortium, 2003).

The organizational context team investigated the status of policies, guidelines, and standards as they related to the organizational or administrative context of clinical telehealth practice in Canada. In addition, this component also explored the directions that telehealth administrative policies, guidelines, and standards should take in order to meet the healthcare needs of Canadians. The purpose of this article is to synthesize the findings of the environmental scan and to summarize the organizational issues and recommendations related to clinical telehealth implementation within organizations.

LITERATURE REVIEW

There is a recognized need for national standards for healthcare professionals and guidelines for the accreditation of healthcare organizations and facilities that provide telehealth services. This lack of standards and guidelines has been considered to be a barrier to the successful integration of telehealth into healthcare facilities. Standards are requirements that an organization must meet in order to earn accreditation and are important, because they provide a benchmark for measuring quality. At present, there are no existing Canadian telehealth accreditation standards. A variety of policies and guidelines were found in the literature review. Although many published papers, reports, and documents were reviewed, few provided insight into organizational policies, standards, or guidelines with respect to the provision of telehealth services nationally or internationally. The majority of the documents reviewed on the subject of organizational policies, guidelines, and standards for telehealth and telemedicine tended to focus on technical aspects. Standards need to be established for the administrative management of telehealth services. In addition, national standards need to be established for the management of privacy, confidentiality, and security, as well as for the documentation of policies and procedures.

The Advisory Committee on Health Infostructure (2001) asserted that several ingredients must be in place if the national health infostructure is to be implemented in an effective manner, including strong leadership; a clear and comprehensive strategy and detailed plan; and a common understanding of federal, provincial, and territorial initiatives. Jennett and Andruchuk (2001) stated that the successful implementation of telehealth services in Canada depends on several key factors: (1) the readiness of the environment; (2) systematic needs analyses, strategic business plans, and diverse, collaborative partnerships; (3) adequate equipment and IT vendors; (4) staged implementation; and (5) evaluation. Jennett and Siedlecki (2001), in a paper looking at the issue of policy development in telehealth, concluded the following:

The success of the telehealth venture is contingent on the development of policy to support and enable the use of technology in delivering quality care and equal access to stakeholders. A successful system is seamless, ubiquitous, and integrated. It supports a single point of access to stakeholders and incorporates clinical, administrative, educational, and transactional functions. The policy enabling this success ensures that privacy, confidentiality, and security are ensured; equality of appropriate access is provided; jurisdictional boundaries are not an impediment to optimal care; cultural diversity and human dignity are respected; stakeholder needs are met; and the telehealth service is timely, cost-effective, and patient-centered.(Jennett & Siedlecki, 2001, p.57–58)

There is a recognized need to develop telehealth policies, guidelines, and standards, but few policies, guidelines, or standards actually were found to exist in practice within Canada. Those policies, guidelines, and standards that do exist are specific to a particular organization, program, or project and are not integrated.

RESEARCH METHODOLOGY

An environmental scan approach was used to address the following research questions: (1) What is the status of policies, guidelines, and standards as they relate to the organizational or administrative context of clinical telehealth practice in Canada? (2) What direction should telehealth administrative policies, guidelines, and standards take to meet the healthcare needs of Canadians? As illustrated in Figure 1, the organizational context environmental scan framework methods included a literature review, stakeholder survey questionnaire, and key informant interviews.

Literature Review

The purpose of the literature review was to (1) identify key issues that need to be explored in the stakeholder survey, specific to the organizational context; (2) help to formulate the survey questions. The literature search was limited to English only, with an emphasis on Canada. The search of articles explicitly dealing with the organizational context (and similar keywords) of telehealth and policy issues, such as standards, guidelines, and protocols, resulted in 260 papers and documents being obtained and reviewed. Of these, 176 references met the review criteria (dealing with organizational context of telehealth and policy issues, such as standards, guidelines, and protocols) and were abstracted. The number of articles explicitly dealing with the organizational context of telehealth and policy issues was minimal. The majority of the literature obtained was related to the topics of organizational readiness, accountability, quality assurance, and continuity within the context of telehealth and/or health technology.

Stakeholder Survey Questionnaire

A mailed survey questionnaire was used to collect the more structured and quantitative information about the organizational context from telehealth stakeholders. Questions were asked about readiness for telehealth, leadership of telehealth programs, quality assurance, policies, processes, accessibility, and coordination of services. The semi-structured questionnaire was developed specifically for the environmental scan survey, using the knowledge gained from the literature review. Face and content validity were established by pilot testing the questionnaire with a sample of nine stakeholders that were representative of telehealth practice in Canada.

A total of 245 questionnaires were mailed to all the individuals listed in the Telehealth Stakeholder Database, a mailing list of individuals and organizations compiled and maintained by a Canadian Secretariat. To facilitate a quality response rate, an e-mail reminder (two weeks after initial mailing) and a second questionnaire mailing (two weeks after the e-mail reminder) were employed. Questionnaires were coded so that the second mailing could be targeted to non-responders. Of the 156 (64%) returned questionnaires, 84 (55%) completed the organizational leadership section of the survey. These 84 were coded and analyzed. Descriptive statistics were calculated using SPSS (Statistical Program for the Social Sciences) Version 11.0. The respondents' open-ended survey responses were transcribed and content analyzed by the research team. The largest portion (36%) of respondents completing the organizational leadership section were nurses (n=30), while administrators (n=18) accounted for 21%, physicians (n=11) accounted for 13%, allied healthcare professionals (n=6) accounted for 9%, and other professions (n=18) accounted for 21%.

Key Informant Interviews

The specific objectives of the 12 key informant interviews were to determine directions that telehealth polices, guidelines, and standards should take, and how they should evolve in order to meet the healthcare needs of Canadians. Organizational key informants were defined as individuals who frequently used telehealth (delivering or receiving telehealth services) within an organization in order to provide care or information related to care and who held management or administrative leadership positions within their organizations. The organizational key informants were selected by the researchers from individuals listed in the Telehealth Stakeholder Database to be representative of organizations offering clinical telehealth services across all Canadian provinces and territories. All (100%) key informants selected for interviews participated. Six physicians, three nurses, two administrators, and one psychologist participated in the interviews.

Semi-structured, open-ended questions were developed and asked during a telephone interview. The interviews were tape recorded with key informants' consents. The key informants were asked to provide their thoughts and opinions about what key organizational factors should be in place for the successful implementation of clinical telehealth services. The tape-recorded interviews were transcribed. The transcriptions and handwritten notes were summarized and analyzed independently by two researchers. Analysis was a multi-step process involving each researcher independently reviewing the transcriptions; becoming familiar with the content; and extracting significant comments, themes, phrases, or ideas. The researchers then met and discussed the independently reviewed transcripts in order to reach consensus on the common themes, phrases, or ideas.

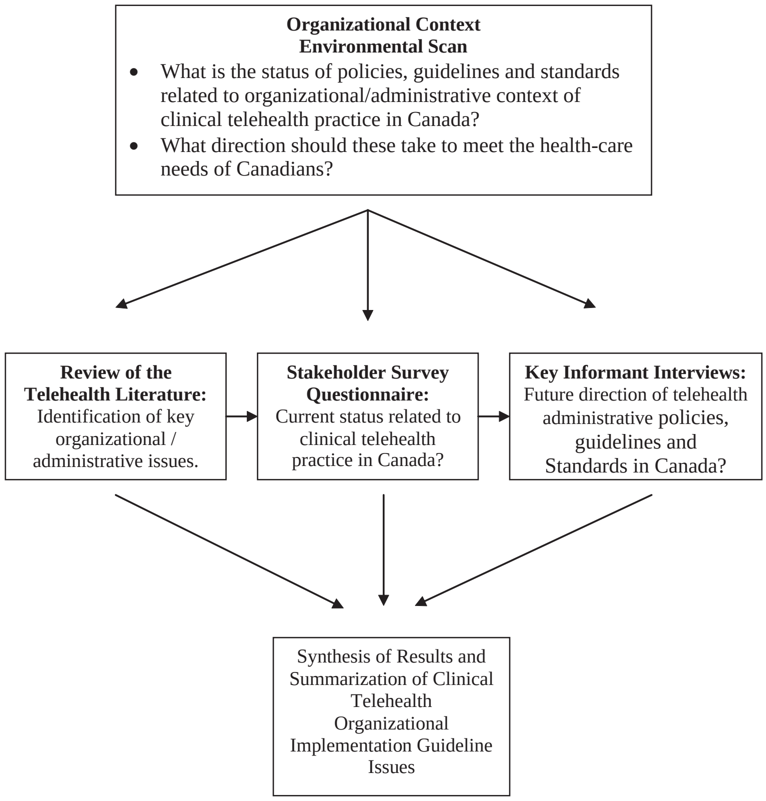
RESULTS

The following summaries synthesize and describe the four key organizational themes identified during the environmental scan: Organizational Telehealth Readiness, Accountability, Quality Assurance, and Continuity.

Organizational Telehealth Readiness

The need to systematically assess an organization's readiness to use telehealth technology was identified. Overall, the survey findings and key informants' responses support the literature, stressing the importance of organizational readiness in ensuring the long-term success of telehealth programs and services. Organizational readiness

Figure 1. Organizational context environmental scan framework



for telehealth is a multi-faceted concept consisting of planning, workplace environment, healthcare systems, and technical readiness.

Planning Readiness

An important component of an organization's telehealth readiness is planning readiness (Advisory Committee on Health Infostructure, 2001; Jennett & Andruchuk, 2001). Planning readiness involves organizations being able to demonstrate that the organizational leadership has completed required upfront work in terms of planning for the implementation and sustainability of telehealth (Alliance for Building Capacity, 2002). Planning readiness includes (1) a telehealth strategic plan, including a needs assessment and analysis, business plan, marketing plan (internal and external), communication plan, and evaluation plan; and (2) leadership, including program champions and collaborative partnerships. The dynamics of the planning process and order in which the steps in the planning process are conducted are very important. A phased approach to implementing the strategic plan was identified as important by the key informants.

A needs assessment and analysis was identified as an important factor in the successful implementation of telehealth. The majority (73%) of survey respondents reported that their healthcare organization or facility had conducted a needs assessment for its telehealth services. All 12 key informants (100%) noted that a needs analysis is a critical component of a healthcare organization's telehealth readiness and that there are several components to a needs analysis, including an assessment of community readiness and provider readiness. The rapidly changing status of telehealth services suggests that an almost ongoing approach to needs assessment should be in place. Organizational policies and guidelines should reflect this unique aspect of telehealth. A strategic business plan is an important component of organizational telehealth readiness. Indeed, the majority (83%) of survey respondents reported that their telehealth services were part of their organizational business plan, or that the business plan was in development. These findings are an indication that telehealth services are becoming recognized as a part of routine healthcare services. A telehealth strategic business plan needs to take into account research and evaluation of telehealth as an innovation. Given the rapidly changing nature of telehealth, the needs assessment and business plan must remain open to change and revision on a regular basis. Two subthemes—cost-benefit/cost effectiveness and financial readiness for sustainability—were identified. An indicator of sustainable funding would be that new investment dollars are put forth by the organization in order to address service gaps that use telehealth.

The survey findings indicated that the marketing of telehealth services is in an evolving phase. Only three of the 84 respondents reported that their organizations had a marketing plan in place or in development. Healthcare organizations currently are marketing primarily internally among healthcare providers within their institution or region or to their partners rather than to external audiences. A need was recognized for an open communication plan and for all stakeholders to be kept informed during the planning and implementation stages. It was felt to be critical for achieving buy-in and support for the planned programs, facilitating the change process, and team building. Since telehealth involves communication between organizations, a good working relationship between them is needed, as well as a sense that the receiving organization is a trusted authority.

Leadership Readiness

One of the key informants used the term *leadership readiness* to refer to the need to have program champions and collaborative partnerships in place and informed prior to beginning a telehealth program (Ash, 1997). Survey respondents (36%) indicated that there is a move toward the involvement of multiple advocates in senior positions or telehealth committees. Different types and levels of program champions and/or primary advocates are needed, based on their ability to influence. Within organizations, two types are needed: a clinical or provider champion and a senior-level administrative champion. At the community level, someone from within the community is required to promote what telehealth can provide (Alliance for Building Capacity, 2002). Senior management support is critical for the success and sustainability of telehealth services in healthcare organizations. All 12 key informants (100%) indicated that collaborative partnerships are an important component of organizational readiness. These partnerships should be in place before beginning and should include a variety of partnerships with different groups. The types of partnerships developed depend on the scope and application of the telehealth services. Partnerships need to be revised over time. Collaborative partnerships demonstrate that an organization has the ability to partner.

Workplace Environment Readiness

Findings also noted that organizations need to take into account the impact that telehealth services have on the workplace environment and work routines (Aas, 1999). The workplace environment should be ready to implement telehealth services (Aas, 2001). The identified components of workplace environment readiness include awareness of legislation; professional, and regulatory requirements; structural readiness; administrative support for telehealth; along with change management, human resources, healthcare system, and technical readiness. Healthcare organizations ready to use telehealth technology are aware of all legislation and professional regulations that may impact the delivery of telehealth services (e.g., health information protection laws that mandate policies and procedures to be in place in order to protect the privacy and confidentiality of sensitive information) as well as regulatory specific policies to determine telehealth coverage.

The workplace environment must be ready for the required telehealth technology and equipment (Kaplan, 1997). An organization with structure, capacity, means, and resources to implement telehealth is required because of the large amount of coordination needed to deliver and to receive requests for services. The workplace setting should have telehealth equipment in locations where it is convenient for providers and where it can facilitate the delivery of patient care (i.e., structural readiness). Support needs to be established for clinical decision making, functioning, and the process of using the telehealth system. These administrative support policies and procedures include mechanisms for the transfer of patients; a standardized, well-defined, easy-to use referral system; a standard and consistent method of recordkeeping at both the receiving and referring sites; and definitions for who gets privileges to use telehealth both at the receiving and referring sites. The introduction of telehealth services into an organization is often disruptive to the work environment (Aas, 1999; Ash, 1997). Readiness also means that the organization's leadership understands this impact and has a change management plan in place (Garside, 1998; Southon, Sauer & Dampney, 1997).

Human Resources Readiness

Human resources readiness involves having adequate and dedicated human resources in order to implement the strategic plan. Clarification of roles and responsibilities provides the required policies and procedures that are related to user characteristics, such as training and specialty. Six key informants (50%) stressed that an organization that is ready to use telehealth technology would have an education or learning plan in place for those individuals who provide care via telehealth.

Healthcare System Readiness

Healthcare system readiness for telehealth involves the healthcare system being ready in terms of providing strategic support and expertise and having policies, guidelines, and procedures in place at provincial/territorial and/or national ministry levels related to the required infrastructures, funding, remuneration, support for innovation, and diffusion processes (Olsson & Calltorp. 1998-1999).

Technical Readiness

Four key informants (33%) stated that the technical requirements and feasibility of providing telehealth services were essential components of organizational readiness with respect to such things as network and local site technical readiness and interoperability, technical feasibility, bandwidth, verification of fidelity of data transmission, and procedures to make sure that it is checked regularly. Technical support must be in place. Two key informants (17%) representing rural and remote telehealth services stressed that the technology support aspect of telehealth equipment requires consideration, particularly related to the maintaining of equipment in communities that do not have easy access to technology support services.

Quality Assurance

The literature review indicated a need to determine whether or not telehealth does improve the quality and accessibility of care (Jennett, Gao, Hebert & Hailey, 2000; Kangarloo, Dionisio, Sinha, Johnson, & Taira, 1999; Office of Health and the Information Highway, 2000; Yawn, 2000). All 12 key informants (100%) and survey respondents reported that telehealth quality assurance is important. Nearly all respondents (83%) indicated that their organizations were collecting some data or had such activities in development. Telehealth services evaluation is evolving; however, there is still a lack of sophistication in data collection. Evaluation of telehealth services currently focuses on the collection of system utilization, technical performance data, patient/client satisfaction, and provider satisfaction data. The more challenging areas related to the provision of telehealth services, such as access to care/services (Jennett, Person, Watson, & Watanabe, 2000), patient/client outcomes, cost-benefit/cost effectiveness, the impact on workflow, team relationships, provider roles, and responsibilities need to be evaluated more frequently. In addition, all aspects of the telehealth encounter should be evaluated, including a patient's physical and psychological comfort, technical quality of the service, quality of communication between caregivers and patients, duration of the consultation, timeliness of the care delivery, degree of confidentiality, and the costs (Hailey & Jennett, 2004; Sisk & Sanders, 1998). Informed, evidence-based evaluation of quality of care and cost effectiveness is still mainly in an early stage. However, the need to have baseline data to show that quality of care and patient safety are at least equal to face-to-face encounters is becoming recognized. The collection of both qualitative and quantitative data to evaluate telehealth services is needed (Lee, 1997). Four key informants (33%) indicated that telehealth services should be evaluated in a continuous and ongoing process. All key informants (100%) expressed that regular review of telehealth services should be part of an organization's overall evaluation process. It was suggested that they be done quarterly for the first year and then semi-annually.

The survey findings indicated that a wide range of organization or program-specific indicators and measures are being used to evaluate telehealth services. Standardization of data collection tools, common measures, and key indicators for telehealth applications are needed. Key performance indicators for telehealth should be defined in order to ensure consistency in how they are monitored, reported, and evaluated. Four key informants (33%) responded that standardized quality indicators, defined at both the program level and the organizational level, are needed related to system utilization, patient and provider satisfaction, and technical performance.

At the organizational level, a condition of accountability is the requirement of monitoring what is occurring through the development of an ongoing evaluation plan. Part of this accountability for quality assurance involves organizations having mechanisms in place in order to make the necessary adjustments to services, based on evaluation results.

Accountability

For each accountability issue, strengths and areas for improvement were identified. Key informants indicated that strengths can be demonstrated by the policies and procedures that organizations have in place and by the accountability mechanisms and structures present to address key requirements. Eight topics were identified: governance framework; privacy, confidentiality, security, and ethics; cultural awareness; document and storage of patient/client telehealth records; liability and risk management; licensure; cross-jurisdictional services and other liability issues; and remuneration/reimbursement.

Governance Framework

Organizations need to develop a governance framework that focuses on the roles and responsibilities of all individuals involved in telehealth activities (Industry Canada, 2001). Two issues related to the governance structure of the organization were identified:

1. Where telehealth is positioned in the organizational structure as well as the lines of accountability for telehealth services were reported to be important by both key informants and survey respondents. The person to whom telehealth is accountable should be at a high senior administrative level and positioned in the organization to make a strategic impact.
2. Administrative processes should be in place in order to support the governance structures and to help those individuals responsible for telehealth to assume accountability. Telehealth services ultimately should be accountable to a board or some other body that would provide governance with the appropriate administrative steps to support that structure: technical, personnel, supervisory, and managerial. To whom and for what telehealth is accountable needs to be defined and documented in writing. There are three levels of accountability: the health system itself, including health regions; the organization (administrative and clinical accountability); and the Privacy Commission (consent, confidentiality, and privacy).

Privacy, Confidentiality, Security, and Ethics

The literature indicated that privacy, confidentiality, and security must be considered when implementing telehealth services (Committee on Maintaining Privacy and Security in Health Care Applications of the National Information Infrastructure, 1997; Privacy Working Group, 2000). The practice of telehealth does not appear to raise any new policy issues regarding confidentiality. Traditional confidentiality policies should be applied in this new context with consideration for the specific technical issues involved in applying these polices for telehealth applications. In 2001, COACH, Canada's Health Informatics Association, published Guidelines for the Protection of Health Information (Canada's Health Informatics Association, 1997) as a framework to assist health organizations in the development and implementation of comprehensive privacy and security programs.

Protection of privacy requires organizational-level policies (Committee on Maintaining Privacy and Security in Health Care Applications of the National Information Infrastructure, 1997; Privacy Working Group, 2000). Most organizations have corporate privacy policies, and thus, 70% of questionnaire respondents and all key informants (100%) felt that a separate privacy policy specific to telehealth was not needed. However, existing privacy policies will have to accommodate the crossing of organizational and facility boundaries and the sharing of highly personal information. Organizational policies are required to govern who can be present during a videoconferencing session and who must declare themselves.

Confidentiality is an important issue that requires organizational-level policies around the protection of health information and the consenting process (American Medical Association, 1996; Privacy Working Group, 2000). Confidentiality is protected under the Health Information Act or similar provincial legislation, and the principles must be reflected in writing in the organization's policies (Canada's Health Informatics Association, 1997). The policies must reflect the impact of the human, organizational, and technical aspects of telehealth.

All key informants (100%) identified security as an important organizational policy issue, an area in which additional policies and standards related to telehealth were needed. Telehealth has some unique aspects that require special mention in the organization's security policies, guidelines, and procedures (Canada's Health Informatics Association, 1997). These policy and procedures issues are similar to those around electronic health records, and key informants believed that in the near future, these would be integrated. In addition, two informants (17%) stated that there should be an educational component in order to deal with the security issues related to human behavior. This was felt to be of particular importance in the remote regions of Canada due to the use of lay healthcare providers and lay telehealth coordinators.

The majority of survey respondents (65%) stated that their organizations did not have telehealth-specific codes of ethics and appeared to rely on their organizations' existing codes of ethics and/or legal guidelines. However, there are several ethical aspects related to telehealth concerning informed consent, protection of confidentiality, privacy and security, handling of confidential electronic information, and so forth that are unique to this method of delivering health services (Iserson, 2000). Respondents' suggestions on what should be included in a telehealth code of ethics support the need to review existing codes of ethics to determine if they include telehealth-specific ethical considerations. All 12 key informants and 13% of questionnaire respondents believed that the same ethical principles that apply in face-to-face care should be applied to telehealth encounters. Existing provincial/territorial policies, standards, guidelines, and procedures should be used. The issue of consent for healthcare via telehealth is one area that was identified as needing special consideration (Joint Interdisciplinary Telehealth Standards Working Group. 1998; Stanberry, 1998). Organizations should have procedures in place to have patients/clients give informed consent before they agree to the telehealth consultation. In addition, policies need to be in place that outline what information needs to be given to patients/clients for informed consent. Two key informants (17%) reported that special ethical needs must be taken into consideration in policies, standards, guidelines, and procedures when providing telehealth services to First Nations and Inuit communities in the more isolated and remote areas of Canada.

Documentation and Storage of Patient/ Client Telehealth Record

Although 56% of survey respondents reported that there was a protocol in place in their organizations or facilities for the documenting and storing of patient information for telehealth, the wide range of survey responses to the location of the patients' medical records (18% referring site; 18% consulting site; 24% copies at both sites; 12% other) indicated that there were no uniform standards that everyone could follow. Record-keeping and documentation for telehealth care was identified as another accountability issue that requires organizational policies to be in place so that what is going on can be tracked and so that two institutions can generate a record for the same patient and same encounter (Huston, 1999). With the coming age of the electronic health record, three key informants (25%) felt that standards should be developed in keeping with the total electronic communication environment.

Liability and Risk Management

The literature review indicated that there are several types of liability relating to the health professionals involved, the technology, the organization or institution, the human resources and training, and the telehealth application (Canadian Nurses Association, 2000; Saltzman, 1998). Organizations should have a written policy statement regarding whether healthcare providers need to be certified or not in order to potentially cover liability issues related to credentialing and privileging. It is recommended that established risk management policies be followed and that new telehealth-specific policies be created for anything not covered under existing policies (Switzer, 2001).

Licensure, Cross-Jurisdictional Services, and Other Liability Issues

Licensure was felt to be an important policy issue (Blum, 2000; Jacobson & Selvin, 2000). However, there was lack of agreement among the survey respondents and key informants as to whether or not it was necessary to require health professionals to be licensed in order to practice telehealth.

A number of policy issues need to be resolved with respect to the provision of cross-jurisdictional telehealth services (Pendrak & Ericson, 1996). Four key informants (33%) and 39% of survey respondents indicated that a common Canada-wide level of agreement should be in place with respect to cross-jurisdictional services. Also, policies regarding other cross-border liability issues need to be established at the national level by professional associations or regulatory bodies before healthcare organizations can put in place their organizational policies and procedures. Until national policies are developed, interim policies and agreements are required in order to address these issues.

Remuneration/Reimbursement

The issue of reimbursement, as it relates to physicians and other healthcare professionals in private practice, is unique to telehealth due to its borderless and cross-jurisdictional nature (Hogenbirk, Pong & Liboiron, 2001; Pong & Hogenbirk, 2000). All key informants (100%) felt that the development of policies in order to address reimbursement for health professionals paid on a fee-for-service basis is not an organizational-level issue and would have to be handled at the provincial level.

Continuity

Overall, the findings indicated that organizational leaders are just beginning to integrate telehealth into healthcare organizations and facilities as a service.

Integrated Telehealth Delivery Model

Organizations have to ensure that they do not develop policies and procedures that are different from existing policies and procedures for a regular visit. An integrated telehealth delivery model should be developed that positions telehealth as a strategic resource, which makes it possible to continuously improve the organization's capacity to deliver services and information across distances. One key informant indicated that a high accreditation rating should be given to an organization in which healthcare providers, during the course of the day, move easily from one medium to another as seamlessly as talking on the phone, going to a meeting, or sending an e-mail.

Administrative Interoperability

Telehealth services should be integrated into existing administrative policies, guidelines, and procedures. There are clearly the following well defined policies and guidelines that need to be in place for telehealth: technical interoperability; clinical practice, including liability; protection of privacy; freedom of information; and financial issues (Jennett & Siedlecki, 2001; Lemaire, 1998-2000). An integrated system of information and communication technologies rather than a focus on just videoconferencing is required in order to enable continuity of care, particularly in the remote areas of Canada. Scheduling of telehealth services is a unique administrative interoperability issue that requires standardization. Two key informants indicated that administrative policies should be put in place that allow all providers access to information about teleconsultations. As telehealth applications become more complex (e.g., homecare, preoperative care at a distance, regionalization of health facilities), some central coordination of telehealth communications will be required. A standard for coordination and linkages should be developed along with administrative structures to facilitate these standards. How do we make it work in a practical sense? How much is centralized? How much is a function of existing administrative structures, and how much needs to be new? Telehealth can provide coordination and linkages to facilitate the continuity of care without physically having to move patients. Telehealth can facilitate a different way of linking together virtually rather than physically, but the coordination has to come from the providers or organizations.

Coordination of Multiple Telehealth Services

Key informants suggested that organizations should look at how they currently deliver several services and deliver multiple telehealth services in the same manner and try as much as possible to integrate into routines that are already in place. The organization of outpatient services was proposed by one key informant as a useful model. A small-core telehealth group is needed to look after and manage the equipment and physical space. Macro-level organizational policies should be in place that are relevant to all telehealth applications (e.g., what has to happen in the event of a failure; maintenance standards; how booking is done; considerations of privacy, ethics, and informed consent). Coordination of multiple telehealth services should be the responsibility of a telehealth management committee, with site coordinators or network coordinators responsible for day-to-day operations. Setting priority criteria and guidelines for use of the services would be an administrative responsibility. The site coordinators would be coordinating the clinical services, using the developed criteria and guidelines.

Strategies and Policies to Ensure Sustainability

Business plans, regular reviews, and reliable, long-term funding strategies were identified as mechanisms around which policies and guidelines should be developed in order to ensure accountability and sustainability of telehealth services within organizations. One key informant stated that the extent to which an organization has demonstrated continuity would be evident by how much telehealth is in place. The majority of survey respondents (83%) and all 12 key informants (100%) recognized the importance of long-range business plans in terms of sustainability of telehealth services within an organization. However, they tended to believe that five years was too long of a period for a telehealth business plan to be in place, as it is difficult to do a long-range business plan with such a vibrant, growing technology. Program-related infrastructure supports need to be developed in order to ensure sustainability (e.g., bridging, scheduling, information repositories, telehealth information repositories, and directories). The budget for telehealth should be an integrated part of the complete budget, not an add-on. Long-term funding was identified by all key informants (100%) as being critical to the sustainability of telehealth services. A sustainability policy with funding to sustain it needs to be in place across Canada.

DISCUSSION AND CONCLUSION

This organizational environmental scan established that there are a number of organizational-level items that require consideration when developing policies, standards, and guidelines for telehealth services (see Table 1). These include the human, physical, and environmental infrastructures required for telehealth; organizational telehealth readiness; the business case specific to telehealth services; the governance framework for telehealth; provincial, territorial, national, and international policy-related activities; different policies, guidelines, and procedures for different types of telehealth activities; quality

Table 1. Summary of clinical telehealth organizational implementation guideline issues

Key Themes	Issues
General Organizational Issues	Concept of a Virtual Organization Integration of Telehealth Policies Telehealth-Specific Policy Issues Flexibility and Sensitivity to Innovation Multiple Types of Clinical Telehealth Applications and Technologies
Organizational Telehealth Readiness	Planning Readiness • Telehealth Strategic Plan • Needs Assessment and Analysis • Strategic Business Plan • Leadership Readiness • Evaluation Plan • Dynamics of Planning Workplace Environment Readiness • Awareness of Legislation, Professional and Regulatory Requirements • Structural Readiness • Administrative Support for Telehealth • Communication Plan • Change Management Readiness • Human Resources Readiness HealthCare System Readiness Technical Readiness
Quality Assurance	Ongoing Evaluation for Quality Assurance Key Performance Indicators
Accountability	Governance Framework Privacy, Confidentiality, Security, and Ethics Documentation and Storage of Patient/ Client Telehealth Records Liability and Risk Management Licensure Cross-Jurisdictional Services Remuneration/Reimbursement
Continuity	Integrated Telehealth Delivery Model Administrative Interoperability Coordination of Multiple Telehealth Services Strategies and Policies to Ensure Sustainability

assurance, including safety and required processes and indicators; and the integration of telehealth services with existing non-telehealth services. In addition, telehealth quality assurance, accountability, and continuity policies, standards, guidelines, and procedures should be integrated as much as possible with those in existence for face-to-face services.

These findings are a demonstration of the current status of telehealth in Canada and are reflective of the organizational implementation issues related to clinical applications in 2003.

The pan-Canadian funding and jurisdictional priorities and envelopes at the time of this work focused considerable attention on telehealth readiness and evaluation. Thus, it is not surprising that the majority of the findings relate to the areas of organizational readiness and accountability. Attention had not yet been directed to telehealth clinical outcomes, quality of care, or patient safety at the time of this study. This may be why less input was provided in these areas.

Telehealth is an innovation with rapidly changing characteristics and is a new alternative way of providing services. Therefore, organizations must move beyond site-specific focus to network considerations. The recommendations arising from these findings provided the groundwork for the development of the Organizational Context sections of the NIFTE framework of guidelines (National Initiative for Telehealth Guidelines, 2003). These sections are Organizational Readiness, Overarching Organizational Leadership Issues, Accountability, Quality Assurance, and Continuity. These are now being operationalized and considered in the organizational hospital accreditation processes.

ACKNOWLEDGMENTS

The Health Telematics Unit (HTU) expresses its gratitude to the Richard Ivey Foundation for funding. The HTU researchers appreciate the collaborative insights and efforts of the NIFTE Research Consortium members, the NIFTE Secretariat, and the NIFTE Advisory and Steering Committees. Appreciation is expressed to all stakeholders who participated in the survey questionnaire and key informant interviews. The research team also would also like to thank the HTU staff, researchers and student research assistants for their work on this project.

REFERENCES

Aas, I.H.M. (1999). Telemedicine and the organization of the health sector. *Journal of Telemedicine and Telecare, 5*(S1), 26–28.

Aas, I.H.M. (2001). A qualitative study of the organizational consequence of telemedicine. *Journal of Telemedicine and Telecare*, *7*(1), 18–26.

Advisory Committee on Health Infostructure. (2001). *Tactical plan for a pan-Canadian health infostructure: 2001 update. Office of health and the information highway.* Ottawa: Health Canada. Retrieved August 17, 2005, from http://www.hc-sc.gc.ca/ohih-bsi/pubs/2001_plan/plan_e.html

Alliance for Building Capacity. (2002). *Framework for rural and remote readiness in Telehealth*. Calgary: University of Calgary.

American Medical Association. (1996). *The promotion of quality telemedicine, Part II. CME resource guide*. Chicago: American Medical Association.

Ash, J. (1997). Organizational factors that influence information technology diffusion in academic health science centers. *Journal of the American Medical Informatics Association*, *4*(2), 102–109.

Blum, J.D. (2000, March). Telemedicine: The cyber physician and credentialing. *Health Care Law Monthly, 2*, 3–12.

Canada's Health Informatics Association. (1997). *Guidelines for the protection of health information*. Toronto: COACH.

Canadian Nurses Association. (2000). Telehealth: Great potential or risky terrain? *Nursing Now*, *9*, 1–4.

Committee on Maintaining Privacy and Security in Health Care Applications of the National Information Infrastructure. (1997). *For the record: Protecting electronic health information*. Washington, DC: National Academy Press.

Garside, P. (1998). Organizational context for quality: Lessons from the fields of organizational development and change management. *Quality in Health Care, 7*(supp.), 8–15.

Hailey, D., & Jennett, P. (2004). The need for economic evaluation of telemedicine to evolve: The experience in Alberta, Canada. *Telemedicine Journal and e-Health*, *10*(1), 71–76.

Hogenbirk, J.C., Pong, R.W., & Liboiron, L.J. (2001). *Fee-for-service reimbursement of tele-*

medicine services in Canada, 1999/2000 (Report prepared for the Office of Health and the Information Highway, Health Canada). Sudbury, Ontario: Laurentian University.

Huston, J.L. (1999). Telemedical record documentation. *Topics in Health Information Management*, *19*, 59–65.

Industry Canada. (2001). *Medical imaging technology. WG 5-Final Working Group Report. Ethical issues related to telehealth*. Ottawa: Industry Canada.

Iserson, K.V. (2000). Telemedicine: A proposal for an ethical code. *Cambridge Quarterly in Healthcare Ethics*, *9*(3), 404–406.

Jacobson, P.D., & Selvin, E. (2000). Licensing telemedicine: The need for a national system. *Telemedicine Journal and e-Health*, *6*(4), 429–439.

Jennett, P., & Siedlecki, B. (2001). Telehealth policy: Building a functional system. *Telehealth Law*, *1*(4), 53–58.

Jennett, P.A., & Andruchuk, K. (2001). Telehealth: "Real life" implementation issues. *Computer Methods and Programs in Biomedicine*, *64*(3), 169–174.

Jennett, P.A., Gao, M., Hebert, M.A., & Hailey, D. (2000). Cost-benefit evaluation of telehealth implementation implications for regions and communities. In *Proceedings of the 2000 International Conference on Information Technology in Community Health*, Victoria, British Columbia, Canada (pp. 146–153).

Jennett, P.A., Person, V.L., Watson, M., & Watanabe, M. (2000). Canadian experiences in telehealth: Equalizing access to quality care. *Telemedicine Journal and e-Health*, *6*(3), 367–371.

Joint Interdisciplinary Telehealth Standards Working Group. (1998). *Telemedicine creating virtual certainty out of remote possibilities* (Report of the Interdisciplinary Telehealth Standards Working Group for the Department of Human Services [State of Victoria]). Retrieved August 17, 2005, from http:www.dhs.vic.gov.au/ahs/archive/telemed/execsum.htm

Kangarloo, H., Dionisio, J.D.N., Sinha, U., Johnson, D., & Taira, R.K. (1999). Process models for telehealth: An industrial approach to quality management of distant medical practice. In *Proceedings of the 1999 American Medical Informatics Association Annual Symposium*, Los Angeles, California, USA.

Kaplan, B. (1997). Addressing organizational issues into the evaluation of medical systems. *Journal of the American Medical Informatics Association*, *4*(2), 94–101.

Lee, M. (1997). *Telehealth in Canada: Clinical networking, eliminating distances*. Ottawa: CANARIE.

Lemaire, E. (1998-2000). *Using communication technology to enhance rehabilitation services. A solution oriented manual*. Ottawa: Institute for Rehabilitation Research and Development.

National Initiative for Telehealth Guidelines. (2003). National Initiative for Telehealth (NIFTE) framework of guidelines. Ottawa: NIFTE. Retrieved August 17, 2005, from http://www.cst-sct.org/resources/FrameworkofGuidelines2003eng.pdf

National Initiative for Telehealth Guidelines Research Consortium. (2003). *Final report of the national initiative for telehealth (NIFTE) guidelines—Environmental scan of organizational, technology, clinical and human resources issues*. Ottawa: NIFTE.

Office of Health and the Information Highway. (2000). *Evaluating telehealth "solutions": A review and synthesis of the telehealth evaluation literature*. Ottawa: Health Canada.

Olsson, S., & Calltorp, J. (1998/1999). *Telemedicine: A tool for organizational and structural* *change in healthcare*. London: Kensington Publications.

Pendrak, R.F., & Ericson, R.P. (1996, December). Telemedicine and the law. *Healthcare Financial Management Magazine*, *50*, 46–49.

Picot, J. (1998). *Sector competitiveness frameworks series: Telehealth industry part 1—Overview and prospects*. Ottawa, Canada: Industry Canada.

Pong, R.W., & Hogenbirk, J.C. (2000). Reimbursing physicians for telehealth practice: Issues and policy options. *Health Law Review*, *9*(1), 3–12.

Privacy Working Group. (2000). *Principles for the privacy protection of personal health information in Canada*. Retrieved August 17, 2005, from http://www.cna-nurses.ca/CNA/documents/pdf/publications

Saltzman, K.M. (1998). Health care technology and the law. *Medical Group Management Journal*, *45*(4), 68–74.

Sisk, J.E., & Sanders, J.H. (1998). A proposed framework for economic evaluation of telemedicine. *Telemedicine Journal*, *4*(1), 31–37.

Southon, F.C., Sauer, C., & Dampney, C.N.G. (1997). Information technology in complex health services: Organizational impediments to successful technology transfer and diffusion. *Journal of the American Medical Informatics Association*, *4*(2), 112–124.

Stanberry, B. (1998). The legal and ethical aspects of telemedicine 4: Product liability and jurisdictional problems. *Journal of Telemedicine and Telecare*, *4*(3), 132–139.

Switzer, D.P. (2001). Network liability: A new frontier for healthcare risk management. *Journal of Healthcare Risk Management*, *21*(2), 3–13.

Yawn, B.P. (2000). Telemedicine: A new framework for evaluation. *Telemedicine Journal*, *6*(1), 55–61.

APPENDIX A. ORGANIZATIONAL FOCUS SURVEY QUESTIONNAIRE

The questions in this section ask about your organization's or facility's readiness for telehealth, leadership for your telehealth program, quality assurance, policies, processes, accessibility, and coordination of services.

1. Did your organization/facility conduct a needs assessment for its telehealth services?
   - ❑ Yes
   - ❑ No
   - ❑ Uncertain (***Please explain***):

2. ***If yes,*** when was this needs assessment done?
   - ❑ Less than one year ago
   - ❑ Between 1 year and 2 years
   - ❑ Between 2 years and 3 years
   - ❑ Greater than 3 years ago

3. Are your telehealth services a part of your organization's overall business plan?
   - ❑ Yes
   - ❑ To be considered in the near future
   - ❑ No
   - ❑ Uncertain (***Please explain***):

4. Do your telehealth services operate on a remuneration basis (e.g., cost-recovery or fee-for-service basis)?
   - ❑ Yes — ALL
   - ❑ Yes — PARTIALLY (***Please explain***):

     What is?

     What is not?
   - ❑ No
   - ❑ Uncertain (***Please explain***):

5. Do you market your telehealth services?
   - ❑ Yes
   - ❑ No
   - ❑ Uncertain (***Please explain***):

6. What is the location of your telehealth services (e.g., large hospital, small hospital, walk-in clinic, physician's office)?
   - ❑ On-site location (***Please specify***):
   - ❑ Off-site location (***Please specify***):
   - ❑ Other (***Please specify***):

7. Do your telehealth services have a program champion (i.e., primary advocate)?
- ❑ Yes ➔ **Go to question 8**
- ❑ No ➔ **Go to question 9**
- ❑ Uncertain (***Please explain***):

8. ***If yes,*** who acts as the telehealth program's primary advocate? What position does he or she hold in your organization/facility (e.g., physician, administrator, etc.)?
Please specify:

9.Do you collect data to evaluate your telehealth services?
- ❑ Yes — all services ➔ **Go to question 10**
- ❑ Yes — some services
- ❑ No ➔ **Go to question 11**
- ❑ Uncertain (***Please explain***):

10. ***If yes,*** how do you evaluate your telehealth services?
Please specify:

10a. Does your organization/facility collect data to determine the impact that telehealth services have on the following?

	Yes	In Development	No	Uncertain
The quality of care/service	❐	❐	❐	❐
System utilization	❐	❐	❐	❐
Patient/client outcomes	❐	❐	❐	❐
Cost of care/service vs. its benefit	❐	❐	❐	❐
Patient/client access to care/service	❐	❐	❐	❐
Provider's satisfaction with care/service	❐	❐	❐	❐
Patient's/client's satisfaction with care/service	❐	❐	❐	❐
Technical performance	❐	❐	❐	❐
System integration with established services	❐	❐	❐	❐
Workflow	❐	❐	❐	❐
Health team relationship	❐	❐	❐	❐
Provider roles and responsibilities	❐	❐	❐	❐
Other (***Please specify***): ________________________	❐	❐	❐	❐

11. Are there baseline data available so that the impact of telehealth services can be compared to data prior to telehealth?
- ❑ Yes
- ❑ No

❑ Uncertain (***Please explain***):

12. The following are some telehealth policy issues that have been identified in the literature. Does your organization/facility have written policies in place for the following areas as they relate to telehealth? ***Please check yes, no, or uncertain for each.***

	Yes	In Development	No	Uncertain
a) Reimbursement	❑	❑	❑	❑
b) Security	❑	❑	❑	❑
c) Privacy	❑	❑	❑	❑
d) Confidentiality	❑	❑	❑	❑
e) Liability	❑	❑	❑	❑
f) Cross-Jurisdictional Services	❑	❑	❑	❑
g) Ethics	❑	❑	❑	❑
h) Other (***Please specify***):				
______________________	❑	❑	❑	❑

13.Does your organization/facility or province/territory have a policy developed for dealing with cross-border licensure and other liability issues?

❑ Yes
❑ In development
❑ No
❑ Uncertain (***Please explain***):

Ethics, Privacy, Confidentiality and Security

15. In your organization, is there a telehealth-specific code of ethics for telehealth personnel?

❑ Yes
❑ In development
❑ No
❑ Uncertain (***Please explain***):

16. What telehealth-specific items would you suggest for inclusion in a code of ethics for telehealth personnel? (***Please specify)***

17. Are there policies regarding the role of non-clinical staff during confidential telehealth activities within your organization? For example, do you have a policy that requires non-medical staff to sign non-disclosure agreements?

❑ Yes
❑ No
❑ In Development
❑ Uncertain (***Please explain***):

18. Does your organization/facility conduct security assessments/audits of its telehealth services?
- ❑ Yes
- ❑ No
- ❑ Uncertain (***Please explain***):

19. Does your organization/facility have procedures in place for dealing with breaches of client/patient privacy or confidentiality?
- ❑ Yes
- ❑ In Development
- ❑ No
- ❑ Uncertain (***Please explain***):

20. Does your organization/facility have a risk management strategy in place for its telehealth services?
- ❑ Yes
- ❑ No
- ❑ Uncertain (***Please explain***):

21. In your organization, which of the following methods are used to ensure patient confidentiality for a telehealth encounter (***Choose as many as apply***):
- ❑ Physical security measures (e.g., locking away equipment when not in use)
- ❑ Confidentiality clauses in employees' contracts
- ❑ Training of employees regarding confidentiality
- ❑ Secure storage of telehealth medical records (e.g., passwords, encryption, etc.)
- ❑ Following the standards for health information privacy developed by the Canadian Institute for Health Information for maintaining patient confidentiality
- ❑ No protocols/procedures to ensure confidentiality for telehealth encounters
- ❑ Other (***Please specify***):

22. In your organization/practice, when is it necessary to obtain consent (verbal or written) for a telehealth encounter?
- ❑ Consent is implied by the patient arriving and participating in a telehealth encounter ➔ ***Go to question 26***
- ❑ Consent is necessary for all telehealth encounters
- ❑ Consent is necessary only under certain conditions

23. In your organization what type of consent is required for the following telehealth parameters? (***choose one of the following options for each parameter by placing a check mark (,) in the box***)

Telehealth Parameter	Verbal Consent Required	Written Consent Required	No Consent Required	N/A	Uncertain
Telehealth as a medium for service delivery					
Diagnosis and treatment recommended					

A permanent videotape is to be kept					
There is a surgical procedure involved					
The patient is part of a research study					

24. In your organization, what information is included in a written informed consent form? **(*Choose as many as apply*)**

- ❑ Potential risks, consequences, and benefits
- ❑ Description of procedures that will be followed
- ❑ Who has case responsibility and their obligations
- ❑ Who has access to patient information (privacy and security measures)
- ❑ Where patient health record will reside
- ❑ How and what information will be transmitted
- ❑ Whether session is to be photographed or videotaped
- ❑ No written consent form is obtained
- ❑ Other (***Please specify***):

25. In your organization, please indicate who is responsible for obtaining consent (verbal or written) for each of the following telehealth parameters. **(c*hoose one of the following options for each parameter by placing a check mark (,) in the box*)**

Telehealth Parameter	**Telehealth Coordinator**	**Referring Physician/ Health Professional**	**Consulting Physician/ Health Professional**	**Both Physician/ Health Professional**	**Other Staff**	**No Consent Required**	**N/A**	**Uncertain**
Telehealth as a medium for service delivery								
Diagnosis and treatment recommended								
A permanent videotape is to be kept								
Surgical procedure is involved								
The patient is part of a research study								

26. In your organization, is there a protocol for documenting and storing patient information for telehealth?

- ❑ Yes
- ❑ No ➔ **Go to question 28**
- ❑ In development
- ❑ Uncertain ➔ **Go to question 28**

27. In your organization, where is the patient's medical record stored, or where will it be stored for a telehealth encounter? (***Choose one***)

- ❑ The referring site (i.e., where the patient is located)
- ❑ The consulting site (i.e., where the consultant is located)

- ❑ Copies of the patient's medical record should be stored at both sites
- ❑ Other (***Please specify***):

28. Do you have additional observations and/or comments regarding the standards and/or guidelines pertaining to the organizational/leadership component of telehealth?

Thank you for your participation!

APPENDIX B. ORGANIZATIONAL FOCUS KEY INFORMANT INTERVIEW QUESTIONS

We are interested in your thoughts and opinions about what the key organizational factors are for the successful implementation of telehealth services. We will be addressing four content areas: (1) Organizational Readiness; (2) Accountability; (3) Quality Assurance; and (4) Continuity.

Organizational Readiness

➢ **Telehealth readiness is defined as "the degree to which users, health-care organizations, and the health system itself are prepared to participate and succeed with telehealth implementation" (The Alliance for Building Capacity, 2002, p. 2).**

1. In your view, what policy and procedures need to be in place to be able to say that an organization is **ready to use telehealth technology**?
2. What do you view to be **key components** of administrative or organizational leadership readiness?

Accountability

➢ **Being accountable is defined as being "responsible; required to answer for conduct, tasks, or activities. This responsibility may not be delegated" (Canadian Council on Health Services Accreditation, 2001, Glossary, p. 1).**
➢ **A number of content areas come under the Accountability heading, including liability and risk management; credentialing and privileging; ethics, governance, and leadership; privacy, confidentiality, and security.**

3. From your perspective, what **written policies/standards/protocols/guidelines** need to be in place in order to ensure accountability in an organization providing telehealth services?
4. What written policies **should be** in place for telehealth?

Quality Assurance

➢ **Quality is defined as "the degree of excellence; the extent to which an organization meets clients' needs and exceeds their expectations" (Canadian Council on Health Services Accreditation, 2001, Glossary, p. 15).**
➢ **Quality Assurance is defined as "strategies to verify that a product, process, or service conforms to and meets specifications or requirements" (Canadian Council on Health Services Ac-**

creditation, 2001, Glossary, p. 16).

5. In your opinion, what policies and procedures specific to quality assurance are needed for organizations that are providing telehealth services?
6. What are various **components of quality assurance** that would **require policies and procedures** (e.g., ongoing data collection)?
7. If yes, **what type of telehealth data** should be collected?

Continuity (coordination, linkages, and administrative interoperability)

➢ **Continuity is defined as "the provision of unbroken services that are coordinated within and across programs and organizations, as well as during the transition between levels of services, across the continuum, over time" (Canadian Council on Health Services Accreditation, 2001, Glossary, p. 5).**
➢ **The continuum is defined as "an integrated and seamless system of settings, services, service providers, and service levels to meet the needs of clients or defined populations" (Canadian Council on Health Services Accreditation, 2001, Glossary, p. 5).**

8. In your opinion, what are the **policies, guidelines, and procedures** that need to be in place in order **to enable coordination** of an organization's telehealth services?
9. If an organization offers **more than one telehealth service** (e.g., telepsychiatry, telehomecare, telelearning), **how should they be coordinated**?
10. In your opinion, what strategies, policies, or procedures need to be in place in order to ensure the

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 1, Issue 3, edited by J. Tan, pp. 24-46, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 4.9
Tele-Medicine:
Building Knowledge-Based Tele-Health Capability in New Zealand

Nabeel A. Y. Al-Qirim
Auckland University of Technology, New Zealand

ABSTRACT

This chapter reviews the strategic planning of health information systems in New Zealand. This step is deemed necessary to identify the main accelerators and/or impediments that influence technology adoption and diffusion in health organisations in New Zealand. This research introduces the tele-medicine technology as one possible solution to provide continuous, quality, and immediate medical care to rural patients and to encourage networking among the different hospitals in New Zealand. This research suggests that in order to realise tele-medicine benefits in health care delivery in New Zealand, certain issues need to be addressed such as implementing comprehensive cost-benefit analysis and identifying the benefits sought from adopting the tele-medicine technology. The New Zealand context is unique and this perspective with respect to tele-medicine adoption and success is addressed in this research.

INTRODUCTION

Information systems (NZHIS, 1995a; NZHIS, 1995b; NZHIS, 1996; Neame, 1995; Austin, Trim & Sobczak, 1995; Conrad & Shorttel, 1996), information technology (IT) (Bomba, Cooper & Miller, 1995) and technology (Little & Carland, 1991) have been emphasised as strategic tools for enhancing health care delivery and for improving performance, leading to optimised services and efficiencies. However, the New Zealand studies (NZHIS, 1995a; NZHIS, 1995b; NZHIS, 1996; Neame, 1995) indicate that the sector is "relatively devolved," with purchasing contracts being the main mechanisms to drive sector-wide change at the provider level. Much of the information needed is unavailable in the form needed or at the time that it is needed most. This in part is related to gaps in the conceptual understanding of service delivery, which is in this sector is a very complex business, spanning what has been an extensive range of relatively autonomous functional areas. But it is also due to a lack of reliable information about outcomes, effectiveness and actual costs on which improvements can be based. Because of this lack of empirical data, the tools for dealing with this complexity, and understanding what happens and why, are deficient (NZHIS, 1996). Various organisational issues and the lack of coordination at the national level were also identified. Expertise in health information management and systems is limited. Currently few health and disability sector personnel have the knowledge and skills to understand the issues or to make informed judgements about the validity of the advice they obtain.

New Zealand is not alone in this situation and different countries are facing similar difficulties although the severity of this situation varies from one country to another (Austin, 1992; Bakos & Tracy, 1986; Shortell, Morrison & Friedman, 1990; Topping & Hernandez, 1991; Conrad & Shortell, 1996). This literature points to different organisational, technological, and environmental impediments in adopting and in making use of IS/IT in organisations (Austin, 1992; Austin, Trimm & Sobczak, 1995; Ward, Griffiths & Whitmore, 1990). The 23 Health and Hospital Services (HHSs)[1] that exist all over New Zealand are no longer in competition with each other or paid in principal according to the number of people they care for. The competition from the few private hospitals has no effect on them. Ever-lacking government funds (Neame, 1995) are faced with further reduction on the medical portfolio by reducing or eliminating less priority and less life threatening services for the sake of introducing important new ones. The various gaps that exist between the various stakeholders and the lack of a leader (e.g., the government) to coordinate between the different HHSs result in having 23 different information systems that eventually do not interact with one another.

Diminishing funds from the government and cost control mechanisms have led to the need for alternative and more cost-effective means of providing care (Edelstein, 1999; Neame, 1995). In many cases, this has become necessary for survival (Edelstein, 1999) in order to sustain the increased competition among health care providers. The business of health care has become so competitive in different countries that many small rural hospitals are trying to align themselves with larger tertiary care centres in a community health-information network, a tele-medicine network, or some other type of partnership in order to survive and to retain their local patients (Huston & Huston, 2000). Within these challenges, tele-medicine emerge as one possible solution to New Zealand health providers in reaching out to rural patients (Charles, 2000; Harris, Donaldson & Campbell, 2001), to areas where patient volumes for certain services are limited (Edelstein, 1999), to conduct administrative and clinical meetings, and to conduct different training courses to patients (smoke treatment centres), doctors, nurses, and other medical staffs (Perednia & Allen, 1995; Wayman, 1994).

Tele-medicine means medicine from a distance where distant and dispersed patients are brought closer to their medical providers through the means of telecommunication technologies (Charles, 2000; OTA, 1995; Noring, 2000; Perednia & Allen, 1995; Wayman, 1994). Noring (2000) provided an interesting comparison between the former definition for tele-medicine and tele-health. This researcher defined the term tele-health as expanding the capacity of tele-medicine to provide the full continuum of care, from health promotion and disease prevention through curative treatment and terminal care. This term also implies including non-physician-based health care providers.

Tele-medicine covers a wide spectrum of benefits through the use of video conferencing (VC) technology in areas such as consultations, diagnostics, therapeutic, transfer of patient-related records, case management, training, and meetings. Researchers envision tele-medicine to be an important building block in the strategic plan of many health care organizations (Charles, 2000). In a rural setting, tele-medicine could help New Zealand health providers in supplying quality, fast, and economical medical services to rural patients and hence save doctors and patients valuable time wasted in commuting long distances. Specialists could utilise this extra time in seeing more patients at the main hospital.

Thus, this research is interested in introducing the tele-medicine technology as one of the strategic building blocks of Health and Disability Information Systems (HDISs) integration across the different health care providers in New Zealand. This research is interested in achieving the following objectives:

1. Highlight the status of tele-medicine in New Zealand.
2. Highlight the importance of tele-medicine to health care providers in New Zealand and elsewhere. Identifying the accelerators and/or the impediments that influence tele-medicine success could assist health care providers in New Zealand and elsewhere in planning their adoption and use of tele-medicine. This research utilises the technological innovation literature as a guiding theoretical framework in highlighting these factors.

Before expanding on the potential benefits of tele-medicine to health care providers, the following section examines the status of tele-medicine in New Zealand.

TELE-MEDICINE IN NEW ZEALAND

One of the early initiatives that emerged in 1993 within Northland HHS was transmitting radiology images between two hospitals using leased telephone lines. Tele-medicine has been investigated by most of the health providers in New Zealand. The entire crown-owned hospitals in New Zealand are managed by regional organisations known as Health and Hospital Services (HHS)[2]. Some HHS regions have one hospital and others have more than one. In the North Island, Waitemata Health has three video teleconferencing systems. Two of those systems are used in psychiatry and the third one is used in administrative and training purposes (ADMN). Health Waikato has two VC systems. Two of these systems are used in dermatology and the third one is used in ADMN. Northland Health Ltd (Whangarei) has two VC systems and another two PC-based VC. These systems are used in psychiatry and ADMN. Starship Children's Hospital (part of Auckland Healthcare Services) is in the process of establishing a national tele-paediatric network across the 23 HHSs and already started with a pilot project. Middlemore Hospital (South Auckland Health) has one group VC system for ADMN. Auckland University School of Medicine has one group VC system for training purposes and for conducting regular clinical meetings with other CHEs in New Zealand. Lakeland Health, LTD has two group video-teleconferencing systems for ADMN purposes.

In the South Island, Coast Health Care, LTD has three VC systems used in paediatrics, psychiatry, and ADMN.—HYPERLINK "canterburyhlth.html"—Canterbury Health, Ltd (Christchurch) has one video-conferencing system for paediatrics and for ADMN purposes. Healthlink South, LTD (Christchurch) has two group systems for psychiatry and ADMN. –HYPERLINK "hlthsthcant.html"—Health South Canterbury, Ltd (Timaru), Healthcare Otago, LTD, –HYPERLINK "nelsonmarlboro.html"—Nelson-Marlborough Health Services, LTD, and Southern Health, LTD each have one group system. Otago University in Dunedin is using it for training in their school of medicine. It is interesting to know that there is a level of cooperation among CHEs in the South Island in the areas of psychiatry and paediatrics.

A stock-take of the VC technology in these hospitals reveals that medical schools in New Zealand were among the early adopters and users of the technology. Out of the 23 HHSs in New Zealand, only 12 have actively adopted tele-medicine. The adopted systems ranged between one and four tele-medicine systems with the majority of HHSs adopting one system only. Those HHSs that adopted one tele-medicine system use it mostly for general purposes such as managerial meetings, case discussions and occasionally for clinical training. Such initiatives were described as being initial and experimental. Where an HHS owned more than one tele-medicine system, it was oriented for clinical purposes such as psychiatry, paediatric, and dermatology areas. Hence, an attempt is being made to adopt tele-medicine to provide prompt, inexpensive, and quality medical care to geographically dispersed patients, which was otherwise not possible.

TELE-MEDICINE: A BACKGROUND

The first tele-medicine initiative employing interactive television sessions for medical purposes emerged in 1959 by using a microwave link for tele-psychiatry consultations between the Nebraska Psychiatric Institute in Omaha and the state mental hospital 112 miles away (Perednia & Allen, 1995). In the late 1980s tele-medicine was being used routinely to deliver general health services to remote regions of Norway (Noring, 2000). In the US, interest in tele-medicine was initially focused on use in the military, in space programs, on off-shore oil rigs, in prisons, and in rural areas (Noring, 2000). It seems the opportunities provided by tele-medicine inspired many innovative ideas. For example, BMI British Midland has become the first airline to install tele-medical technology on planes making long-haul flights to the US Virgin Atlantic has also purchased the system and will start installing it in long-haul aircraft this year. The device monitors blood pressure via a wrist cuff, pulse rate, temperature via an ear probe, electrocardiogram, blood oxygen and carbon dioxide levels. Using a modem, this data is sent to physicians who can advise the crew on what action to take (Anonymous, 2002).

Since tele-medicine's inception in the 1950s, Perednia and Allen (1995) reported limited growth and pointed to the fact that only few tele-medicine projects were instituted in the 1970s and 1980s at several sites in North America and Australia. They confirmed that none of the programs begun before 1986 has survived. Although data is limited, the early reviews and evaluations of these programs suggest that the equipment was reasonably effective at transmitting the information needed for most clinical uses and that users were for the most part satisfied. However, when external sources of funding were withdrawn, the programs disappeared, indicating that the single most important cause of their failure was the inability to justify these programs on a cost-benefit basis. Other issues, such as limited physician acceptance, played a less significant role in their downfall (Perednia & Allen, 1995).

The views about tele-medicine effectiveness in the medical area vary from one adopter to another. Depending on one's viewpoint, tele-medicine can be seen as a valuable tool for providing immediate specialty care services to rural areas, a more efficient use of existing medical resources, a way to attract patients living outside a hospital's normal service area, a way of bringing international health care dollars. On the other hand, others could see it as a serious misallocation of increasingly scarce health care dollars (Perednia & Allen, 1995). This challenge needs to be resolved by highlighting the importance of the tele-medicine technology in a rural setting specifically.

In review of the literature it was observed that despite the rapid growth and high visibility of tele-medicine projects in health care (Grigsby & Allen, 1997), few patients are actually being seen through tele-medicine for medical purposes. In almost every tele-medicine project, tele-consultation accounts for less than 25% of the use of the system (Perednia & Allen, 1995). The majority of the online time is used for medical education and administration (Wayman, 1994; Perednia & Allen, 1995; Hassol, 1996). The low level of usage can be explained in part by the federal government's position on reimbursement for tele-medicine consultations (Hassol, 1996). However, this author did not find any association between reimbursement and tele-medicine utilisation in this research. Other issues need to be resolved first before the different benefits of tele-medicine can be realised. These unresolved important issues revolve around how successful tele-medicine can be in providing quality health care at an affordable cost and whether it is possible to develop a sustainable business model that could maintain profitability over time. This depends on (Perednia & Allen, 1995): (1) clinical expectations, (2) matching technology to medical needs, (3) economic factors like reimbursement, (4) legal (e.g., restrictions of medical practices across state lines (licensure) and issues of liabilities), and social (e.g., changing physician behaviours and traditional practices and workflow) issues (Anderson, 1997), and (5) organisational factors. These issues are discussed in the following sections.

TELE-MEDICINE ADOPTION: A THEORETICAL FRAMEWORK

Larsen (1998) states that information systems innovations are considered key enablers for business innovativeness. On the other hand, he asserted that the increased rate of failure among IS projects within organisations calls for an increased understanding about IS innovations (Moore & Benbasat, 1991). Innovation can be defined as an idea, practice or product that is perceived as new by the potential adopters even if it had existed earlier elsewhere (Rogers, 1995). The recent success and emergence of tele-medicine using the video conferencing (VC) technology in the early 1990s is an innovative approach of its adoption for medical purposes. In search for models that would explain technologies adopted by organizations, Rogers' (Rogers, 1983, 1995) model appeared to be the most widely accepted model by researchers in identifying critical characteristics for innovations (Moore & Benbasat, 1991, 1996; Premkumar & Roberts, 1999; Thong, 1999). Rogers classical adoption model is made up of the following factors: relative advantage, complexity, compatibility, observability, and trialability (*Table 1*). The results of using tele-medicine (observability) are quite observable to the different hospitals and have been highly publicised by the literature (Al-Qirim, 2002; Grigsby & Allen, 1997). Cost has been outlined as an important determinant of adoption by Rogers (1995) and other researchers (Bacon, 1992; Elliot, 1996; Tornatzky & Klein, 1982) (*Figure 1*).

In the following, the different factors that influence tele-medicine success are reviewed in the light of the technological innovation literature. Further identification of these factors that facilitate and/or hinder adoption and diffusion of tele-medicine could help HHSs and policymakers in New Zealand and elsewhere in adopting tele-medicine and in overcoming different barriers that could hinder its adoption.

Table 1. Innovation characteristics

Innovation Characteristics
1. Relative advantage: the degree to which using technology is perceived as being better than using its precursor of practices.
2. Complexity: the degree to which technology is perceived as being easy to use.
3. Compatibility: the degree to which using technology is perceived as being consistent with the existing values, and past experiences of the potential adopter
4. Trialability: the degree to which technology may be experimented with on a limited basis before adoption
5. Observability: the degree to which the results of using technology are observable to others.
6. Cost: the degree to which technology i s perceived as cost effective.

Relative Advantage

Table 2 identifies some of the stakeholders involved in tele-medicine adoption and depicts their perception about tele-medicine advantages in comparison with their earlier practices (Al-Qirim, 2002; Perednia & Allen, 1995; Wayman, 1994).

1. issues concerning patient's privacy;
2. specialists may reject extending their knowledge to rural doctors (profession protection) (Gammon, 1994);
3. some physicians prefer ambulatory services as a way to break a way from the hospital's stressful environment; and
4. some busy doctors tend to see tele-medicine as an extra burden (Gammon, 1994).

The most extensive use of tele-medicine has been to support continuing education over VC. Many more doctors and nurses are now able to attend conferences or classes because the direct (travel) and indirect (out-of-office) costs are decreased (Wayman, 1994). This raises concerns about the main objective(s) of tele-medicine and how it is perceived in the minds of the adopters. Is it a teaching utility or a patient care utility? Although the former leads eventually to better care and stable physicians at rural areas, the latter is envisioned to be the ultimate goal.

Cost Effectiveness

There will be no advantage in adopting tele-medicine if the investment is not justifiable in terms of

Table 2. Tele-medicine advantage

Stakeholder	Practices with telemedicine	Earlier practices	Remarks
Specialists	Seeing patients one-on-one through poor clarity video screen, ISDN connection keeps disconnecting	One-on-one, in person	Disadvantage
	Studio scheduling and moving to studio	No need, regular appointments in physician's office.	Disadvantage
	Second opinion from expert	No or not easy or cheap option	Advantage
	Case discussion and management	Offline, local	Advantage
	Immediate reporting from video	Immediate and accurate from patient	Disadvantage
	Could record session (legal & judicial reviews)		Advantage (1)
	Meeting rural doctors	Difficult	Advantage (2)
	Training rural doctors	Difficult	Advantage (2)
	Time optimisation (reduced ambulatory)	Longer time	Advantage (3,4)
	Less time waiting for rural patients	Longer waiting time	Advantage
	Fewer referrals to specialists	Increased	Advantage
	Reduced turnover of medical staff at rural areas—they feel less isolated from the outside world		Advantage
	Logistics:		
	Seeing rural patients over video reducing waste (time, traveling, lodging, etc.)	Travel to patient	Advantage (3)
	Lost supplementary income	Supplementary income	Disadvantage
Management	Meeting rural managerial staff, managerial training	Difficult	Advantage
Patients	See specialist over the video with the presence of rural medical staff	Commute to central specialist or wait for specialist to come to rural	Advantage

its acceptable financial returns. On the other hand, healthcare providers in general, including New Zealand, emphasise the importance of providing quality care and preserving human's well-being. Thus, making the cost element one factor in the adoption decision of tele-medicine, but not the most important one. On the other hand, hospitals have limited resources and hence, it is feared that hospitals would assess the tele-medicine technology as a non-priority medical technology (Al-Qirim, 2002), which emphasises the need to strike a balance between its financial feasibility and its advantages. Let us not forget that tele-medicine is an ideal technology in a rural setting, which makes the cost element insignificant to a certain degree for various reasons. Rural areas typically experience a severe shortage in specialist staff. This is due to vast rural areas to service, poor payment levels for this specialty, significant difficulty in recruitment and retention, and the professional isolation specialists experience in rural areas are some of the reasons (Al-Qirim, 2002; Charles, 2000; Harris et al., 2001).

The potential importance of tele-medicine in impacting the health and the welfare of rural communities is tremendous. The estimates of the percentage of tele-medicine patient candidates who can be retained by the rural hospital vary between 50% and 80% (Wayman, 19945). This patient retention not only has a direct, positive effect on the rural hospital and rural physician, but also on the area economy. The average dollar in a rural community circulates seven times; each patient trip to receive speciality consultation encounters costs in gas, lodging, and food, not to mention lost wages for the patient and assisting family/friends (Wayman, 1994). Early intervention has often proved to be less expensive in the long run, so access to specialty care via tele-medicine can ultimately result in decreased healthcare costs to rural areas (Huston & Huston, 2000).

Despite the ongoing evaluations of tele-medicine feasibility in healthcare and the different attempts to measure the costs and benefits of using the tele-medicine technology, there are still not enough data to provide accurate estimates of tele-medicine costs (Huston & Huston, 2000). Another important point to indicate here is that sometimes and according to the internal policy of certain hospitals the financial benefits are hidden or intangibles and difficult to calculate. Direct costs such as the cost of equipment, training, maintenance and ongoing support, upgrades, and ongoing telecommunications are quite obvious and indeed resemble the major costs involved in any tele-medicine project. In a recent research study in the US it was suggested that reimbursement and the cost of telecommunications were identified as the most important barriers of tele-medicine adoption (Grigsby & Allen, 1997).

Perednia and Allen (1995) indicate that current literature tackling tele-medicine has failed in addressing its cost-effectiveness on the basis of its use: as a diagnostic tool and/or as a therapeutic tool and/or case management tool. This further complicates the cost/benefit formula. The concern would be whether tele-medicine is used in the correct place initially and whether it is being used effectively in any of the above three areas. It is essential in the process of adopting certain tele-medicine technology to choose the right one that matches the hospital needs. Using tele-medicine to transmit video when the case necessitates that (i.e., psychiatry, physical and occupational therapy, orthopedics, some neurology applications, etc.), text when appropriate (i.e., lab reports, patient histories and physicals, dietary evaluations, insurance information, medication histories, etc.) using email, still images (i.e., X-rays, pathology slides, lab specimen slides, CTs, MRIs, lung scans and ultra sounds) using scanners or still digital camera or fax machines, and audio (i.e., electronic stethoscope) using telephone or email. It is of utmost importance to remember that most diseases do not move and hence, there is no need for real time and full motion VC equipment (Perednia & Allen, 1995; Wayman, 1994).

Resources available for medical care are limited and the use of expensive tele-medicine technologies will reduce the total number of tele-medicine sites that can be installed regionally or nationally. While use of the latest and most powerful equipment may seem logical, the early adoption of two-way, full-motion video may be unrealistic for many rural and under-served areas (Perednia & Allen, 1995). This is due to logistical reasons concerning the quality of telecommunication lines provided to rural areas (Charles, 2000).

Evidence suggests that other forms of cost control initiatives in the US such as the establishment of health maintenance organisations (HMO) and managed care have a significant effect in the adoption of more technologies and systems than others (Baker & Wheeler, 1998). In the US, most cost-effective applications are those that are paid for by insurers, such as the use of tele-medicine for radiology, prisoner healthcare[3], psychiatry, and home healthcare. Other applications enhance access to care but are not cost effective because third-party payers (e.g., HMO) do not pay for related costs for professional fees or the implementation of the technology (Charles, 2000; Huston & Huston, 2000). However, this situation does not apply to New Zealand because healthcare providers are paid ultimately by the government to provide healthcare services to patients in the regions they serve.

There are different ways to reduce the huge expenses in setting and running the tele-medicine sites and recent research suggests that some hospitals tend to promote other services through their tele-medicine network (i.e., like renting the facility to private entities and offices in rural areas to connect to their central offices) or encourage other health providers within and outside the hospital to join them and share the facility (Al-Qirim, 2002; Perednia & Allen, 1995).

Compatibility

Perhaps the greatest dilemma associated with all these changes is how to ensure that innovations in the process of delivering care are not achieved at the expense of sacrificing widely accepted values of what constitutes the humanity of care (Noring, 2000).

Tornatzky and Klein (1982) in their meta-analysis found compatibility features to be an important determinant of adoption. Rogers' (1995) compatibility characteristic is highly stressed here because past research in healthcare (Austin, 1992; Austin, Trimm & Sobczak, 1995) considered the problem relating to physicians accepting IS for clinical purposes. The use of computers and cameras suggests a shift away from the one-on-one personal interaction now viewed as essential and desirable in healthcare delivery (*Table 2*). Issues such as "Surgeon syndrome" which is conservatism towards all other technologies than one's own was highlighted as an impediment to the adoption of tele-medicine (Gammon, 1994). Some view it as an alternative to lack of care or treatment altogether; others view it as a dramatic quality and efficiency improvement to current levels of care (Wayman, 1994). On the other hand, these significant changes to the physician's and patient's treatment environment (i.e., one-to-one through the VC *Vs*. one-on-one and in person) are expected to create resistance for the tele-medicine technology (Al-Qirim, 2002). For example, recent research found that in an extreme case, one of the clinical staff in the rural area almost fainted when she saw herself in the television screen of the VC equipment (Al-Qirim, 2002). It is important that the changes introduced by tele-medicine be compatible with the hospital's value, practice, and experience to ensure its successful adoption. Issues such as gradual introduction and motivation are highly envisaged here.

The concern is that physicians are reluctant to accept the technology. Gammon (1994) suggested that doctors in general and specialists, specifically, would be the single most influential group affecting the diffusion process of tele-medicine. Gammon (1994) distinguished between two groups of physicians: expertise delivering physicians and expertise receiving physicians. The first group consists primarily of specialists at central hospitals who either, have ambulated (conducted outside treatment visits or operations) visits and enjoyed supplementary income[4] or have contributed significantly to hospital income by treating guest patients from neighboring municipalities. By carrying out distant consultations, they experience a reduction in income both for themselves and for their hospital. Furthermore, some physicians have shown a negative attitude toward allowing receiving (primarily rural) hospitals to participate (via VC) in medical meetings, while they are positive towards participating in similar meetings when they themselves are recipients. Physicians working in environments characterised by overload and stress will most likely perceive implementation of tele-medicine as an additional burden. This, coupled with tele-medicine's replacement of ambulatory, as well as a possible fall in income will muster their resistance to tele-medicine, if not absolute sabotage towards the implementation of tele-medicine (Gammon, 1994). Development of reasonable incentives as well as measures for integrating tele-medicine into the working environment will be an important step.

Security and legal issues are some of the impediments to the success of tele-medicine in New Zealand (*Table 2*). The security of patient's data and VC encounters are important issues and need to be addressed. Legal issues relating to operating a tele-medicine network including corporate practice of medicine, patient confidentiality and privacy, malpractice, informed consent, licensure and credentialing, intellectual property, funder's (Medicare and Medicaid) payment, fraud and abuse, medical device regulation, and antitrust (price fixing) (Edelstein, 1999) are some of the major issues, which could bring any tele-medicine project to a complete halt. However, countries such as the US are well ahead in addressing these issues (Edelstein, 1999). Edelstein (1999) reports that in one session of the American Congress, 22 pieces of legislation relating to tele-medicine were introduced. Four tele-medicine-related bills have been introduced in the 106th Congress.

Complexity

Complexity in the case of tele-medicine could be divided into two parts: Firstly, complexity relating to technical limitations in the VC equipment and bandwidth. Buying the wrong or ill-specified technology such as not presenting clear picture and voice or the different components in the VC equipment are not fully compatible with each other may result in frustrating the users. Selecting the right communication technology and bandwidth such as ISDN, ADSL, and Bridges is essential in order to have clear images and voice, file download/upload, and a valid and reliable connection, respectively. It should be noted here that installing the VC equipment needs a special room with specialised lighting and acoustic system. Secondly, user's complexity is related to the ease-of-use and to the simplicity of the VC equipment. Providing appropriate training courses earlier on could assist in removing lots of the misperceptions about the complexity of the equipment. This part should not be confused with the compatibility characteristic above, such as convincing the specialists moving from their offices or clinicians to the VC theatre. It could be argued here that the perceived complexity of the VC equipment could lead to incompatible perceptions amongst clinicians. Recent research examining the acceptance of the tele-medicine technology amongst physicians suggests that perceived usefulness was found to be a significant determinant of attitude and intention to adopting tele-medicine. However, the perceived ease of use was not significant (Hu, Chau, Sheng & Tam, 1999).

Trialability

Another important point here is the issue of trying and experimenting with the tele-medicine system before adopting it. Tele-medicine involves a considerable investment in buying the equipment and the training to know how to use it. Most importantly, it should be integrated into the hospital environment in general and in the clinical area specifically to witness its effectiveness. According to the reported challenges above, trying the system in the intended real setting could provide initial and vital signs, which could assist decision-makers in making accurate decisions concerning adopting or rejecting the technology. This makes the limited trialing of the system not sufficient to yield useful results. The issue here is whether the different HHSs are willing to make this involving step or whether the suppliers will accept lending their equipment for a considerable time to HHS to trial their system — and hence remains unanswered.

DISCUSSION AND CONCLUSIONS

If tele-health services are properly introduced and based on evidence of effectiveness, "… *telehealth has the capacity to improve the quality of healthcare, provide equity of access to healthcare services, and reduce the cost of delivering healthcare*" (Noring, 2000).

This research introduced the importance of adopting the tele-medicine technology by healthcare providers in New Zealand specifically. Tele-medicine represents a great opportunity for healthcare providers in New Zealand to network and to provide integrated healthcare and administrative services to rural areas specifically. This could provide important surrogates to the different gaps existing in the HDIS of the different HHSs in New Zealand. The lack of coordination and cooperation between the different HHSs in New Zealand needs to be resolved in order for this integration to succeed.

This research relied on the technological innovation literature to guide the development of the different factors that could accelerate or impede tele-medicine adoption. These factors are important to the strategic adoption and use of tele-medicine by the different HHSs in New Zealand and elsewhere. Tele-medicine introduces various clinical and administrative advantages as highlighted in this research and these advantages could be focused in delivering quality care and services to rural patients and physicians. However, as a precursor to adoption, it is important to identify the particular advantages sought from tele-medicine, taking into consideration the adoption context and the actual needs of the different HHSs. In a worst-case scenario, tele-medicine enables specialists in the main hospital give second opinions or allow for efficient follow-up and case management. If the costs involved in adopting and running the tele-medicine project were not planned well, the whole project could be brought to a complete standstill. Considering issues such as running costs and hidden costs needs to be identified earlier on. Issues pertaining to the complexity of the technology and to its compatibility with the physician's working environment needs to be addressed with more emphasis put on the latter as it could prove detrimental to the whole success of the tele-medicine project. Therefore, providing a framework where physicians are encouraged to accept and use the technology in providing healthcare services to patients and to other rural physicians is highly stressed here.

New Zealand's small area and population (3.82 million) (NZStat, 2001) could lessen the impact of many of the big challenges that hinder tele-medicine adoption in countries such as the US. Issues like licensure and reimbursement are major impediments in the US but not in New Zealand. There is one legal system in New Zealand and hence, interstate legalities and boundaries are large issues in the US but not in New Zealand. New Zealand has a sophisticated telecommunication (networks, mobile) infrastructure, which could serve the large-scale diffusion of tele-medicine across the different HHSs. This research demonstrated that tele-medicine provides various advantages and its potential in reducing costs at different levels has been established. It is important to further assess the strategic importance of tele-medicine within the different HHSs and at the national level in New Zealand. This task is quite achievable and requires the involvement of the different stakeholders highlighted in this research. Creating a national strategic plan aiming at identifying opportunities with respect to speciality-care, rural coverage and medical needs, and other administrative objectives could assist in driving the health sector forward and in providing fast and quality care to rural patients in the first place and to all New Zealanders eventually. Creating this integrated tele-medicine network amongst the different HHSs could prove viable to healthcare delivery in New Zealand.

REFERENCES

Al-Qirim, N. (2002). Enabling electronic medicine in New Zealand: The case of video conferencing adoption in KiwiCare. In F. Tan (Ed.), *Cases on Global IT Applications and Management,* (pp. 186-203). Hershey, PA: Idea Group Inc.

Anderson, J. (1997). Clearing the way for physicians: Use of clinical information systems. *Communication of the ACM, 40*(8), 83-90.

Anonymous. (2002). Telemedicine flying high. *Professional Engineering, 15*(9), 47.

Austin, C. (1992). *Information Systems for Health Services Administration*. Michigan: AUPHA Press/Health Administration Press.

Austin, C., Trimm, J. & Sobczak, P. (1995). Information systems and strategic management. *Healthcare Management Review*, *20*(3), 26-33.

Bacon, C. (1992, September). The use of decision criteria in selecting information systems/technology investments. *MIS Quarterly,* 369-386.

Bakos, J. & Treacy, M. (1986). Information technology and corporate strategy: A research perspective. *MIS Quarterly, 10*(2), 107-126.

Charles, B. (2000). Telemedicine can lower costs and improve access. *Healthcare Financial Management Association, 54*(4), 66-69.

Edelstein, S. (1999). Careful telemedicine planning limits costly liability exposure; *Healthcare Financial Management*, *53*(12), 63-69.

Elliot, S. (1996). Adoption and implementation of IT: An evaluation of the applicability of Western strategic models to Chinese firms. In K. Kautz & J. Pries-Heje (Eds.), *Diffusion and Adoption of Information Technology* (pp. 15-31). London: Chapman & Hall.

Harris, K., Donaldson, J. & Campbell, J. (2001). Introducing computer-based telemedicine in three rural Missouri countries. *Journal of End User Computing, 13*(4), 26-35.

Hu, P., Chau, P., Sheng, O. & Tam, K. (1999). Examining the technology acceptance model using physician acceptance of telemedicine technology. *Journal of Management Information Systems, 16*(2), 91-112.

Huston, T. & Huston, J. (2000). Is telemedicine a practical reality? *Association for Computing Machinery. Communications of the ACM, 43*(6), 91-95.

Josey, P. & Gustke, S. (1999). How to merge telemedicine with traditional clinical practice. *Nursing Management, 30*(4), 33-36.

Kwon, T. & Zmud, R. (1987). Unifying the fragmented models of information systems implementation. In R. Borland & R. Hirschheim (Eds.), *Critical Issues in Information System Research* (pp. 252-257). New York: John Wiley.

Larsen, T. (1998). Information systems innovation: A framework for research and practice. In T. Larsen & E. McGuire (Eds.), *Information Systems Innovation and Diffusion: Issues and Directions* (pp. 411-434). Hershey, PA: Idea Group Inc.

Little, D. & Carland, J. (1991). Bedside nursing information system: A competitive advantage. *Business Forum Winter*, 44-46.

Mathews, J. (2000). Satellites Filling Telemedicine Niche. *Aviation Week & Space Technology, 153*(18), 10.

Moore, G. & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research, 2*(3), 192-221.

Moore, G. & Benbasat, I. (1995). Integrating diffusion of innovations and theory of reasoned action models to predict utilisation of information technology by end-users. In K. Kautz & J. Pries-Heje (Eds.), *Diffusion and Adoption of Information Technology* (pp. 132-146). London: Chapman & Hall.

Neame, R. (1995). *Issues in Developing and Implementing a Health Information System*. Wellington: Ministry of Health.

Noring, S. (2000). Telemedicine and telehealth: Principles, policies, performance, and pitfalls. *American Journal of Public Health, 90*(8), 1322.

(NZHIS) New Zealand Health Information Service. (1995a). *Health Information Strategy for the Year 2000: Stocktake of Current Position and Future Plans.* Wellington: Ministry of Health.

(NZHIS) New Zealand Health Information Service. (1995b). *Health Information Strategy for the Year 2000: Gaps, Overlaps, and Issues Report.* Wellington: Ministry of Health.

(NZHIS) New Zealand Health Information Service. (1996). *Health Information Strategy for the Year 2000.* Wellington: Ministry of Health.

(NZStat) Statitics New Zealand. (2001). *A report on the Post-enumeration survey 2001.* Retrieved November 9, 2002 from the World Wide Web: www.stats.govt.nz/domino/external/pasfull/.

Office of Technology Assessment U.S Congress (OTA). (1995). *Bringing Health Care On Line: The Role of Information Technologies*, OTA-ITC-624. Washington, D.C: US Government Printing Office.

Perednia, D. & Allen, A. (1995, February). Telemedicine technology and clinical applications. *The Journal of the American Medical Association (JAMA), 273*(6), 483-488.

Premkumar, G. & Roberts, M. (1999). Adoption of new information technologies in rural small businesses. *The International Journal of Management Science (OMEGA*), *27*, 467-484.

Rogers, E. (1995). *Diffusion of Innovation*. New York: The Free Press.

Shortell, S., Morrison, E. & Friedman, B. (1990). *Strategic Choices for America's Hospitals.* San Francisco, CA: Jossey-Bass.

Thong, J. (1999). An integrated model of information systems adoption in small businesses. *Journal of Management Information Systems, 15*(4), 187-214.

Topping, S. & Hernandez, S. (1991). Health care strategy research, 1985-1990: A critical review. *Medical Care Review, 48*(1), 47-98.

Tornatzky, L. & Klein, K. (1982). Innovation characteristics and innovation adoption implementation: A meta-analysis of findings. *IEEE*

Transactions on Engineering Management, 29(11), 28-45.

Ward, J., Griffiths, P. & Whitmore, P. (1990). *Strategic Planning for Information Systems*. London: John Wiley & Sons.

Wayman, G. (1994). The maturing of telemedicine technology Part I. *Health Systems Review, 27*(5), 57-62.

ENDNOTES

[1] Recently, hospitals are now the provider arm of the District Health Boards, e.g., Waikato District Health Board (W-DHB).

[2] Recently, hospitals are now the provider arm of the District Health Boards, e.g., Waikato District Health Board (W-DHB).

[3] Avoiding the costs of transportation and additional security measures.

[4] However, it seems this has no adverse effect on New Zealand doctors as rural trips are considered part of their activities and hence, not paid any extra wage (Oakley et al., 2000).

This work was previously published in Creating Knowledge-Based Healthcare Organizations, edited by N. Wickramasinghe, J.N.D. Gupta & S. Sharma, pp. 193-206, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 4.10
Adopting and Implementing Telehealth in Canada

Penny A. Jennett
University of Calgary, Canada

Eldon R. Smith
University of Calgary, Canada

Mamoru Watanabe
University of Calgary, Canada

Sharlene Stayberg
Alberta Health and Wellness, Canada

INTRODUCTION

Canada spans 9,976,140 square kilometers and has an approximate population of 32 million people (Statistics Canada, 2001). More than 90% of Canada's geography is considered rural or remote (Government of Canada, 2001). Despite the highly dispersed population, and, indeed, because of it, Canada is committed to the idea that a networked telehealth system could provide better access and equity of care to Canadians. Growing evidence of the feasibility and affordability of telehealth applications substantiates Canada's responsibility to promote and to develop telehealth.

Telehealth is the use of advanced telecommunication technologies to exchange health information and provide healthcare services across geographic, time, social, and cultural barriers (Reid, 1996). According to a systematic review of telehealth projects in different countries (Jennett et al., 2003a, 2003b), specific telehealth applications have shown significant socioeconomic benefits to patients and families, healthcare providers, and the healthcare system. Implementing telehealth can impact the delivery of health services by increasing access, improving quality of care, and enhancing social support (Bashshur, Reardon, & Shannon, 2001; Jennett et al., 2003a). It also has the potential to impact skills training of the health workforce by increasing educational opportunities (Jennett et al., 2003a; Watanabe, Jennett, &

Watson, 1999). Therefore, telehealth has a strong potential to influence improved health outcomes in the population (Jennett et al., 2003a, 2003b).

Fourteen health jurisdictions—one federal, 10 provincial, and three territorial—are responsible for the policies and infrastructure associated with healthcare delivery in Canada. This article presents a telehealth case study in one of Canada's health jurisdictions—the province of Alberta. The rollout of telehealth in Alberta serves as an example of best practice. Significant milestones and lessons learned are presented. Progress toward the integration of the telehealth network into a wider province-wide health information network also is highlighted.

BACKGROUND

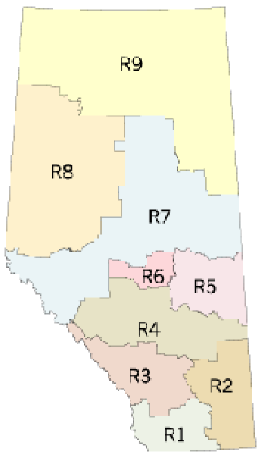
Canada's province of Alberta has a geography that is well suited to using telehealth technologies. Previous telehealth pilot projects throughout the province provided evidence of potential benefits from telehealth applications (Doze & Simpson, 1997; Jennett, Hall, Watanabe, & Morin, 1995; Watanabe, 1997). Alberta is the westernmost of Canada's prairie provinces with a total area of 661,188 km^2. Approximately 3 million people live in this western province, with two-thirds of the population living in two major cities in the lower half of the province; 19.1% of the population is distributed over the northern remote areas and southern rural communities (Figure 1) (Alberta Municipal Affairs, 2004; Statistics Canada, 2001).

In Alberta, health regions assume responsibility for acute care facilities and continuing and community-based care facilities, including public health programs and surveillance. The population sizes of health regions vary, as do their service census populations (Figure 1). The province has approximately 100 hospitals and more than 150 long-term care facilities. Health professionals are located largely in the urban centers, leaving many rural and remote areas with limited access to a variety of healthcare providers and services. In the mid-1990s, there was a physician-to-population ratio of 1:624 (Alberta Health, 1996). Most physicians are compensated on a fee-for-service basis by the provincial government (Alberta Health, 1997).

During the last decade, health reforms and restructuring have taken place in Canada, both at the federal (Kirby, 2002; National Forum on Health, 1997; Romanow, 2002) and provincial (Clair, 2000; Fyke, 2001; Mazankowski, 2001; Ontario Health Services Restructuring Commission, 2000) levels. These reforms were conducted in response to important trends challenging the Canadian Medicare system, such as escalating costs for new technologies and drugs, aging population, and increasing public expectations. Furthermore, health reforms addressed the issue of access to healthcare services for some groups, such as Aboriginal people and populations living in rural and remote parts of the country (Romanow, 2002).

Major challenges include the scarcity and isolation of healthcare professionals in many communities because of Alberta's large landmass; varied, extreme, and unpredictable climate; and population dispersion. Such realities were recognized as principal factors motivating the consideration of a provincial health information network. Alberta began to plan such a network in the mid-1990s, with the objectives of exploiting health information technologies and linking physicians, allied health professionals, hospitals, clinics, health organizations, and Alberta Health and Wellness (the provincial government's department of health) (Government of Alberta, 2003; Jennett, Kulas, Mok, & Watanabe, 1998). This network, entitled Alberta Wellnet, was a joint initiative by Alberta Health and Wellness and stakeholders in the health system. Initially, a core set of priority initiatives consisted of a pharmacy information network, telehealth, a healthcare provider office system, continuing and community care services, service

Figure 1. Health regions in Alberta (Government of Alberta, 2004)



Health Region	Region # on Map of Alberta	Population*	% of Alberta Population	# of Videoconference Sites
Chinook	R1	152,636	4.9	8
Palliser	R2	98,074	3.1	9
Calgary	R3	1,122,303	35.9	34
David Thompson	R4	286,211	9.2	27
East Central	R5	109,981	3.5	10
Capital Health	R6	978,048	31.3	28
Aspen	R7	176,580	5.7	18
Peace Country	R8	130,848	4.2	24
Northern Lights	R9	69,063	2.2	16
Total		3,123,744	100	174

event extract, population health and surveillance, and diagnostic services information sharing. This article describes the development of the provincial telehealth network in Alberta.

THE PROVINCIAL DEVELOPMENT OF THE TELEHEALTH NETWORK

Just prior to the establishment of Alberta Wellnet, an anonymous philanthropic foundation expressed interest in exploring the feasibility of developing an integrated, province-wide telehealth network based on need, capability and support. One of the authors (ES) initiated discussions with the foundation through an intermediary, and a small team of content experts was assembled to begin dialogue. The foundation requested a letter of intent and a presentation of concept to substantiate the benefits of allocating funds to telehealth. Following internal review and external consultation, the foundation allocated $525,000 for the preparation of a telehealth business plan. A project planning team was formed to serve as the province-wide planning group and to oversee the preparation of the business plan. This team incorporated representatives from the RHAs, the health sector, the universities, and the community. Two province-wide focus groups were held with representatives from the stakeholder groups to discuss content and approach, and a number of meetings were held between the CEOs (chief executive officers), their RHAs, and boards. The 17 regions and the two provincial boards at that time indicated their support in principle for a provincial telehealth network. The foundation's acceptance of the Provincial Telehealth Business Plan provided a commitment of up to $14 million to fund a provincial telehealth network pursuant to the satisfaction of a number of conditions (Table 1). The milestones associated with the development of the network are outlined in Table 2.

Specifically, the provincial telehealth network began its rollout in 1999 and continues to evolve and develop. A Provincial Telehealth Committee (PTC) comprised of health region, physician, and Alberta Health and Wellness representatives was formed in 1999 to provide strategic direction to this provincial initiative. Figure 1 outlines the current status of the network. In October 2004, there were 261 active telehealth systems that were networked and centrally connected through a network consisting of a combination of ISDN, SW56, and AGNPac. Applications were clinical, educational, and administrative in nature, and sessions were charted between 2002 and 2004

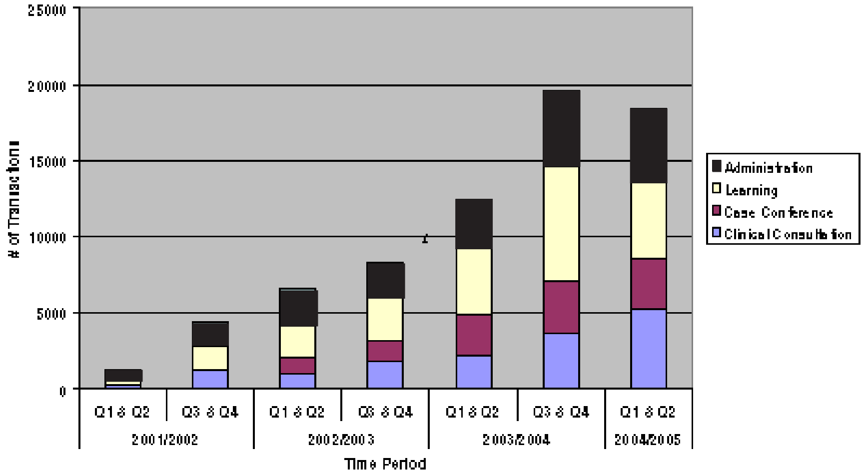
Table 1. Donor foundation conditions

- Funds to be used for capital purposes only
- 2 to 1 matching funds required ($7,000,000)
- Province-wide network (a minimum of 75% of the population, at least 12 Health Regions)
- Improve access (remove distance, especially for remote, isolated rural)
- Incorporation of all current pilot projects
- Sustainable, integrated into the health information system

Table 2. Milestones

1994	Health reform and restructuring in Alberta established 17 Regional Health Authorities (RHAs), down from >200.
1995	An anonymous philanthropic donor foundation expressed interest in exploring the feasibility of developing an integrated province-wide telehealth network based on need, feasibility, and support.
1996	Initial Business Plan supported the concept of a provincial telehealth plan. Potential allocation of $14 million, providing a number of conditions could be met.
1996	Initial plans for a provincial health information network (Alberta Wellnet) began.
1997	Telehealth Coordination (ATCC) and Implementation Planning Committees (ATIPT) formed with action plans regarding policy, infrastructure implementation, and operational solutions. Vision and overall management Report produced by the Telehealth Working Committee, with input from all stakeholders.
1997	$7 million one-time commitment, plus $1 million operating costs commitment from Alberta Health.
1997	Provincial Health Authorities' Association re-established a Charitable Foundation to receive and administer funds.
1997-8	The Alberta Medical Association's Committee on Fees and Alberta Health opened the way to potential payment for telehealth services.
1997-8	Telehealth handbook compiled to provide guidelines for Regions initiating telehealth activities. ATIPT merged with ATCC.
1998	Joint Planning Committee held with Alberta Telehealth Coordinating Committee (ATCC), the RHAs, and the Provincial Boards resulting in four priority telehealth applications: psychiatry; emergency care; radiology; and continuing education. Frameworks for three issues were discussed: 1) Needs assessment models, 2) Funding models, and 3) Provincial telehealth infrastructure and network.
1998	"Decision Document" produced for the Minister of Health, the RHAs, the anonymous philanthropic foundation, and the Alberta Medical Association. Business models for the four priority areas evolved, along with working groups assigned to each.
1999-00	Establishment of Provincial Telehealth Committee and coordinator.
1999-04	Central support services were initiated and deployed, including an interoperability standard testing framework for vendors, a scheduling system, followed by bridging and gatewaying services, and core evaluation and costing frameworks, processes, and tools.
2003	Alberta's RHAs reduced from 17 to 9 Health Regions.
2003	30 different clinical disciplines included in Telehealth.
2003-6	Provincial Telehealth Strategic Business Plan approved. Five critical success factors were articulated. A number of proposed initiatives were put in place.
2003-6	Telehealth Clinical Grant Fund established. Two calls for proposals have occurred, resulting in the funding of approximately 20 clinical projects and anticipated approval for an additional 20 projects.
2004	10,000th telehealth clinical consultation occurred.
2004	Clinical telehealth service extended to First Nations Communities.
2004	Telehealth scheduling system extended to the Territory of Nunavut.
2004	First provincial clinical telehealth forum held. Second one approved for 2005.

Figure 2. Growth of Telehealth—Increased administrative, learning, case conference, and clinical consultation transactions



(Figure 2). In 2004, the network's 10,000th clinical consultation occurred, and clinical services were extended to the First Nations' communities.

Since 2001, central supporting services have been established to support all stakeholders. For example, an Internet-based telehealth scheduling system initially was deployed in 2001 to facilitate coordination of telehealth activities throughout the province and has been used successfully to support approximately 30,000 telehealth events. The system now is being shared with several other jurisdictions (Nunavut, Saskatchewan, some mental health sites in British Columbia, Alberta First Nation's and Inuit Health Branch sites). Other core services included an interoperability standards testing site for vendors at Alberta Research Council, bridging and gatewaying services as well as costing models and core evaluation frameworks. For example, a home telehealth business case template was completed to facilitate health region analysis of this mode of service, while a newly developed site-based costing model provided support for further research and the development of a sustainable economic model for telehealth in Alberta. Further, a comprehensive provincial evaluation process is in development. To this end, an Inventory of Telehealth Evaluation Activities in Alberta was completed in order to establish a baseline and to inform refinements to the existing evaluation framework. Associated indicator standards are being aligned with the Alberta Health and Wellness Quality Framework. The Health Information Standards Committee of Alberta recently approved a new Alberta telehealth videoconferencing standard, as technological standards are continuously evolving.

A Provincial Telehealth Strategic Business Plan (2003-2006), developed by the PTC on behalf of health regions and approved by the Council of CEOs, articulates five critical success factors to realize the full potential of the Alberta Telehealth network. These five foci include (1) providing timely access to quality care, (2) collaborating effectively, (3) securing and effectively managing resources, (4) creating a culture that fosters change, and (5) formalizing the evaluation process and enhancing telehealth research.

At present, all of Alberta's Health Regions have telehealth programs in place. Due to required health system and provider buy-in rate, the clinical applications were slightly slower to evolve. As a consequence, a project that established incentives for the development of clinical telehealth applications was initiated. Alberta Health and Wellness funded a Telehealth Clinical Grant Fund (overseen by the PTC) resulting in more than 20 clinical telehealth projects, including cardiology, emergency, diabetes, urgent mental health, and pediatric services. A recent call for proposals is expected to result in funding of 20 additional projects, some of which will be extended to First Nations Communities.

SIGNIFICANT LESSONS LEARNED AND OBSERVATIONS

There were many lessons learned during the development of the Provincial Telehealth Program. They fall into two broad categories: (1) Business, Funding, and Sustainability (four items) and (2) Policy, Operational, and Infrastructure Issues (15 items) (Table 3). A number of key observations also emerged. These included the need to:

- Understand the nature and the implications of health reform and restructuring;
- Identify and clarify policy and operational issues around specific applications;
- Incorporate adequate telecommunications infrastructure;
- Respect diversity in priorities as well as preparedness and state of readiness among Health Regions;
- Ensure provision for sustainable operational support; support preparation of business models for shared telehealth needs in the Health Regions;
- Pay special ongoing attention to human factors, including the reengineering and business change management work place processes; and
- Involve end users, including physicians, nurses, and administrators, early in the planning process and conceptual design.

FUTURE TRENDS AND DIRECTIONS

Today, a major focus for the province is integrating the provincial telehealth program with other provincial health information initiatives. These include the Electronic Health Record Initiative, the Physician Office Support Program (POSP), and the Pharmacy Information Network (PIN). A second focus is on migrating from the current platforms to a high-speed, high-capacity IP broadband backbone (i.e., the SuperNet project) (Government of Alberta, 2002). In parallel, the province is continuously examining new evolving technologies and partnerships that can optimize the use of ICTs within the health sector, with a focus on quality of care, patient safety, and evidence-based decision making at the point of care.

CONCLUSION

The deployment of the provincial telehealth network continues to grow and to strengthen. Many milestones were encountered in the development of the Alberta Telehealth Network (Table 2). Several significant lessons learned also were documented (Table 3). While significant progress has been made, challenges and impediments remain for the creation of a functional, sustainable, and integrated province-wide telehealth network. Specifically, the successful integration of telehealth into a broad provincial health information and health system demands attention to human, system, and workplace factors. The significance of several critical factors cannot be ignored: different degrees of regional and sectoral readiness to adopt telehealth as well as diversity in selected telehealth priorities, potential cultural shift in the way providers work and learn, training needs,

Table 3. Lessons learned

Business, Funding, Sustainability

- The preparation of business models for shared telehealth needs in the Health Regions was helpful to ensure buy-in and progress.
- There is a great deal of work remaining after program initiation.
- Funding opportunities to cover the initial and ongoing operating costs of Telehealth require ongoing vigilance.
- Accountability for all network components required articulation.

Policy, Operational and Infrastructure Issues

- Raising awareness of telehealth and its potential continues to be challenging, as telehealth encompasses a variety of areas and technical complexity, which is difficult to reduce to simple language.
- A number of policy and operational issues around specific applications were identified and required further study and clarification.
- Cross-jurisdictional policy work is extremely time consuming. For example, the number of editions of the consent policy framework was extremely high.
- Respect for diversity of “state of readiness”, priorities, and partners/collaborators among Regions was required.
- A representative governance structure was required. All stakeholders needed to participate in the planning and decision-making during the early stages of rollout.
- A “decision document” for the Minister of Health, the Health Authority, the anonymous philanthropic foundation, and the Alberta Medical Association was a tool to enhance successful integration and buy-in.
- A clear understanding of the nature and implications of health reform and restructuring are required before telehealth can be integrated into a regional health system. Telehealth must align with and support major health reform initiatives to continue to be supported by and successful in the provincial funding structure. It must also take into consideration the convergence of technologies as it evolves.
- Alberta Health, the Ministry of Health, was a key player in initial successful implementation.
- Consideration of the province’s overall health information approach was important. Multiple reporting requirements are needed.
- In early stages, policy and integration committees were helpful to move agendas along. Commitment was central as expertise, time, as well as expenses were volunteered.
- Continual revisiting of critical issues with RHAs and Provincial Boards, along with ongoing active participation, was required. Perseverance in planning and development is rewarded.
- Adequate infrastructure (telecommunications, trained staff, etc.) for a network was critical.
- Identification of “pressure points,” along with provincial barriers and obstacles was difficult, but essential to move the agenda along.
- Implementation requires close attention to human, system, and workplace factors.
- Ongoing evaluation and monitoring is required for continuous quality improvement.

and required public/private/academic sector collaborations.

It also was observed that telehealth and e-health initiatives make substantial impacts in key areas of health reform. Integration requires close attention to the ongoing healthcare restructuring movements as well as to the convergence of technologies. There is a tendency to simply add technology to existing processes; however, some of these processes need to be reengineered in order to realize the full potential of telehealth. Further, changing funding processes to better accommodate telehealth is extremely difficult; the amount and quality of evaluation research on the impact of investments in telehealth must be increased to support the decision makers.

The provincial telehealth network initiative provides ongoing progress reports as health reform and restructuring continue and as integration with other provincial health information initiatives continues.

ACKNOWLEDGMENTS

The authors wish to acknowledge the commitment and many volunteer hours of members on the various committees, planning teams, and working groups associated with the rollout of the Alberta Telehealth Initiative. The authors acknowledge the assistance of Winston Teo, Heidi Brandstadt, and the Health Telematics Unit in the preparation of this manuscript. The authors also wish to acknowledge both the foresight and generosity of the anonymous donor foundation that served as the catalyst for this initiative.

REFERENCES

Alberta Health. (1996). *Inventory of health workforce in Alberta, 1995*. Calgary, Alberta: Alberta Health.

Alberta Health. (1997). *Health care '97—A guide to health care in Alberta* (2nd ed.). Calgary, Alberta: Alberta Health.

Alberta Municipal Affairs. (2004). *2004 official population list*. Retrieved February 2, 2005, from http://www.municipalaffairs.gov.ab.ca/ms/pdf/2004Pop.pdf

Bashshur, R. L., Reardon, T. G., & Shannon, G. W. (2001). Telemedicine: A new health care delivery system. *Annual Review of Public Health*, *21*, 613-637.

Beolchi, L. (Ed.). (2003). *European telemedicine glossary of concepts, standards, technologies, and users* (5th ed.). Brussels, Belgium: European Commission, Information Society Directorate-General.

Clair, M. (2000). *Les solutions émergentes—Rapport et recommandations.* Québec: Gouvernement du Québec.

Commission of the European Communities. (2004). *E-health—Making healthcare better for European citizens: An action plan for a European e-health area*. Brussels, Belgium: COM (2004) 356 final.

Doze, S., & Simpson, J. (1997). *Evaluation of a telepsychiatry pilot project*. Edmonton, Alberta: Provincial Mental Health Advisory Board, Alberta Heritage Foundation for Medical Research.

Eysenbach, G. (2001). What is e-health? *Journal of Medical Internet Research*, *3*(2), e20. Retrieved February 18, 2005, from http://www.jmir.org/2001/2/e20

Fyke, K.J. (2001). Caring for Medicare—Sustaining a quality health system. Saskatchewan: Saskatchewan Commission on Medicare.

Government of Alberta. (2002). *Alberta SuperNet: Restructuring and government efficiency.* Retrieved February 7, 2005, from http://www.albertasupernet.ca

Government of Alberta. (2003). *Alberta Wellnet*. Retrieved February 2, 2005, from http://www.albertawellnet.org

Government of Alberta. (2004). *Health and wellness: Health regions*. Retrieved February 11, 2005, from http://www.health.gov.ab.ca/regions/map_lookup.htm

Government of Canada. (2001). *Annual report to Parliament 2000-2001: Enhancing the quality of life for rural Canadians*. Canadian Rural Partnership. Retrieved February 7, 2005, from http://www.rural.gc.ca/annualreport/2001/report_e.phtml

Jennett, P., et al. (2003a). *Socio-economic impact of telehealth: Evidence now for healthcare in the future* (Vol. 1). Retrieved February 7, 2005, from http://www.fp .ucalgary.ca/telehealth/AHFMR_State_of_the_ Science_Review.htm

Jennett, P. A., et al. (2003b). *Socio-economic impact of telehealth: Evidence now for health care in the future final report* (Vol. Two) Retrieved February 7, 2005, from http://www.fp.ucalgary.ca/telehealth/AHFMR_ State_of_the_Science_Review.htm

Jennett, P. A., Hall, W. G., Watanabe, M., & Morin, J. E. (1995). Evaluation of a distance consulting service based on interactive video and integrated computerized technology. *Journal of Telemedicine and Telecare*, *1*, 69-78.

Jennett, P. A., Kulas, D. P., Mok, D. C., & Watanabe, M. (1998). *Telehealth: A timely technology to facilitate health decision making and clinical service support*. In J. Tan & S. Sheps (Eds.), *Health decision support systems* (pp. 353-369). Gaithersburg, MD: Aspen Publishers.

Kirby, M. (2002). *The health of canadians—The federal role*. *Final report (Volume 6): Recommendations for reform*. Ottawa: Standing Senate Committee On Social Affairs, Science and Technology, Government of Canada.

Mazankowski, D. (2001). *A framework for reform*. *Report of the Alberta Premier's Advisory Council on Health*. Edmonton, Alberta: Preier's Advisory Council on Health. Retrieved May 5, 2006, from http://www.health.gov.ab.ca/resources/publications/PACH_report_final.pdf/

National Forum on Health. (1997). *Canada health action: Building on the legacy* (Vol. 1), Final Report. Ottawa: Health Canada.

Ontario Health Services Restructuring Commission. (2000*). Looking back looking forward: A legacy report*. Quebec: Quebec Ministry of Health and Social Services.

Quebec Ministry of Health and Social Services. (2001). *Report of the Ministerial Table on Telehealth of the Quebec Ministry of Health and Social Services.*

Reid, J. (1996). *A telemedicine primer: Understanding the issues*. Billings, MT: Artcraft Printers.

Romanow, R. (2002). *Commission on the future of health care in Canada*. *Building on values: The future of health care in Canada—Final report*. Ottawa: Government of Canada.

Statistics Canada. (2001). *Population counts, for Canada, provinces and territories, and census divisions by urban and rural, 2001 census—100% data*. Retrieved February 2, 2005, from http://www12.statcan.ca/english/census01/products/standard/popdwell/Table-UR-D.cfm?PR=48

Watanabe, M. (1997). *Towards a province-wide partnership in telehealth: Report of the telehealth working committee*. Edmonton, Alberta: Alberta Health and Wellness.

Watanabe, M., Jennett, P., & Watson, M. (1999). The effect of information technology on the physician workforce and health care in isolated communities: the Canadian picture. *Journal of Telemedicine and Telecare*, *5*(2), S11-19.

KEY TERMS

AGNPac: AGNPac is a government of Alberta IP network.

Bandwidth: The range of frequencies transmittable through a transmission channel. If the channel is unique, the band is called baseband. If the channel is made multiple through a process of multiplexing, the band becomes wideband and can support data, voice, and video at the same time. (1) The difference between the highest and the lowest frequencies in a data communication channel. (2) The capacity of a channel. (3) A measure of the amount of information and, hence, its transfer speed, which can be carried by a signal. In the digital domain, bandwidth refers to the data rate of the system (e.g., 45 Mbit/sec) (Beolchi, 2003). Bandwidth is a practical limit to the size, cost, and capacity of a telehealth service.

Broadband: A popular way to move large amounts of voice, data, and video. Broadband technology lets different networks coexist on a single piece of heavy-duty wiring. It isolates signal as a radio does; each one vibrates at a different frequency as it moves down the line. Its opposite is baseband, which separates signals by sending them at timed intervals (Beolchi, 2003). Future networks, like those being deployed in Alberta, will carry these higher speed communications (i.e., Broadband ISDN).

E-Health: An emerging field in the intersection of medical informatics, public health, and business, which refers to health services and information delivered or enhanced through the Internet and related technologies. In a broader sense, the term characterizes not only a technical development but also a state of mind, a way of thinking, an attitude, and a commitment for networked, global thinking to improve healthcare locally, regionally, and worldwide by using information and communication technology (Eysenbach, 2001). The Commission of the European Communities (2004) declares that e-health tools or solutions include products, systems, and services that go beyond simply Internet-based applications and include tools for both health authorities and professionals as well as personalized health systems for patients and citizens.

IP: Internet Protocol is the network layer protocol for the Internet (Beolchi, 2003).

ISDN: Integrated Services Digital Network is a set of international digital telephone switching standards that can be used to transmit voice, data, and video. It offers the advantages of error-free connections, fast call setup times, predictable performance, and faster data transmission than possible when using modems over traditional analogue telephone networks. Basic ISDN services combine voice and data, while broadband services add video and higher-speed data transmission. ISDN offers end-to-end digital connectivity (Beolchi, 2003).

SW56: Switch56 is digital telecommunication that transmits narrow-bandwidth digital data, voice, and video signals.

Telehealth: Encompasses the use of advanced telecommunications technologies to exchange health information and to provide healthcare services across geographic, time, social, and cultural barriers (Reid, 1996). Telehealth refers to all healthcare or social services, preventative or curative, delivered at a distance via a telecommunication, including audiovisual exchanges for information, education and research purposes, as well as the process and exchange of clinical and administrative data (Quebec Ministry of Health and Social Services, 2001). Homemade systems or systems used outside of official health information networks are included in the latter definition.

This work was previously published in Encyclopedia of Digital Government, edited by A. Anttiroiko and M. Malkia, pp. 26-33, copyright 2007 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 4.11
IS Implementation in the UK Health Sector

Stuart J. Barnes
Victoria University of Wellington, New Zealand

INTRODUCTION

Since the mid-1980s, the UK public sector has been the subject of wide-ranging reforms involving the introduction of IS and IT. Change has been sought in the ways that services are managed and delivered, the evaluation of the quality of aforesaid services, and in accountability and costing. One of the most predominant of such changes has been the introduction of competition for services, the motivation of which has been to invite efficiency, effectiveness, and related benefits ensuing from the accrual of economies.

Pivotal to such change has been an explosion in the introduction of a variety of information systems to meet such challenges. Focusing on health care, a large part of the work of the health service involves collecting and handling information, from lists of people in the population to medical records (including images such as X-ray pictures), to prescriptions, letters, staffing rosters and huge numbers of administrative forms. Yet until recently, the health service has been woefully backward in its use of the technology to handle information by the standards of private industry.

This has been quickly changing in recent years and by 2003 the National Health Service (NHS) spent £2.8 billion annually on capital in hospitals (Department of Health, 2003a), around 10% of which was for IT. In the last 20 years, IT has added 2% to overall health expenditure (Wanless, 2001). This investment is still small by the standards of the private sector, but is all the more significant when we consider that health care is an industry which has been slow to adopt IT and one which presents some of the biggest IT opportunities (Department of Health, 2002).

BACKGROUND: INFORMATION SYSTEMS IN THE NHS

The implementation of IT in the UK health sector has been fraught with difficulties. In fact, estimates suggest that problems with the first wave of projects in the public sector, from the mid-1980s to mid-1990s, cost over £5 billion (Collins, 1994).

There are a number of high-profile examples of IS failure in the NHS, including that of Wessex Regional Health Authority's Regional Information Systems Plan (£63 million), the London Ambulance Service's Computer Aided Dispatch system (£1.1 to £1.5 million), and more recently, various Resource Management Initiative (RMI) Case Mix failures (£1 to £3 million) (Barnes & Targett, 1999).

In this section we will focus on RMI and Case Mix as an illustration, since this was one of the first initiatives for wide reaching IT-induced organizational transformation and information integration in the health sector. The systems implemented have been an important support for financial developments in the management of UK hospitals. During 2003, the importance of Case Mix systems has again come to the fore as the NHS has attempted to implement prospective payment systems (PPS) (Department of Health, 2003b).

RMI was a driving force in the move towards information systems and cultural change in the NHS. First announced in 1986, RMI was going to help clinicians and other hospital managers to make better-informed judgments surrounding how the resources they control can be used most effectively (DHSS, 1986). The Initiative was not only aimed at persuading clinicians to own the management process, but to provide them with accurate, up-to-date and relevant information which could be used to cost medical activities and improve patient care. The response to this need for improved information services available to hospital units was the development and implementation of a sophisticated and extensive package of IT referred to as the "Case Mix" information system (CMIS), an idea borrowed from the U.S., with the purpose of clinical and management audit.

Prior to RMI, the introduction of IT in the NHS was patchy and limited. Where systems existed, the technology was very varied, incompatible, archaic, and dependent upon regional computer departments to deliver necessary operational systems. The development of CMIS, with its dependence on data fed from other systems such as the Patient Administration System (PAS), radiology, pathology, theatre and nursing systems, provided a catalyst for the adoption of operational systems throughout the hospital (Barnes, 2001).

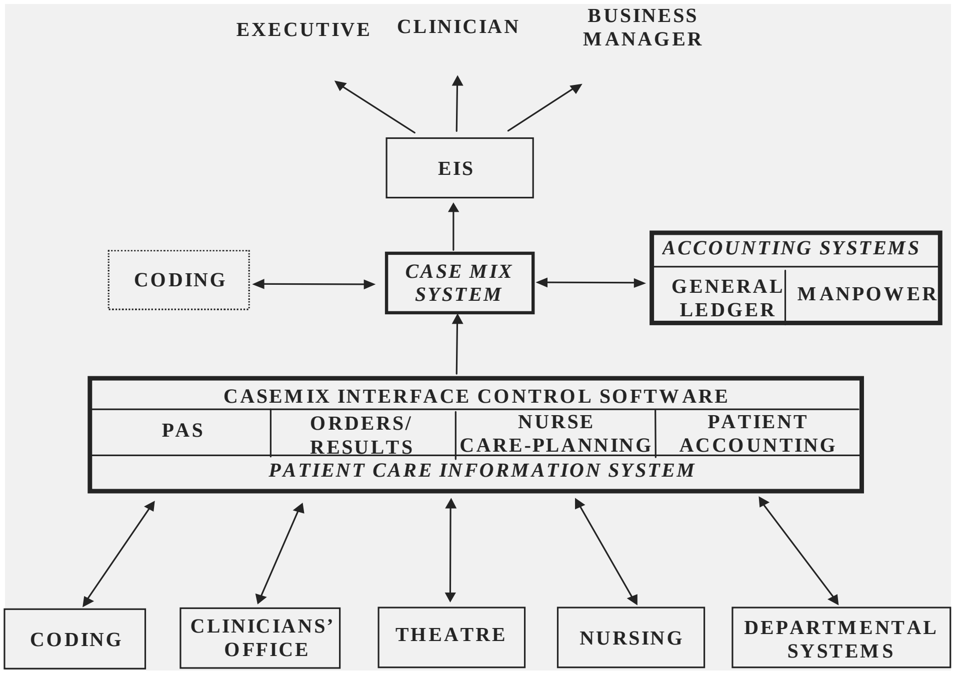
CMIS takes a central position in the hospitals' IT infrastructure, as shown in Figure 1 (an illustrative example - specifications may vary depending upon the hospital), providing a tool for collecting and analyzing data from all areas of hospital operations using an executive information system (EIS). As we can see from Figure 1, there are two main types of data feed: financial and medical. The financial feed consists of pulling data from the general ledger and humanpower systems, particularly standard costs and budgets. This contrasts with the other main feed to CMIS, that of the "patient care information system" - a label given to the array of feeder sub-systems providing information on all aspects of patient treatment and care.

Each of the feeder systems is interfaced with CMIS, so as to provide appropriate data in an acceptable format. Such data are accumulated by CMIS within the care profile sub-system: this stores the actual tests, treatments, costs, number of cases and so on to be compared with expected "ideal" profiles or projected activity levels as drawn from the financial data, enabling financial audit. Regarding clinical audit, CMIS provides the tool for assessment of the professional clinical practices of each clinician.

EXPERIENCES IN INFORMATION SYSTEMS IMPLEMENTATION

In addition to traditional problems of IS implementation, RMI also reveals a number of interesting and contrasting influences on the early implementation of strategic IS in UK hospitals:

Figure 1. The role of case mix in NHS hospitals



- *Central influence* - Regional authorities (RAs) had a substantial influence on the development of systems in a number of areas, for example investment justification, what to procure, objectives, financing, and project management milestones. Importantly, the original "standard" systems were not sensitive to the needs of individual hospitals, affecting stakeholder support and the need for project redefinition. Where relationships with RAs were tenuous, RMI was looked upon with suspicion. Initiatives imposed subsequently changed the shape and direction of the project, and created other priorities within hospitals, while recommendations about clinical coding were never clear. Central influence is, interestingly, both a reason for the existence of the project, and for many of its problems: paradoxically, it is both an enabler and an inhibitor.
- *Project purpose* **-** The problems in communicating the purpose of CMIS, and in approaching locally sensitive designs, affected the attitudes of stakeholders: the project was very much an imposed directive. Many individuals were unclear about the rationale for CMIS, and this was compounded by the traditional absence of IT within hospitals. While RM was aimed at improving resource allocation, ironically, many saw IT spending as a waste of money as opposed to direct patient care.
- *Clinicians, management and CMIS* **-** Hospitals are distributed organizations, with a variety of fragmented groups and cultures, each with their own defined roles. RMI and CMIS cut a swathe through many of these, particularly clinicians and management, and sat uncomfortably in hospitals. Where the system impinged upon the territory of individuals or groups, friction was encountered. Hospitals, for many years without IT, and with a defined structure and set of roles, found change difficult. While

politics occurred in a variety of areas, the clinician-management divide is perhaps the most interesting. Each of these groups found it difficult to adjust to a new culture and structure, and for apparently conflicting reasons: top management was often opposed to devolvement, while clinicians were wary of CMIS as a "tool of management" and a control device. Thus, paradoxically, CMIS was seen both as a tool of centralization and of devolution.

CONCLUSIONS AND FUTURE TRENDS

The complexities surrounding RMI compounded early experiences in strategic IS implementation in the NHS. Evidence suggests that success has been difficult to achieve, and that hospitals found it difficult to achieve any tangible benefits (e.g., see Barnes, 2001; Barnes & Targett, 1999; Lock, 1996). Evaluating investments is made all the more difficult by the fact that published evidence is scarce. Lock (1996) found that only 5% (by value) of all investments were the subject of published assessment, which may be an indicator of the poor value of such investments. Even published studies are far from conclusive: the "official" resource management evaluation conducted by Brunel University found "no measurable patient benefits" in its review of the six pilot sites, while costs were more than double those expected (Heath Economics Research Group, 1991).

One of the key impacts of case mix has been on the financial management of hospitals. The sustainable financial management of hospitals has recently come into question (Audit Commission, 2003). During 2003, the UK Department of Health has outlined the Government's desire to implement PPS, as outlined in its *Payment by Results* White Paper (Department of Health, 2003b). This would allow the trust financial regime (TFR) of hospitals to carry deficits and surpluses, spurring greater efficiency. This has led to some concerns that it may lead to under-investment and reduction in quality (Medicare Payment Advisory Commission, 2003; Rosenberg & Browne, 2001; Topping, 1997). On balance, the evidence from the U.S. and elsewhere seems to suggest that these systems are not detrimental to the quality of health care (Dismuke & Guimaraes, 2002; Shen, 2003).

REFERENCES

Audit Commission. (2003). *Achieving the NHS plan*. HMSO, London.

Barnes, S.J., & Targett, D. (1999). A framework for strategic IS implementation in the UK health sector. *Topics in Health Information Management, 19*(4), 62-74.

Barnes, S.J. (2001). Experiences in SIS implementation in UK healthcare. In T. Spil & R. Stegwee (Eds.), *Strategies of healthcare information systems* (pp. 11-30). Hershey, PA: Idea Group Publishing.

Collins, L.W. (1994). CIOs must look beyond the IS horizon. *Computers in Healthcare, 12*(5), 39-40.

Department of Health. (2002). *Delivering 21st century IT support for the NHS*. HMSO, London.

Department of Health. (2003a). *Implementing payment by results and investment in technology and drugs.* Technical Paper, DoH, London.

Department of Health. (2003b). *Payment by results.* HMSO, London.

DHSS (Department of Health and Social Security). (1986). *Health services management: Resource management (management budgeting) in health authorities.* HMSO, London.

Dismuke, C.E., & Guimaraes, P. (2002). Has the caveat of case-mix based payment influenced the quality of inpatient hospital care in Portugal? *Applied Economics, 34,* 1301-1307.

Health Economics Research Group. (1991). *Final report of the Brunel University evaluation of resource management.* Brunel University, Uxbridge.

Lock, C. (1996). What value do computers provide to NHS hospitals? *British Medical Journal, 312,* 1407-1410.

Medicare Payment Advisory Commission. (2003). *Report to Congress: Medicare payment policy.* US Congress, Washington DC.

Rosenberg, M.A., & Browne, M.J. (2001). The impact of the inpatient prospective payment system and diagnostic-related groups: A survey of the literature. *North American Actuarial Journal, 5*(4), 84-94.

Shen, Y. (2003). The effect of financial pressure on the quality of care in hospitals. *Journal of Health Economics, 22,* 243-269.

Topping, S. (1997). The impact of health care reform on capital acquisition for hospitals. *The Financial Review, 32*(3), 751-778.

KEY TERMS

Case Mix: On admission to a hospital, the individual patient is provided with a variety of services, each of which expends resources. Each encounter or episode has a mix of, for example, tests, treatments, accommodation, and so on. Case mix refers the combination of services and resources used for each patient episode.

Case Mix Information System (CMIS): An information system, fed with data from an array of hospital subsystems, for the principal purpose of clinical and financial audit of patient cases. Further developments to CMIS have seen other functionality added, such as contract planning, quality assurance, waiting lists, and clinical support.

Clinical Audit: The assessment of professional clinical practices. General information routinely provided includes lengths of stay, deaths and re-admissions. More specific information such as drugs administered and operative procedures performed are available via the appropriate feeder systems. The data allow aspects of case mix, clinical management, diagnostic accuracy and patient outcomes to be compared.

Clinical Coding: Categorises diagnoses and procedures for patient episodes according to a detailed clinical standard index. The most common standard is the International Classification of Diseases (ICD), particularly for financial audit, but a UK standard, Read, is very popular with medical staff. These are grouped to provide aggregate information for audit.

Executive Information System (EIS): An IS that provides high-level information tailored to the needs of decision-makers. Typically a graphics-oriented system, in the context of CMIS the EIS aggregates and uses summary data to present information to clinicians, executives and business managers in the form of graphs, histograms, pie charts, tables and so on.

Financial Audit: The budgetary assessment of the cost of hospital care. CMIS collects data on the actual tests, treatments, costs, number of cases and so on. These are then compared with expected “ideal” profiles or projected activity levels as drawn from the financial data, typically based on best practices or historical data.

Resource Management Initiative (RMI): A UK initiative providing clinicians and other hospital managers with information required to use resources to maximum effect, generally by the introduction of new IT. This was aimed at encouraging clinicians to take more interest and involvement in the management of hospital and community units, making them responsible for operational and strategic decisions.

This work was previously published in Encyclopedia of Information Science and Technology, Vol. 3, edited by M. Khosrow-Pour, pp. 1669-1672, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 4.12
Applying Adaptive Structuration Theory to Health Information Systems Adoption:
A Case Study

Dana Schwieger
Southeast Missouri State University, USA

Arlyn Melcher
Southern Illinois University, USA

C. Ranganathan
University of Illinois at Chicago, USA

H. Joseph Wen
Southeast Missouri State University, USA

ABSTRACT

Adaptive Structuration Theory (AST) is rapidly becoming an important theoretical paradigm for comprehending the impact of advanced information technologies. In this paper, a modified AST model was designed to illustrate the changing inter-relationships among the variables affecting the adoption and application of a new technology into a medical organization setting. Using findings from a case study conducted over a 10-month period, the authors apply the case to the model to illustrate the complex interactions between medical billing technology and organizational processes. As the organization attempted to install and implement the new system, they found that in order to maintain daily operations, they would have to modify and adapt several aspects of the organization, technology, and operations. As the system was slowly integrated into operations and the organization's

needs evolved through the adaptation process, the study, in turn, found that different iterations of the model could emphasize different structures. The case illustrated that the capacity to manage health information systems (HIS) often requires the organization to prioritize its needs and focus its energies on a critical structure while temporarily disregarding others until the primary healthcare processes are under control.

INTRODUCTION

Driven by a need to improve utilization of information and productivity, information technology (IT) has become pervasive in the healthcare industry. Some of the areas in clinical medicine in which technology has been successfully employed include billing and scheduling, practice management, laboratory result reporting, and diagnostic systems. The use of computer technology and information technology in healthcare and its delivery is called medical informatics, which began with the computerization of hospital administration tasks in the 1960s. These systems are best thought of as cost reducing and/or quality improving technologies.

Increased demands for electronic exchange of data have been driven by both internal and external pressures. Hospitals are comprised of a multitude of specialized departments and suppliers requiring that large amounts of clinical as well as financial data be exchanged. External forces consisting of insurance company regulations and guidelines (Hagland, 1998), government mandates and restrictions as well as Medicare deadlines (Straub, 1998) have pushed organizations to adopt technologies to automate their operations. Automating these processes may reduce costs as less paper is generated, as fewer mistakes are made, and as information is transferred faster.

A Health Information System (HIS) can also increase the quality of medical care. This was the goal of many of the pioneers in medical informatics or clinical systems development. The quality improvements from hospital information systems would emerge from the improved record-keeping and decreased mistakes engendered by more administrative systems, as well as from clinical systems designed to aid in the provision of medical care.

Today, the role of HIS in medical care has expanded at an ever-increasing pace. As a result, health care professionals' familiarity with medical informatics as well as the adoption of HIS is crucial for the delivery of higher quality care. However, the challenges of applying IT to healthcare are very real. Concerns of privacy and confidentiality of data, lack of national standards for protecting medical data, the need for large scale investments, and the requirement for behavioral adaptations on the part of patients, physicians, and organizations are just a few of the impediments to the adoption and use of IT in healthcare.

Rural area medical practices are especially feeling squeezed by the demands being placed upon the use of technology in the medical field. Although their use of HIS is limited, governmental regulations and the demands of insurance clearinghouses are forcing these clinics to adopt automated billing technologies. Some clinics, unable to afford billing technology capable of electronic data exchange, have been forced to merge with other practices or close their doors. Those clinics that could afford the technology experienced the challenges associated with adopting this new billing system into their business operations.

This paper uses a version of Adaptive Structuration Theory (AST) to examine the challenges faced by a rural medical clinic as it adopts new billing technology. AST provides a conceptual change model that helps capture the longitudinal change process. This paper proposes a modified AST model, which provides a theoretical framework that explains the appropriation process of medical electronic billing systems (MEBS). In recent years, MEBS has become a critical tool for supporting healthcare services. The appropriation of MEBS in a medical center involves a great deal of change, which, if not carefully considered, could result in significant difficulties. Using a case study approach, this research identifies appropriation issues when planning and evaluating MEBS usage in medical centers.

LITERATURE REVIEW

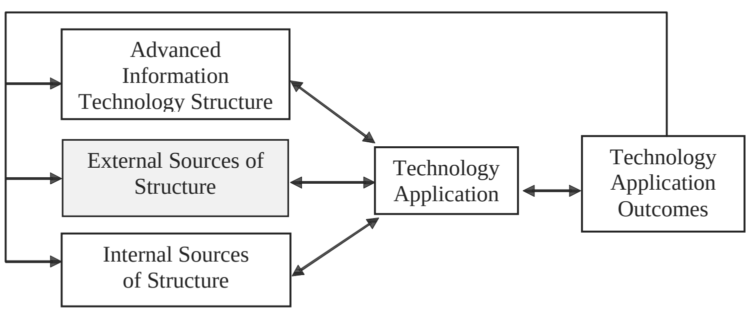
Changes in information technologies cannot be viewed as isolated events; rather, one must be mindful of the interdependent, reciprocally structuring relationships that exist between the information technology and the organization (Lucas & Baroudi, 1994; Orlikowski & Baroudi, 1996). One strand of research dealing with this type of incorporative change process is adaptive structuration theory originally posited by DeSanctis and Poole (1994) as an extension of Anthony Giddens' structuration theory (Giddens, 1979, 1994). Adaptive structuration theory focuses upon the interrelated dynamics embedded in the application/creation of the technology that is in use by the organization through the combined processes of human interaction, technology, and organizational social structures (Griffith, 1999; Lucas & Baroudi, 1994). This study attempts to extend the research in adaptive structure theory through the development of a modified version of DeSanctis and Poole's 1994 model. The model is depicted in Figure 1.

DeSanctis and Poole found that AST, over time, led to changes in the rules, processes, and procedures which were used within group decision support system social interactions. In their study, DeSanctis and Poole (1994) defined AST as "an approach for studying the role of advanced information technologies in organization change" (p. 121). Their research, however, focused upon a "snapshot" of a meeting in which group decision support systems were being used to study interaction at the micro- or individual level of the organization rather than at the institutional level. In light of their definition, AST could best be applied from a longitudinal perspective rather than just an instance in time.

This study extends the scope of research in the area of emergent organizational structure by applying a modified version of DeSanctis and Poole's model to a management information system at the macro-level of organization structure. Case study research is applied to describe and analyze the intraorganizational structural changes that develop in a multi-clinic, Midwest-based medical organization as a result of the introduction and implementation of medical billing technologies into daily operations.

Structuration theory has been approached from three different schools of thought: institutional,

Figure 1. Research model adapted from DeSanctis and Poole's 1994 model



decision-making, and socio-technical. The institutional school of thought viewed technology as an opportunity for change rather than as a causal agent of change (Barley & Tolbert, 1997; Kling, 1980). Under this perspective, technology did not determine behavior. Instead, people generated social constructions of technology (DeSanctis & Poole, 1994; Orlikowski, 1992).

The literature focusing on the decision-making school of thought emphasized the cognitive processes associated with rational decision-making and adopted a psychological approach to the study of technology and change (DeSanctis & Poole, 1994). The socio-technical school of thought combined the institutional and decision-making perspectives by incorporating the power of existing social practices with the influence of advanced technologies for shaping interaction. This, in turn, was thought to bring about organizational change (DeSanctis & Poole, 1994).

DeSanctis and Poole (1994) continued along the socio-technical line of thought with AST. This theory accounted for the structural potential of technology while maintaining focus upon the use of technology as a primary determinant of technology impacts. Their study examined the effects that occurred when advanced technologies were brought into social interaction to affect behavioral change.

Few studies have examined the affects of the diffusion process from the perspective of AST. The diffusion process was considered from beginning to end using the diffusion model posited by Kwon and Zmud (1987) and later modified by Cooper and Zmud (1990). Diffusion literature was applied in parallel to the foundational constructs of the adapted research model to define structure of advanced information technology, external sources of structure, internal sources of structure, technology application, and finally, technology application outcomes.

RESEARCH MODEL

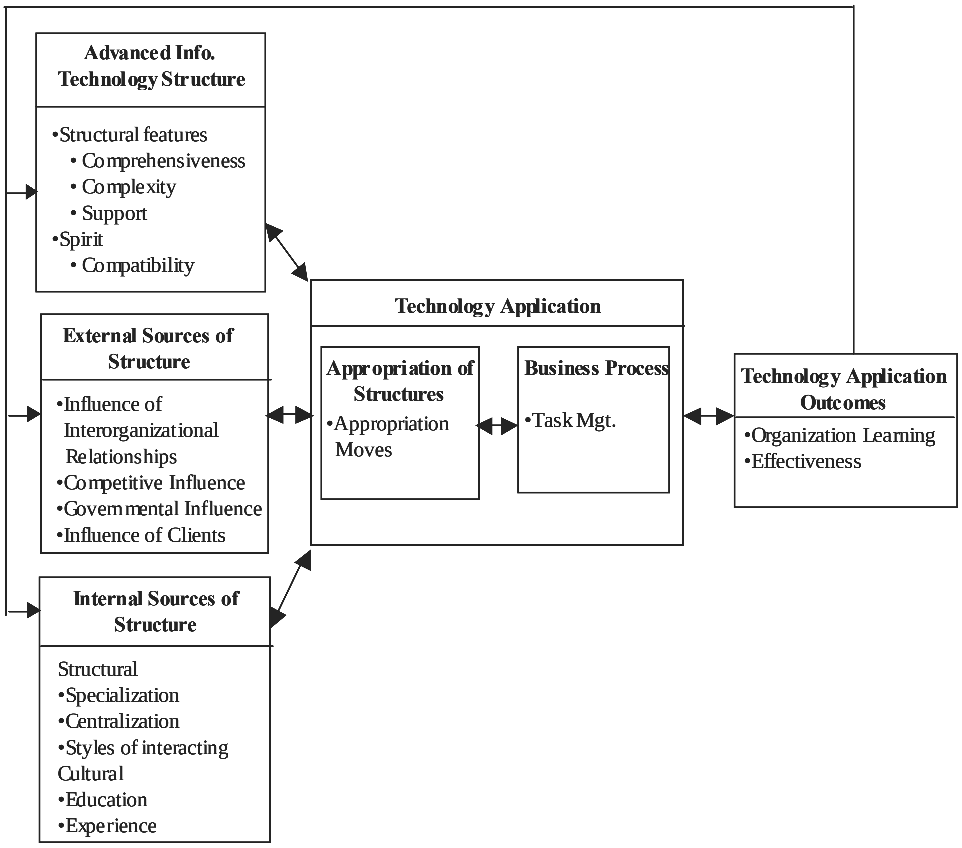
By combining AST (DeSanctis & Poole, 1994) with diffusion theory (Kwon & Zmud, 1987), a model encompassing both areas is proposed. The constructs of the model are defined, for the most part, using Kwon and Zmud's (1987) literature review on diffusion theory and a simplification of DeSanctis and Poole's original model. The research model displaying construct definitions is presented in Figure 2. Each construct of the model is defined in the following paragraphs and illustrated in the accompanying tables associated with each variable.

Advanced Information Technology Structure

Advanced information technology structure is defined as the structural features and the spirit of the technology that outline the expected use of technology within an organization. One of the two elements of structural features includes the technology's comprehensiveness in terms of the number and variety of features and capabilities that users are offered. These resources direct how information is gathered, managed, and manipulated by users (DeSanctis & Poole, 1994). The second element is the level of complexity defined as the degree of difficulty that users experience in using the new technology (Kwon & Zmud, 1987). Complexity can be affected by the ease of use of the application, ease of learning through tutorials, user-friendly screens, help screens, documentation as well as customer service and support.

Spirit of the technology is defined as the "general intent of the technology with regard to values and goals underlying the structural features" (DeSanctis & Poole, 1994, p. 126). Spirit defines the behaviors appropriate in using the technology and helping users understand and interpret its meaning to the organization (DeSanctis &

Figure 2. Research model



Poole, 1994). The spirit of advanced information technology is defined by compatibility. Compatibility is related to the "fit" of an innovation to an organization in relation to the organization's needs and its underlying goals and values (Avolio & Dodge, 2001; Kwon & Zmud, 1987). The construct ranges defining advanced information technology structure are illustrated in Table 1.

Table 1. Advanced information technology structure construct definition range

Advanced Information Technology Structure	Traditional	———	Customizable
Structural Features	Basic	———	Full-featured
Comprehensiveness	Basic	———	All-Inclusive
Complexity	Elementary	———	Advanced
Support	Poor	———	Rich
Spirit	Unfavorable	———	Favorable
Compatibility	Unsuited	———	Well-suited

Table 2. External sources of structure construct definition range

External Sources of Structure	Reactive	———	Proactive
Influence of Interorganizational Relationships	Reactive	———	Proactive
Influence of Competition	Reactive	———	Proactive
Influence of Government	Reactive	———	Proactive
Influence of Clients	Reactive	———	Proactive

External Sources of Structure

External sources of structure consist of interorganizational relationships, governmental influence, competitive influence, and client influence. Interorganizational relationships examine the influence exerted on an organization through affiliations with other organizations. Governmental influence examines the effect that government has on technology application decisions. Competitive influence explores the effect that competitive organizations have on technology application decisions. Client influence assesses the influence patients have on the organization. Thus, the external environment is related to the variability of the environment in regards to instability and turbulence provided by outside sources and the actions taken in regards to these variations (Kwon & Zmud, 1987). See Table 2.

Internal Sources of Structure

Internal sources of structure are divided into two parts: structural and cultural (Table 3). Structural sources of structure consist of specialization and centralization. Specialization is defined as a diversity of specialists within an organization in which individuals or groups of individuals have special functional skills (Kwon & Zmud, 1987). Centralization relates to the concentration of decision-making activity (Kwon & Zmud, 1987). A centralized focus increases the decision-making activity of upper management while a more decentralized organization can increase the decision-making capabilities of lower-level employees.

Cultural sources of structure consist of education, experience and styles of interacting. The organization members' style of interacting can

Table 3. Internal sources of structure construct definition range

Internal Sources of Structure	Less Developed	———	Developed
Structural	Less Developed	———	Developed
Specialization	Generalized	———	Functional
Centralization	Low	———	High
Cultural	Less knowledgeable	———	Knowledgeable
Education	High School Diploma / GED	———	Advanced Degree
Experience	Little	———	Extensive
Styles of interacting	Informal	———	Formal

be influenced by a centralized structure, the formality of management operations, and the boundaries of hierarchical or functional levels. Members of the organization are influenced by the knowledge, education, and experience with technology, both actual and perceived, that others in the organization have.

Technology Application

Technology application is defined as the means by which the technology application outcomes are formulated. This portion of the model consists of two interacting variables, appropriation of structures and business processes (Table 4). The technology application process of the model is influenced by, and likewise influences, advanced information technology structure, external sources of structure, internal sources of structure, and technology application outcomes.

Appropriation of Structures

Appropriation of structures is the degree to which the members of the decision-making body agree on what structures should be accepted through adaptive moves as well as their application. Thus, the technology is defined through the use of its features by the organization rather than just its features alone (Reinig & Shin, 2002; Salisbury, Chin, Gopal, & Newsted, 2002). This is influenced by the intended purposes and meanings associated with applying the technology. The technology structures can be used directly, related to other structures, replaced by other structures, or interpreted as they are applied. Increased agreement on appropriation of structures can lead to greater consistency in the organization's usage patterns of advanced information technology.

Business Processes

Business processes are the actual courses of action in which businesses perform in daily operations. Hunt (1996) defined business process as "a series of steps designed to produce a product or service" (p. 3). Uses of technology and management of tasks can be in-line with the appropriation of structures. As the appropriation of structures vary over time and across adopters, decision processes and technology application outcomes can likewise vary. The business processes determine how the technology can be implemented within the organization. Likewise, the management of the technology can affect implementation of the technology to accommodate business processes.

Technology Process Outcomes

Technology process outcomes are defined as effectiveness and organizational learning (Table 5). Organizational learning may be defined as gaining the necessary knowledge, skills (Attewell, 1992), and expertise (Premkumar & Ramamurthy, 1995) to effectively utilize a technology and apply it to business process. Cooper and Zmud (1990) defined effectiveness as the ability to

Table 4. Technology application construct definition range

Technology Application	Traditional	———	Innovative
Appropriation of Structures	Traditional	———	Innovative
Appropriation moves	Traditional	———	Innovative
Business Processes	Traditional	———	Innovative
Task Management	Traditional	———	Innovative

Table 5. Business process outcomes construct definition range

Decision Outcomes	Low Quality	High Quality
Organizational Learning	Less Successful	Successful
Effectiveness	Less Favorable	Favorable

take full advantage of the information system being implemented. This sub-property examines the fit of the technology as it has been applied to the needs of the organization and its related technological needs. The technology application outcomes are affected by the information gained through the technology application process as well as the implementation of the technology application outcomes. This, in turn, affects technology application processes as well as the medical electronic billing technology structure, external sources of structure, and the internal sources of structure.

RESEARCH METHOD

A case study approach was applied to a multi-physician family practice located in the Midwest and identified as "Family Medical". This organization was selected for the study based upon its current level of growth and application of technology to business operations. Family Medical had been formed six years prior to the study when two separate practices, "Practice A" and "Practice B" consolidated their operations into Practice A's larger medical facility. Upon consolidation, Practice B's medical records were manually coded into Practice A's electronic billing system. Over the 6-year time period, Family Medical had faced tremendous change and technological upheaval as they first combined the operations of two separate facilities and then realized that a new billing application would be necessary to accommodate the aging technology and continuously swelling database. The combined medical practice had grown to provide services to over 22,000 patients and as Practice B's medical records were gradually added to the database, the efficiency of the system slowly declined.

The director of MIS and operations and a billing clerk were interviewed using a focused interview technique. The flexibility of the open-ended questions was supportive of the exploratory nature of the study and allowed for both the collection of facts as well as opinions. Through the utilization of multiple interviews, a more thorough clinic perspective was provided via observation through multiple viewpoints contributed by each medical office interviewee (Straub, 1998).

Documentation and statistics published by the U.S. Department of Health and Human Services, Health Care Financing Administration, American Medical Association, and Medicare were reviewed as they related to medical clinic mandates and regulations. Through site visits, direct observations were made of the medical billing technology and billing procedures employed in the medical office. Information regarding processes, environmental conditions, office attitudes, and structure were observed.

RESULTS

The application of new technology into the business operations of Family Medical illustrated the complex interactions between advanced information technologies, internal and external sources of structure, the application of technology to business processes, and, ultimately, the outcomes of the application of the technology to business processes.

The main contact person at Family Medical was "Steve," the business manager. He was in charge of overseeing several aspects of business operations including, accounting, billing operations, and information systems. Although Steve recognized the benefits and necessity of technology in the medical office setting, he was still recovering from the stressful toll placed on his company by the latest technology overhaul and regretted losing the stability of the former system.

The previous system was a fairly fail-safe system. Employees just typed into dumb terminals on a Unix platform. It was very easy to use and there was little margin for user error. The old application caused very few organizational problems unlike the new software. Actually, our old software matched the practice more so than the new software that we purchased.

The decision to purchase the new billing technology arose after the two practices joined forces and realized that they needed to update their outdated billing technology to accommodate current medical billing trends as well as the heavier usage of the system. Newer systems were capable of providing features such as system integration, graphical user interfaces, statistics and sophisticated charting and reports. Although Steve was more familiar with the managerial side of operations rather than the technical side, the Board of Directors elected him to lead the management team in the search for a medical billing system replacement.

It took 2 years for us to choose our billing software and we still chose a bad system. There were a lot of changes in the medical billing technology industry during the time we were trying to select a new system. It was like trying to hit a moving target. There was an evolution going on in the industry. We would go to one conference and the software and company would be there, then, we would go to another conference and the company would have been bought, sold or absorbed by another company. By the time we had gotten down to the actual selection, there were really only two or three integrated software packages to choose from. This was a bad time to have to be choosing a software package. There was too much volatility.

At the urging of the Board of Directors, the management team finally settled upon an application to recommend. They gathered their research and formalized a report to present at the next board meeting.

Ultimately, management chose the billing software with approval of the Board of Directors. The software was chosen based upon the features that it provided, specifically, the electronic medical records module of the medical clinic management features. The organization did not focus upon the features of the billing technology [in the practice management portion] of the application. Ironically, we have never been able to get the medical clinic management features to work correctly and the only part of the application that we utilize is the billing portion. In retrospect, we feel that we should have paid more attention to the practice management portion of the application.

Although the selected package came with numerous bells and whistles, Family Medical failed to focus on the features that most fit the needs of the organization and its application to business processes. Family Medical also failed to involve a contract lawyer in carefully delineating the responsibilities of the vendor and the associated timeline in which all parts of the implementation process were to be fulfilled. Several implementation factors were orally discussed and supposedly settled by representatives from Family Medical and the software vendor. However, as the implementation process unfolded, Family Medical found that several services provided by the vendor such as training, conversion, installation support, and post-installation support did not quite meet their expectations.

The conversion and implementation processes were worse than they should have been. The installation team was interested in getting the application installed, but not necessarily right, and then moving on. We had so many problems that it was amazing that anything got done.
The changeover [to the new system] was to take place over a long weekend. Complications arose during the installation process and several of the modules would either not work at all or not interoperate with other modules in the system. … Half way through the patient record upload, the system froze. The technicians tried repeatedly to upload the patient records, but due to "compatibility" issues between the new server and the data upload module of the application, the records would not upload and had to be entered manually. In the process of uploading the data, they had corrupted both the patient information as well as the old system rendering both systems unusable. On top of that, no backups had been created to use to restore the old system. Once most of the modules of the system appeared to be up and running, the vendor's installation team took off.
We didn't really know what would and wouldn't work until we started trying to use the system. When we started filing claims, we found out a lot of things were not right, we were not getting acknowledgments back and claims were not getting to where they were supposed to go. By that time, it was too late. The installation team had left and when we would call the vendor's support line, they would either not answer the phone or not return our messages. … It took a whole year to track down every problem one by one and figure out why it happened.

Internal structure: The internal structure of Family Medical experienced tremendous upheaval. Before the new medical software was installed, the organization was functionally-oriented around payer source with employees specializing in specific billing areas such as Medicare, Medicaid, big insurance companies, and so forth. In order to keep operations afloat, the formalized organization structure disintegrated and job responsibilities were assigned on an as-needed basis. People were shifted from various positions to input records into the system.

At the beginning, it was a matter of trying to stay afloat. After the new technology was implemented, we frantically tried to enter all of the patients into the system and be prepared to bill patients on a daily basis. The most important thing that we had to do was to make sure that the patients that we had that day were in the system, covered and billable. The system was used for very little other than processing the day's billing. During this time period, everyone had to be able to help out in other positions. Except for the phone operators, there was no specialization. We moved toward a system where everyone had to do everything. The organization lost in terms of efficiency and speed. The levels disappeared and it was a matter of the company CEO working directly with the billing clerks in order to get the bills processed. Managers and senior account representatives who had more experience and expertise in terms of handling patients, helped in getting the charges and payments posted because we could not afford to bring in a lot of outside help to get everything entered. Patient demographic data is now being entered by reservationists rather than anyone who can be spared and we hope to go back to being more functionally oriented.

Employee morale and respect for various levels of the organization dwindled as problems continued to surface and blame was cast between departments. Areas throughout the organization were affected by the problems caused by the new system.

We went through a period of time when the clinic (medical side also) was waiting on the billing and insurance office to catch up. Everyone perceived the billing department as being very unfamiliar with medical billing technology as well as technology in general. They took some flack for not being caught up because cash flow was so affected. We did not have enough money to give raises for over a year.

At first, if you would have asked the employees if the managers knew what they were doing, they would have said, "No! They have no clue." We were seen as being pretty incompetent in this system, billing processes as well as other forms of technology. The managers were there on nights and weekends, struggling with the billing clerks to get patient records entered. The employees were influenced by seeing the managers get in there and work side-by-side with the people in the billing department. Now that things are starting to settle down in the billing department with the new technology, the attitude toward the competence of this department is no longer as harsh.

Family Medical utilized both informal and formal communication channels throughout the diffusion process. The organization took the standard training and then modified it and condensed the documentation to make it easier for the average person to read. Before the system was installed, they formally discussed with the employees the benefits of getting the new system. They also encouraged the employees to be supportive in assisting each other in trying to learn and use the new system. After the system was installed and the clinic started experiencing problems, management held formal meetings to discuss the problems with the employees as well as actions being taken to remedy the situation.

We did a lot of ground work and training to try to prepare the employees for the new system. We presented the system to them as being an opportunity to receive training that they would never get anywhere else. We tried to let them know that they were being trained so that they could become familiar with the technology and be able to use any system…. Everyone was willing to share information. What one person learned, that person was ready and willing to share to make everyone's lives easier. …Management also had to be very upfront and honest with their people and admit to the implementation problems and their level of difficulty. We had many meetings where the problems were discussed and we had to prioritize the issues.

Technology appropriation: Although the system that was purchased had numerous features for both the business and practice sides, Family Medical was unable to utilize its full capabilities. Not only did they have to deal with the installation problems, they were also faced with the fact that they had purchased the wrong application to fit their business operations. Family Medical was resigned to using the features that they could get to work with minimal to no problems and that fit their operations, ignoring the features that did not fit their organization and continuing to attempt to configure those pieces of the application that they thought would fit their organization but were not yet functional.

The application has a lot of nice features if we could ever learn to use them. It's just too complex. It is more appropriate for a hospital rather than a medical clinic setting. It even has the wrong type of accounting system for what we need in a medical clinic. We purchased the application due *to the medical record portion, but we've never been able to get it to work. We decided to leave the medical record portion of the application alone since it would not work. …We could use some of the information for medical informatics if that part of the system worked. The potential is there, but we just have too many problems to use it.*

Application to business processes: As the technology was applied to the business processes, Family Medical realized that they would have to adapt their daily business operations in order keep the business running. Their main concern was processing the patients to be seen that day so that the patient's information would be available when he/she arrived at the clinic and available for bill processing at the end of the day. They focused their energies on using the billing portion of the application and adapted their business processes to accommodate the technology that worked and the resources that were available. Ancillary operations were foregone to accommodate the more pressing business need of bill processing and payment. Without this function, Family Medical would be unable to finance continued operations.

It is difficult to separate where the problems were. There were some problems with the clearinghouse, some problems getting the information from the clearinghouse to the carrier, and some problems with the feedback process in place to make sure that all of the claims were electronically sent and received by the carrier. We have had to alter the way that we did things and our processes, to some extent, to fit the way that the software worked. The billing process changed due to the problems with the technology. We had everyone go through training for using the application for their job and then we had to move people to different positions in order to get jobs done. The management team had to decide what the priorities were and what could be done and what could be let go. There will probably still be things to catch up on a year from now. …We let a lot of things go that would normally be done on a daily basis just because they were the priorities to keep the business running and to take care of patients. We were too busy trying to deal with the day's patients...

External influences: Overcoming internal application appropriation and implementation issues does not define the limits of the issues Family Medical faced in the application of the technology to business processes. There are several external factors which affected the appropriation of the billing technology to business processes. Due to glitches in the software, some of the bills were not being electronically sent to certain insurance carriers. Family Medical did not realize this omission of data until two to three weeks after the bills had been processed. This prompted them to start developing a system of checks and balances to implement in coordination with the insurance companies to insure that bills were being correctly received.

Their business processes were also heavily influenced by their interactions with Medicare as well as customers. As their biggest external business partner, procedural changes in Medicare's electronic claims handling increased delays in claims processing and, thus, receipt of payments. Time gaps between the patients' office visits and when they would receive their bills added to the confusion of an already complicated document. In response, Family Medical prepared and trained additional staff to specifically handle customer billing inquiries.

… Some of our bills were not getting processed due to errors in the data tables. We didn't find this out until several weeks had passed. We are in the process of changing the system of checks and balances due to the problems that we have encountered with the new software. …Most of the issues that arise from external sources are generated by Medicare. They are the most well-

established of our external associates. Most of our problems with Medicare arise when they change fiscal mediaries or carriers. This affects the clearinghouse also. It really causes problems with our cash flow because we won't get paid on our processed claims for a while. When Medicare changes to a new carrier, Medicare stops processing claims for about six weeks, thus, this adds to the amount of time that it takes to get the money. Medicare also influences the software developers. They are the ones who have to look at the governmental mandates and write patches to keep their users in compliance.
The customer responses also drive our decisions. The customers are still not happy with the statements that they are getting. This will influence our purchase decision when we replace the current application. There are too many business problems that arise from people not understanding their bills.

Outcomes: Family Medical learned that they had to adapt their use of technology, business processes, and organizational structure in order to maintain a minimal level of business operation effectiveness. They continue to wade through the numerous problems created by the implementation of their new medical technology and long for the proficiency of their original system. Although the new system has not been completely installed and configured, they have already begun discussing the possibility of purchasing a replacement system.

The employees liked the previous billing system and its ease of use and navigation. Even though the new system has a graphical interface, after all of the problems we have been through, they still prefer the previous system and would like to go back to it. … We hope to be using another application in five years.

DISCUSSION

There are several lessons that can be learned from the technology adaptation process experienced by Family Medical regarding choosing a technology, the purchase process, and working with vendors and employees.

Solicit the input of effected employees in the purchase decision and process. Family Medical's employees resented having a new technology pushed upon them and especially the fact that they found out about the new technology shortly before the changeover occurred.

Match technology features to business needs. Family Medical chose technology based upon extra features that impressed them rather than the features that would fit the needs of the organization. Once the technology was installed, they found that the features that they needed were not what they had wanted but acceptable and the features that had enticed them to purchase the package were not even functional.

Establish priorities and be willing to adapt as needed. Family Medical was forced to prioritize needs and make immediate and sometimes substantive adaptations in order to keep operations running. Job functions changed across several levels of the organization as the CEO moved from plotting long-term strategies to assisting with data entry. The technology implementation goals of the organization changed from implementing an organization-wide comprehensive health information system to getting a few of the integral billing modules operational so the billing office could perform daily functions.

Obtain the services of a contract lawyer. Several of the promises that the vendor had made to Family Medical were not in writing and were forgotten after the software was purchased. Training, documentation, installation, data integration, maintenance, support, and updates are some of the factors that need to be determined before a purchase contract is signed.

CONCLUSION

Examination of the model through case study analysis illustrated the interconnected relationships between advanced information technology, external sources of structure, and internal sources of structure on the technology application process. The research found that organizations might adapt any or all of those structures to accommodate the business processes or the appropriation of technology to the business processes. Business processes could also be adapted to accommodate the appropriation of the technology and vice versa.

As illustrated by Family Medical's discontinued use of the medical record module and its continued work with the configuration of the non-working modules, part of the technology appropriation process could be learning what features of the application do not work as well as what to do to accommodate and overcome technical difficulties. The capacity to manage business process tasks might require the organization to focus its energies and resources on one structure or certain aspects of structures and temporarily disregard others. To overcome their numerous technological problems, Family Medical found that it had to prioritize its business operations and focus its energies and resources on those processes that were integral to maintaining daily operations. Auxiliary processes were disregarded until the primary business processes were under control.

The adaptive structuration process is a cyclical process in which the organization continues to adjust to its structures as well as to structural changes. The study found that different iterations of the model could emphasize different characteristics of the variables. When the new technology was first introduced, for the most part, the organization utilized a functionally-oriented internal structure with employees assigned to completing the business processes for a particular payer. Due to the onslaught of problems associated with the introduction of the new system, the internal structure essentially disintegrated as available employees were assigned to work on the highest priority task at hand. As the problems became better controlled, the managers started to return the organizational structure to the functional focus.

One concern regarding previous AST research was the lack of focus on one level of the organization. While analyzing the case study data and applying the data to the model, it was difficult to focus on only one level of the organization. Elements from the individual, departmental, and organizational levels all seemed to provide influence in this analysis of AST especially as the levels consolidated.

From observations of Family Medical, the researchers found that it is oftentimes best for an organization to delay the purchase of a technology if they are unable to find an appropriate system that provides a good fit with the organization and matches their most significant needs. However, this alternative is only a viable option if the current system is functional and supplementary resources can be used to maintain or achieve the required level of operations. Problems that arise in a system that is poorly matched to the organization will magnify the employees' level of dissatisfaction.

Communication to employees, both implicit and explicit, plays a significant role in the diffusion process. In addition to normal information sharing methods, when problems arise during the implementation process, it may be necessary for upper-level management to conduct frequent formal meetings to keep employees up-to-date on the status of the situation. As illustrated in the Family Medical case, the employees were greatly affected by the dedication and support management communicated as they worked side-by-side with the billing clerks.

By making adjustments to their internal and external structures, business processes, technology, and application of technology, Family Medical was able to persevere through the technology diffusion process. Although several of the results of the diffusion process were rather negative, application of the case to the model verifies the reciprocal relationships between the constructs.

REFERENCES

Attewell, P. (1992). Technology diffusion and organizational learning: The case of business computing. *Organization Science, 3*, 1-19.

Avolio, B., & Dodge, G. E. (2001). E-leadership: Implications for theory, research, and practice. *Leadership Quarterly, 11*(4), 615-668.

Barley, S. R., & Tolbert, P. S. (1997). Institutionalization and structuration: Studying the links between action and institution. *Organization Studies*, *18*(1), 93-117.

Cooper, R. B., & Zmud, R. W. (1990). Information technology implementation research: A technological diffusion approach. *Management Science, 36*(2), 123-139.

DeSanctis, G., & Poole, M. S. (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization Science, 5*(2), 121-147.

Giddens, A. (1979). *Central problems in social theory: Action, structure and contradiction in social analysis*. Berkeley, CA: University of California Press.

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration.* Cambridge: Polity Press.

Griffith, T. L. (1999). Technology features as triggers for sensemaking. *Academy of Management Review, 24*(3), 472-488.

Hagland, M. (1998). IT for capitation: Getting the whole picture. *Health Management Technology, 19*(9), 22-26, 45.

Hunt, V. D. (1996). *Process mapping: How to reengineer your business processes.* New York, NY: John Wiley and Sons, Inc.

Kling, R. (1980). Social analyses of computing: Theoretical perspectives in recent empirical research. *Computing Surveys, 12*(1), 61-110.

Kwon, T. H., & Zmud, R. W. (1987). Unifying the fragmented models of information systems implementation. In J. R. Boland, & R. Hirschheim (Eds.), *Critical issues in information systems research* (pp. 227-251). New York, NY: John Wiley.

Lucas, H. C., Jr., & Baroudi, J. (1994). The role of information technology in organization design. *Journal of Management Information Systems, 10*(4), 9-23.

Orlikowski, W. J. (1992). The duality of technology: Rethinking the concept of technology in organizations. *Organization Science, 3*, 398-427.

Orlikowski, W. J., & Baroudi, J. J. (1996). Studying information technology in organizations: Research approaches and assumptions. *Information Systems Research, 2*(1), 1-28.

Premkumar, G., & Ramamurthy, K. (1995). The role of interorganizational and organizational factors on the decision mode for adoption of interorganizational systems. *Decision Sciences, 26*(3), 303-336.

Reinig, B. A., & Shin, B. (2002). The dynamic effects of group support systems on group meetings. *Journal of Management Information Systems, 19*(2), 303-325.

Salisbury, W. D., Chin, W. W., Gopal, A., & Newsted, P.R. (2002). Research report: Better theory through measurement-developing a scale to capture consensus on appropriation. *Information Systems Research*, *13*(1), 91-103.

Straub, K. (1998). Financial systems: The next generation. *Health Management Technology, 19*(7), 12-16.

This work was previously published in the International Journal of Healthcare Information Systems and Informatics, edited by J. Tan, Volume 1, Issue 1, pp. 78-92, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 4.13
Adoption of Mobile E-Health Service:
A Professional Medical SMS News Service in Finland

Shengnan Han
Åbo Akademi University, Finland

Pekka Mustonen
The Finnish Medical Society Duodecim, Finland

Matti Seppänen
The Finnish Medical Society Duodecim, Finland

Markku Kallio
The Finnish Medical Society Duodecim, Finland

ABSTRACT

This study investigates physicians' willingness to adopt a professional medical SMS news service in the Finnish healthcare sector. A concise survey using SMS mobile technology was conducted on March 5, 2003. Two hundred and fifty-nine out of 685 responded within 24 hours, and 90% of these answers were received within 6 hours after the survey was sent out. The response rate was 38%. Findings from this simple SMS survey showed that physicians had positive perceptions of the SMS news service. Nearly 60% of the respondents have used it. Some of the answers included spontaneous feedback about the SMS news service, which revealed valuable comments and suggestions regarding further improvements to it. The SMS survey as a new data collection technique needs academic attention. Implications and future research are briefly discussed.

INTRODUCTION

Mobile commerce (m-commerce) has been an important focus of research in recent years. Generally, m-commerce is defined as the extension of electronic commerce (e-commerce) from wired to wireless computers and telecommunications, and from fixed locations to any time, anywhere, and anyone (Keen & Mackintosh, 2001), that is, the use of mobile technologies and devices to provide, sell, and buy convenient, personalized, and location-based services. Many healthcare organizations are turning to m-commerce or wireless solutions in order to achieve better, more effective, and efficient practice management (Wickramasinghe & Misra, 2004). A number of companies are extending their Internet services for physicians for use with personal digital assistants (PDAs) or other mobile devices. For example, the use of PDAs among doctors is rising, and had reached 27% by 2001 in the United States (Harris Interactive, 2001). In Europe the leaders, in terms of the percentage of general practitioners who used PDAs in their practices, were The Netherlands (31%), the United Kingdom (18%), Spain (17%), France (11%), and Germany (10%) (Harris Interactive, 2002). Mobile e-health services might offer an answer to healthcare challenges in the 21st century (Goldberg & Wickramasinghe, 2003).

Text-based technology or short messaging service (SMS) is one of the underlying technology platforms for m-commerce. Compared with wireless Web-based technologies, that is WAP, another m-commerce platform, SMS has a simple user interface and is supported by most mobile phones. Recent years have seen the adoption of SMS worldwide in many sectors of commerce, for example, news, weather forecasting, retail, entertainment, and so forth.

The rapid diffusion of SMS has also inspired some applications in the healthcare industry, for example, a professional medical SMS news service. Medical knowledge is changing constantly. It is not easy for physicians to keep their knowledge and information up to date to help in their patient care and patient management efficiently on the one hand, and to maintain the level of their professional competence on the other (Jousimaa, 2001). A professional medical SMS news service might help physicians keep their knowledge up to date and provide information about recent medical development trends and new discoveries. The aim of this chapter is to investigate how ready physicians are to adopt an available professional medical SMS news service that is currently implemented in the Finnish healthcare sector. First, we present the theoretical background underlying the study. In the following, we present the study context and the survey administration. The method we used for data collection is described in Section 4. In section 5, the results are reported. Discussions and conclusions are at the end.

PHYSICIANS' ADOPTION OF TECHNOLOGY

Users' perceptions of and intentions to adopt an information system (IS) and the rate of diffusion and penetration of technology within and across organizations are two important foci in IS research (e.g., Straub, Limayem, & Karahanna-Evaristo, 1995). They are understood to constitute an essential aspect, property or, value of information technology (Orlikowski & Iacono, 2001). It is generally accepted that the usage of information systems at work could increase employees' productivity in their work, and improve individual and organization performance. System usage is an important dimension for measuring IS success (DeLone & McLean, 1992, 2003). In particular, physicians' adoption of IS in healthcare is aimed to improve the health quality of human beings. In the past few decades, the conclusions of many studies based on different theoretical approaches and research methods have proved and confirmed that usefulness of a supportive medical IS is very important in influencing a physician's decision to use it. For example, Chau and Hu (2002a, 2002b) based their study on classical IS adoption models (e.g., Davis, Bagozzi, & Warshaw, 1989; Moore & Benbasat, 1991) and collected data by means of a questionnaire to investigate physicians' behavior towards adopting new technology. They found that usefulness was one factor determined a physician's adoption of telemedicine technology. Jayasuriya (1998) conducted his research in a similar style. His study indicated that health professionals were willing to use technology in their jobs when they perceived it to be useful for their performance. Many studies have also been conducted using qualitative methodologies. For example, Mayer and Piterman (1999) studied the attitudes of Australian general practitioners towards evidence-based medicine by using a focus group method. They concluded that physicians' readiness to accept technology was influenced by its relevance (usefulness) to general practice, as well as the local contextual and patient factors. Using a prototype simulation study for 1 week in a hospital in Germany, Ammenwerth et al. (2000) successfully found that in order to meet the diverse requirements of different professional groups, the developer was needed to design "multi-device mobile computer architecture". Then, the usefulness of the mobile information and communication system in clinical routines was more likely to be improved.

Obviously, results from different studies by different research methods seem to confirm that usefulness is a fundamental determinant of physicians' readiness to adopt different technologies. The selection of different research methods varies according to the individual IS researcher and largely depends on the problem he/she aims to solve. As suggested by Moody and Buist (1999), the real question is not whether the research method is appropriate per se, but whether it is appropriate in order to answer the question being asked.

STUDY CONTEXT AND THE SURVEY ADMINISTRATION

The Finnish Medical Society Duodecim is a leading provider of medical knowledge and information in Finland. It has put much effort into improving the quality of medical knowledge. It has also adapted new technologies to distribute knowledge to physicians, for instance, CD-ROM, intranet, the Internet, and wireless technology (e.g., Han, Harkke, Mustonen, Seppänen, & Kallio, 2004a, 2004b, 2004c; Jousimaa, 2001).

Duodecim Publishers Ltd., owned by the Finnish Medical Society Duodecim, has developed and provided the SMS professional medical news service (SMS news service for short) for physicians currently employed in the Finnish healthcare sector, including hospitals, healthcare centers, private doctors, and so forth. The news team at Duodecim Publishers Ltd. maintains the service. The physicians subscribe to a specific number of services. After the service has been set up, the physicians receive one item of news a day. Currently, they do not need to pay for the service because of support from Pfizer Finland Ltd. Limitations of SMS technology mean that the service does not include any pictures. With the rapid diffusion of multimedia message (MMS) technology, pictures will be included as well. The SMS news service began in 2002.

After the SMS news service had been implemented for about 1 year, the developer, Duodecim Publisher Ltd. and its supporter Pfizer Finland Ltd., conducted a survey to investigate the physicians' acceptance of the SMS news service and their feedback. Time is of the essence for the service providers in order for them to respond to physicians' changing needs effectively and efficiently. Traditional methods of conducting surveys, for example, mailed paper questionnaires, usually take weeks or months to complete. Therefore, the use of a new medium to distribute the survey was needed. The practitioners com-mended SMS technology. The SMS survey, thus, was conducted on March 5, 2003. It was sent to the target group (685 physicians; among those, 259 subscribers to the SMS news service) as an SMS message, and the answers were gathered via the same medium.

Concerning the limitations of SMS technology, the survey had only one question, that is, “How do you perceive the SMS medical news service by Duodecim?” with three predefined answer alternatives, which were:

a. “The service is really good and useful.”
b. “It is OK, but needs to be improved.”
c. “I haven’t subscribed to the service.”

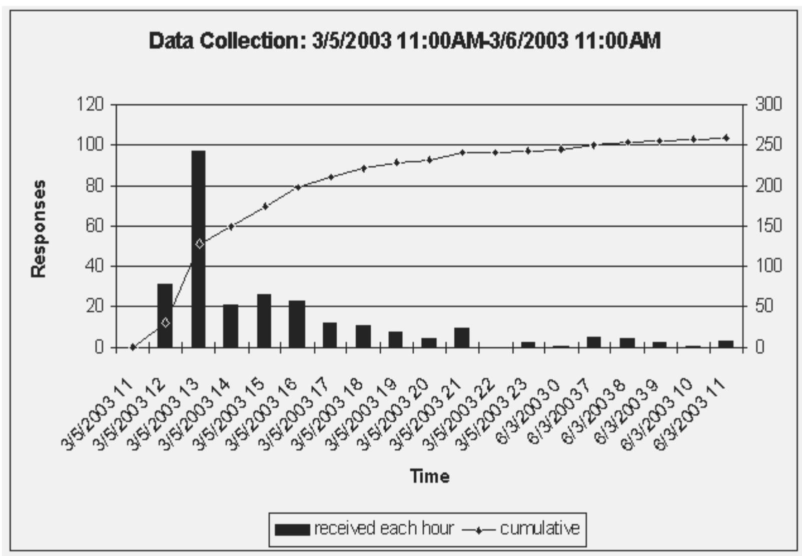
The survey was administered by a company that has experience of mobile marketing and is located in Helsinki. The question was sent out on March 5, 2003, at 12:52 p.m. The whole time span for data collection was 24 hours. On March 6, 2003, 12:00 p.m. sharp, the procedure was finished.

DATA COLLECTION

In Figure 1, we present the time span and numbers of answers collected between March 5, 2003, 11:00 a.m. and March 6, 2003, 11:00 a.m. The x-axis indicates the time series, categorized by hour. The left y-axis shows the numbers of responses received each hour and the y-axis on the right illustrates the cumulative sums of all received answers. Altogether 259 answers (out of 685) were received and the response rate was 38%. Two responses contained empty replies. Among the 259 who subscribed to the service, the response rate was extremely good, over 50%. Some of the answers included spontaneous feedback about the service, which was very valuable.

The majority of answers (>85%, n = 221) were gathered within 6 hours (before March 5, 2003, 19:00) of sending out the question, over 50% (n = 149) within 2 hours (before March 5, 2003, 14:00). Before midnight March 5, we received 243 answers, nearly 95% of the responses; only 16 were sent back on March 6. We soon noticed

Figure 1. Data collection procedure



that the answers received each hour reached a peak between 1 and 2 p.m. on March 5, 2003. The number of responses during this time span was close to 38% (n = 97) of the total.

RESULTS

A basic summary report of our quantitative data is shown in Figure 2. Of 257 usable answers, 31% (n = 79) perceived that the SMS news service was really good and useful. Seventy-three (28%) thought some improvement was necessary. Thirty-two percent of the physicians had not yet subscribed to it. We received 22 answers that did not select any of the predefined alternatives, and we therefore assigned them to D.

We analyzed the content of the feedback provided spontaneously with the three alternatives and other comments categorized as D.

- **Arguments with A**
  Physicians who replied with the A option really liked the service and thought it was very useful for their work. They also showed a positive future intention towards the service. As one physician wrote:

  "[The service] is very useful and I would like to receive the messages in the future.…"

  Although acknowledging the usefulness of the SMS news service, physicians still expected information from traditional channels. A physician replied:

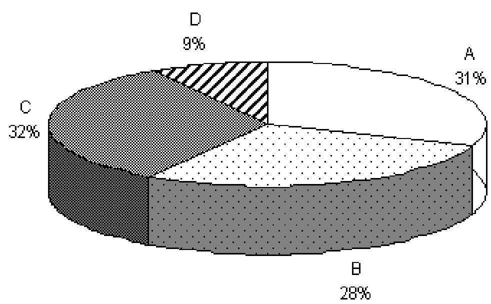
  "The messages are excellent for me and I would be happy to receive other newsletters as well."

- **Arguments with B**
  Most of those who selected B did not specify very much about how to develop the system. Several physicians did give suggestions, however. One physician mentioned that

  "Occasionally one gets news from a short message. [It is] a different picture of the matter compared with the original text [e.g., in newspapers]."

  This message indicated that, as an emerging information distribution channel, mobile SMS was expected to be compatible with other traditional media, such as the texts

Figure 2. Survey results



published in newspapers or on the Internet. On the other hand, physicians also welcomed the SMS news:

"I have during weekdays several times had the time to read the same news on the Internet, but it's good to have repetitions."

Three physicians suggested that the SMS news might include more extensive sources or specialized fields.

"… [the news might come] along with Internet references or more extensive sources."

"… Develop the messages for specialized fields, i.e., information to the correct address...."

"… More accurate sources of references could make it easier to find the article (e.g., about the disadvantage of epilepsy medication)."

Those suggestions shed light on how to improve the content of the SMS news service. Reference sources that help physicians go further to check the latest medical information are quite demanding. Physicians also need information from their specialized fields that are better suited to meet their personal requirements as to the specific knowledge they want to acquire.
Some physicians pointed out that the SMS medical news should be longer; the shortness of SMS was considered a drawback. Comments were received such as:

"The messages could be even longer."

"Apparently the shortness of SMS causes limitations."

- **Arguments with C**
  The feedback came with option C provided several reasons why physicians did not subscribe to the services. The first and obvious reason was they did not need the SMS news service, for example:

"Not used and not needed."

"Not used and I don't want it."

The second reason was some physicians still stick to other distribution channels, such as newspapers or the Internet.

"I haven't followed the text message news. I read newspapers."

"It is easier to read from the Internet."

The third reason was that technical problems with the SMS news service and physicians' lack of relevant technical knowledge hindered their usage of the SMS news service. Complaints included:

"I have not received any news even though I have subscribed."

"I have tried to get the service without success.…"

"Not used since I don't know how.…"

The fourth reason was the small screen of the mobile device: *"I would prefer to read them as e-mails due to the small screen of my mobile phone … probably better with the communicator."* The size of the screen is not comparable with PC-based media apparently.
Some physicians also lacked the time to subscribe to the service: *"I'm in principle interested, but haven't had time to pick up the instructions."* Physicians are usually

very busy. Lack of time is frequently cited as a reason for not adopting new information systems (Berg, 1999).
Besides those negative comments, we also found positive intentions from some non-adopters:

"I will activate."

"I would like to have in use…"

"Reading news as SMS can in principle be feasible."

- **Arguments collected from D**
A content analysis of the feedback we classified as D was conducted as well. We found that there were three main suggestions from the physicians:

1. The service still had some bugs; problems with receiving and opening the service were common complaints. For example:

"None of the text messages that I ordered have been activated, so I have not received any news at present."

"I have not received any news even though I got confirmation of my order."

"For some reason, I have not received news for a long time."

"I acted according to instruction but didn't receive any messages."

"How is it activated?"

These messages indicated that the developer and the service content provider have to pay more attention to the technical problems associated with the SMS news service. The growing number of complaints of the difficulties in activating and receiving it might be a severe barrier for physicians to continue using it.

2. The current length of the message (160–200 characters) was considered both a limitation and an advantage. We got freely formulated feedback of the kind:

"160 characters is quite a limited amount. But on the other hand, one doesn't bother to read too long stories."

"During office hours a short and concise [message] is better...."

Obviously the limited length means that the news has to be concise. This means that they can then save time when reading the news during a busy working day.

3. The news service is quite general; more categories of specialties and more in-depth news are required. For example:

"Seems like a good idea. Personally, I would like to choose those fields that interest me."

"I would wish for more information about urology.…"

These messages revealed that physicians demand more services that meet their "personal" needs. The current service is targeted at all physicians as one uniform group. The developer has to take "personalization" into consideration as a further improvement to the SMS news service.

DISCUSSIONS

On the Survey Question

The findings from this simple SMS survey showed that physicians had positive perceptions of the service. Those responses for options A and B indicated that, generally, the ones (31%) who were happy with the service really liked it, and an almost equal number (28%) thought the service should be improved. The total answers in categories A and B reached 59%.

Arguments came with the predefined alternatives and those categorized as D overlapped in some cases. In brief, they highlighted several important issues concerning how to improve the service in the future. First, technical problems have to be solved immediately. It constitutes a barrier to physicians' usage of the service continuously, or hinders their subscriptions. Possible guidance on how to open and use the service should be improved as well. Ease of use is always considered a crucial driver for the success of an information system. Mobile technology is no exception (Han, Mustonen, Seppänen, & Kallio, 2004d). Second, the length limitation of SMS technology means not only that the news has to be concise, which was considered a time-saving way to read it, but also made it impossible to include more information in one message. With the development and diffusion of MMS, technically the news items could easily be made longer. Third, the physicians demand "personalization." The fact that the current service included only general topics did not exploit to the full advantages of mobile technology, that is, personalized services to mobile users. Therefore, it might undermine the value added by the SMS medical news service. A personalized service for physicians' news needs would enhance their knowledge, which would be highly relevant to their specialties, and provide the information that they desire most. Fourth, as an emerging technology, SMS competes with traditional channels of distributing medical news. The traditional channels, for example, newspapers and the Internet, have dominated information delivery for a long time. Currently, SMS is a complementary channel, but far from the dominating one. A good thing is that it is warmly welcomed by most of the physicians. It has strong growth potential for the future. It adds comparatively more value to physicians' work in terms of flexibility and personalization or localization.

Nearly one third of the respondents had not yet subscribed to the SMS professional medical news service. Attention has to be paid to the reasons why they did not subscribe to the services in order to increase the speed of its diffusion. Compared with the important issues for improving the service referred to above, we find that strong efforts to improve the service might remove the negative "reasons" that impede some physicians from subscribing to it.

On the Data Collection Method

The survey was supported by SMS mobile technology, namely, an SMS survey. We received a 38% response rate within 24 hours, and 90% of responses came within 6 hours after the survey was sent out. We acknowledge the pros and cons of the SMS survey.

Pros

- It is a time-efficient method. The response time is very short compared to other media. Most of the respondents answered within 6 hours after the question was sent out. Within 24 hours, or 1 day, the procedure was completed.
- It is a method that attracts more responses. Compared to traditional survey methods, which usually receive response rates lower than 20%, the survey received a 38% response rate, which was very encouraging. It is also worth highlighting that we received such a high response in 1 day. This is not possible via other media.
- It is a cost-effective method. By using SMS technology, the practitioners have saved a lot of money compared with surveys conducted by other media, for example, paper mail. In Finland, an SMS message costs less than 0.16 Euro.
- It is a method that makes a survey focus on the most important questions. So far, the

limitation of the SMS technology, only allowing 160–200 characters, does not permit inclusion of many questions.

- It is a method supported by new mobile technology. With the development of new technologies, IS researchers try to adapt them to conducting specific studies. For example, tape recording, video recording, Internet, and so forth. The adoption of SMS mobile technology is a trend.
- It is a method containing both quantitative and qualitative data. In the survey, we obtained answers to our predefined alternatives as well as freely formulated feedback. This feedback provided us with fruitful and interesting suggestions, for example, how to improve the SMS medical news service, or why someone did not subscribe to the service.

Cons

- It is a method restricted by technological limitations. The length limitation, 160–200 characters, made it impossible to include more questions that might cover more aspects of a research focus.

Beside those pros and cons of the SMS survey, conducting the survey itself to investigate physicians' acceptance of the SMS medical news service was a very successful experience. Time is crucial for the practitioners to adjust their business models according to the changing demand of customers. Apparently, adoption of mobile technology to support this research is more efficient than other media.

CONCLUSIONS

Users are a key driver for the success of mobile commerce as well as mobile e-health services. It has become increasingly important for IS researchers and practitioners alike to investigate the important factors triggering users' adoption behavior in the new technological context. This chapter is intended to investigate physicians' acceptance and opinions regarding the SMS professional medial news service, a simple mobile e-health service. Physicians' opinions were evaluated using data collected from 257 respondents practicing in the Finnish healthcare sector. Some implications can be drawn from the findings of the study.

An important further development of the mobile e-health service would be to enhance "usefulness" by "personalizing" it. User adoption theories assert that usefulness is the main driver for an individual to adopt a specific information technology. The perceptions of usefulness of the SMS medical news by physicians demonstrate strong possibilities for continuing usage in the future. Providing personalized service is a better way to enhance the "usefulness" of the service. A successful mobile service should be flexible, add value, and take full advantage of mobile technology (e.g., Carlsson & Walden 2001, 2002a, 2002b). It is quite demanding to personalize the SMS professional medical news service according to physicians' different information needs. System developers and contents providers alike should move beyond the limits of current possibilities and devise a service that physicians must have in their working and daily lives. Physicians' work is characterized as busy, pragmatic, and lacking in time (Berg, 1999). A personalized SMS medical news service, on the other hand, fits their work practice. Concise and desirable personalized SMS medical news will both save their time in a busy working day and provide them with new medical knowledge at any time and in any place.

From a managerial standpoint, the findings of this study reveal that, in order to facilitate physicians' motivations to use a technology, it is important to have proper training programs or information sessions to help overcome potential technical difficulties, and also to focus on how the medical news service can help keep their professional knowledge updated. It is better to encourage and cultivate positive intentions towards using the SMS medical service in the work practice. We have to be aware that it is the users and their use, not advanced mobile technology that will drive m-commerce growth to a new level (Jarvenpaa, Lang, Takeda, & Tuunainen, 2003). Obviously, it is equally important for any mobile e-health service as well.

Using SMS to do market research, so-called mobile marketing (Haig, 2002), is not new. The survey reported in this chapter has revealed the benefits of such research. It is a new data collection technique that is time efficient and cost effective. It provided answers in a very short time. The medical news service developer and content providers received "instant" results that help them investigate the problems and find solutions almost at the same time. The SMS survey is not like traditional surveys that usually take months or years; it is a new attempt to do the survey in hours or days with the aid of new mobile technology. It also reduces the cost of the survey. Mail surveys are usually very expensive, as are Internet surveys, which charge a monthly fee. The main weakness of the SMS survey is obviously its length limitation; it excludes the possibility of many questions in the survey. However, it means concentrating on the most important message that practitioners want to know immediately. Such instant feedback might represent high business value for practitioners.

From the management perspective, the SMS-based data collection technique is a very important form of two-way communication and an interaction mechanism for contacting the users—physicians—in the study directly. It is a much more powerful way to speak with physicians and learn their experiences in a relatively short time. It would be very beneficial for service providers to improve the SMS medical news service according to physicians' real needs. The responses from nonadopters revealed that service providers also could adapt the SMS text-messaging survey to inspire awareness of this new mobile medical service, which might lead to mass adoption quickly (Rogers, 1995).

Future research efforts are needed to address these questions rigorously. We might define a complete set of alternative answers for the question (from positive to neutral to negative). We also could ask permission from more physicians to participate in such surveys in order to receive more information both for service development and management purposes. The technical limitations of SMS made it impossible to investigate more relevant questions. Therefore, possible integration with other data collection methods might be necessary.

The research reported here could be seen as a preliminary study reflecting physicians' reactions to a mobile e-health service, especially an SMS professional medical news service. Since the vast majority of the respondents had probably used the SMS professional medical news service in their daily work, their appreciation of the service may be considered as a solid basis for future improvement of the mobile service in healthcare. As far as the external validity of our results is concerned, we have to be cautious since the findings describe the situation in Finland, a country characterized by a very high penetration of mobile phones, and a strong national "mobile fever." Physicians in the Finnish healthcare sector seem to be no exception to this general rule.

ACKNOWLEDGMENT

Our warmest thanks go to Pfizer Finland Ltd. for its support for the survey. The first author would also like to thank Ms. Stina Störling-Sarkkila for her support in translating the Finnish data into English.

REFERENCES

Ammenwerth, E., Buchauer, A., Bladau, B., & Haux, R. (2000). Mobile information and communication tools in the hospital. *International Journal of Medical Informatics, 57,* 21–40.

Berg, M. (1999). Patient care information systems and health care work: A sociotechnical approach. *International Journal of Medical Informatics, 55*, 87–101.

Carlsson, C., & Walden, P. (2001, August 6–12). The mobile commerce quest for value-added products and services. *Proceedings of SSGRR 2001,* L'Aquila, Italy.

Carlsson, C., & Walden, P. (2002a, June 17–19). Mobile commerce: A summary of quests for value-added products and services. *Proceedings of the 15th Bled Electronic Commerce Conference, e-Reality: Constructing the e-Economy,* Bled, Slovenia.

Carlsson, C., & Walden, P. (2002b, July 4–7). Extended quests for value-added products and services in mobile commerce. *Proceedings of the International Conference on Decision Making and Decision Support in the Internet Age*, Cork, Ireland.

Chau, P.Y.K., & Hu, P.J. (2002a). Investigating healthcare professionals' decisions to accept telemedicine technology: An empirical test of competing theories. *Information & Management, 39*(4), 297–311.

Chau, P.Y.K., & Hu, P.J. (2002b). Examining a model of information technology acceptance by individual professionals: An exploratory study. *Journal of Management Information Systems, 18*(4), 191–229.

Davis, F.D., Bagozzi, R.P., & Warshaw, P.R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982–1003.

DeLone, W., & McLean, E. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research, 3*(1), 60–95.

DeLone, W., & McLean, E. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems, 19*(4), 9–30.

Goldberg, S., & Wickramasinghe, N. (2003, January 6–9). 21st century healthcare—the wireless panacea. *Proceedings of the 36th HICSS*, Big Island, Hawaii.

Haig, M. (2002). *Mobile marketing: The message revolution.* London: Kogan Page.

Han, S., Harkke, V., Mustonen, P., Seppänen, M., & Kallio, M. (2004a, May 23–26). Physicians' perceptions and intentions regarding a mobile medical information system: Some basic findings. *Proceedings of the 15th IRMA international conference*, New Orleans, LA.

Han, S., Harkke, V., Mustonen, P., Seppänen, M., & Kallio, M. (2004b, June 13–15). Professional mobile tool: A survey of physicians' perceptions of and attitude towards a mobile information system. *Proceedings of the 5th GITM conference*, San Diego, CA.

Han, S., Harkke, V., Mustonen, P., Seppänen, M., & Kallio, M. (2004c, June 13–15). Mobilizing medical information and knowledge: Some insights from a survey. *Proceedings of ECIS 2004,* Turku, Finland.

Han, S., Mustonen, P., Seppänen, M., & Kallio, M. (2004d, August 5–8). Physicians' behavior intentions regarding a mobile medical information system: An exploratory study. *Proceedings of AMCIS2004,* New York.

Harris Interactive. (2001, August 15). Physicians' use of handheld personal computing devices increases from 15% in 1999 to 26% in 2001. *Health Care News, 1*(25).

Harris Interactive. (2002, August 8). European physicians especially in Sweden, Netherlands and Denmark, lead U.S. in use of electronic medical records. *Health Care News, 2*(16).

Jarvenpaa, S.K., Lang, K.R., Takeda, Y., & Tuunainen, V.K. (2003). Mobile commerce at crossroads. *Communications of the ACM, 46*(12), 41–44.

Jayasuriya, R. (1998). Determinants of microcomputer technology use: Implications for education and training of health staff. *International Journal of Medical Informatics, 50*, 187–194.

Jousimaa, J. (2001). *The clinical use of computerised primary care guidelines.* Published doctoral dissertation, University of Kuopio, Finland.

Keen, P., & Mackintosh, R. (2001). *The freedom economy: Gaining the m-commerce edge in the era of the wireless Internet.* Berkeley, CA: Osborne/McGraw-Hill.

Mayer, J., & Piterman, L. (1999). The attitudes of Australian GPs to evidence-based medicine: A focus group study. *Family Practice, 16*, 627–632.

Moody, D.L., & Buist, A. (1999). Improving links between information systems research and practice—lessons from the medical profession. *Proceedings of the 10th ACIS* (pp. 645–659).

Moore, G.C., & Benbasat, I. (1991). Development of an instrument to measure the perception of adopting and information technology innovation. *Information Systems Research, 2*(3), 192–223.

Orlikowski, W.J., & Iacono, C.S. (2001). Research commentary: Desperately seeking the "IT" in IT research—a call to theorizing the IT artefact. *Information Systems Research, 12*(2), 121–134.

Rogers, E.M. (1995). *Diffusion of innovations* (4th ed.). New York: Free Press.

Straub, D.W., Limayem, M., & Karahanna-Evaristo, E. (1995). Measuring system usage: Implications for IS theory testing. *Management Science, 41*(8), 1328–1342.

Wickramasinghe, N., & Misra, S.K. (2004). A wireless trust model for healthcare. International *Journal of Electronic Healthcare, 1*(1), 60–70.

NOTE

An earlier and short version of the paper is published in the proceedings of the 4th International Conference on Electronic Business (ICEB2004), December 5–9, 2004, Beijing, China.

Chapter 4.14
Computer-Based Health Information Systems:
Projects for Computerization or Health Management?
Empirical Experiences from India

Ranjini C. R.
Lancaster University, UK

Sundeep Sahay
University of Oslo, Norway

ABSTRACT

Large investments are being made to reform the health sector in developing countries as the potential of ICTs in achieving health goals is being increasingly recognized. However, there have been various reports that indicate this potential of ICT is not being fully realized on the ground in particular settings. In this chapter, an empirical investigation of the introduction of health information systems in the primary health-care sector in India is reported. Three cases—the India Health Care Project, Family Health Information Management System, and Integrated Health Information Management Systems—are presented. The authors argue against adopting a technocentric approach during the development of the HIS and suggest that these efforts should be sensitive to the sociotechnical context. Furthermore, a variety of constraints are identified. The chapter concludes with a discussion on the potentials of integration to address some of the identified constraints.

INTRODUCTION: THE CHALLENGES OF HIS IN DEVELOPING COUNTRIES

The plural of datum is not information.

~ Anonymous

Information and communication technologies in primary health-care settings offer a number of opportunities to enhance the efficiency of administration and improve the delivery of health-care services. Health information systems (HISs), geographic information systems (GISs), telemedicine, Web-based initiatives, and the development of health-care databases (see Bodavala, 2002) are examples of a few ICT-based initiatives currently ongoing in the primary health sector in India, and also in other developing countries. Despite the undoubted potential that ICTs provide, obtaining their practical benefits on the ground is a very complex undertaking, and there are various reports of total, partial, sustainability, and replication failures (Heeks, Mundy, & Salazar, 2000). Contributing to this unrealized potential are a number of complex interrelated issues such as inadequacies in both the computer-based infrastructure (for example, Nhampossa, 2004) and physical infrastructure of roads and transportation, which are required for the transmission of reports (Mosse & Nielsen, 2004); that is, there is a lack of proper network infrastructure, the persistent presence of legacy systems embroiled with different institutional interests (Nhampossa), weak human-resource capacities both in numbers and skills (Chilundo, 2004), heavy workload of health staff who need to give priority to providing care over administrative tasks like reporting (Mosse & Sahay, 2003), and a culture of information use that sees periodic reports as primarily fulfilling the needs of the bureaucracy rather than using the information to support action (Quraishy & Gregory, 2005). All these contextual influences make the challenge of introducing ICTs in the health-care sector a very difficult one in practice (Sahay, 2001).

The critical role that HISs can play in public health has been emphasized since the early 1980s in attempts to integrate data collection, processing, reporting, and use to strengthen management at all levels of health services (Lippeveld, Sauerborn, & Bodart, 2000). HISs in most developing countries have been described by researchers and also policy documents emerging from international agencies as being grossly inadequate (for example, Lippeveld, Foltz, and Mahouri, 1992; World Health Organisation [WHO], 1987). Sauerborn and Lippeveld (2000) argue that this ineffectiveness stems from various reasons including the irrelevance and poor quality of data being gathered, duplication and waste among parallel HISs, lack of timely reporting and feedback, and poor use of information. As a result, what we typically find is HISs that are data led rather than action led (Sandiford, Annett, & Cibulski, 1992); in this sense data is an end in itself rather than a basis for planning, decision making, and evaluating interventions. Institutionally, HISs in developing countries are situated in rather centralized structures (Braa, Heywood, & Shung King, 1997; Braa & Nermukh, 2000; Braa et al., 2001) in which local use of information is not encouraged (Opit, 1987). This has led Sandiford et al. to comment that what is needed is not necessarily more information but more use of information.

While as a part of various health reform efforts, including HIS, ICTs are being actively introduced by international agencies and national and local governments, what is often found is that the focus of such efforts are primarily on the means, computerization, rather than the ends of what needs to be achieved: strengthening information support for health management. Introducing ICTs in the development of HIS is not necessarily the silver bullet that solves the efficiency problem of the health services (Sandiford et al., 1992), and over the years research has emphasized that critical issues to be addressed in the implementation of IS are social and organisational, not solely technical (Anderson & Aydin, 1997; Walsham, 1993). As Helfenbein et al. (1987) have argued, changing the way information is gathered, processed, and used for decision making implies making changes in the way an organisation operates. They also suggest that producing and utilizing information more effectively will affect the behaviour and motivation of all personnel.

Policy makers (for example, under the sector-wide approach policy) and also researchers (for example, Piotti, Chilundo, & Sahay, 2005) propose that the integration of an HIS provides mechanisms to address some of the existing inefficiencies of HIS and to develop new forms of health data analysis. For example, the management of HIV programmes requires data from various sources such as the mother, child, and HIV-specific programmes. Rather than developing new HISs, it is argued that the existing systems should be effectively integrated. However, integration of HIS is a complex task. In a study on a national STI (sexually transmitted infections)/HIV/AIDS programme in Mozambique, Piotti et al. (2005) analysed the multiple reporting systems that were in place. Systems for the prevention of mother-to-child transmission, volunteers, counseling and testing, blood-transfusion services, AIDS patients under follow-up for the treatment of infections and/or antiretroviral therapy, rural-district AIDS inpatients, and AIDS patients under follow-up for home-based care were among the many reporting systems that they identified. Despite formal instruments to support the reform efforts to integrate the HIV/AIDS HIS, the multiplicity of reporting systems continued to exist, each with its own practices and processes. The integration of the HIS is not merely a technical or social task, but it is sociotechnical in nature.

We argue that taking a technological determinist perspective on HIS development implies a lack of sensitivity to the sociopolitical and institutional context, and contributes significantly to the unrealized potential of the computerization of HIS in developing countries. There is an urgent need to shift this focus from computerization itself to the question—What is it that we want to achieve through computerization?—in order to strengthen these HIS-focused reform efforts. Analysis of this question helps to develop an argument against the adoption of a technocentric approach to the development of HIS and to instead approach this effort with a focus on integration and sensitivity to the sociotechnical context. Introducing computers does not in itself lead to better handling and use of information, and the challenge is to initiate a parallel informational reform process where information is seen as a resource for action rather than for fulfilling the needs of the bureaucracy.

The present chapter analyzes empirically some of these challenges, drawing upon a study of three specific HIS initiatives introduced in primary health-care settings in Andhra Pradesh (AP; a state in southern India): The India Health Care (IHC) project, the Family Health Information Management System (FHIMS), and Integrated Health Information Management Systems (IHIMS). In the following section, we provide a brief summary of the broad context of the health sector and HIS in India. We next describe the research setting followed by a narrative of the HIS projects. Later, we present our analysis and discuss the challenges and opportunities inherent in shifting the focus of HIS efforts from computerization itself to how it can support health management. Finally, some brief conclusions are presented.

THE INDIAN HEALTH SECTOR CONTEXT

The goal of health policy in India has been stated as supporting universal and free primary health care for all (Gupta & Chen, 1996). A primary health centre was conceived in India as an institutional structure to provide integrated preventive, promotional, curative, and rehabilitative services for the entire rural population. The primary health sector includes primary health centres (PHCs), subcentres (SCs), community health centres, and upgraded PHCs in some areas. Each PHC is expected to provide health care for a population of about 30,000, whereas in practical terms, this ranges from 20,000 to 70,000. More details on the health sector in Andhra Pradesh are provided later.

The health staff in the PHCs and SCs generate considerable amounts of data about the services that they render, collate them on a weekly or monthly basis, and report them in prescribed formats, sometimes duplicating the same report format to different departments or programme managers. These informational activities take up a considerable amount of the health staff's time. Moreover, frontline health assistants are mere producers of data rather than its users. While the focus on reporting serves to fulfill the needs of the bureaucracy and their dependencies on international aid agencies, and the potential need for epidemiological analysis, information is rarely ever used to guide local action at the level at which the data are collected (Sahay, 2001). There are also quality-related issues that inhibit the active use of data. For example, data collected at every level are aggregated, and by the time these data reach the higher levels of the health system, the specific situations in the peripheral areas are completely masked (Quraishy & Gregory, 2005). The flow of data is largely unidirectional from the lower levels to the higher ones (see Figure 1), with very little constructive feedback being given to the local levels to strengthen their work processes. Instead, often the only feedback is in the form of a reprimand, where health staff are pulled up for not meeting their (rather unrealistic) targets. Mechanisms to facilitate mutual exchange of information are also minimal.

Acknowledging these limitations of the existing HIS during the launch of a new programme named the National Rural Health Mission, the prime minister of India, Dr. Manmohan Singh, called for reorienting the HIS to support local action (April 12, 2005):

The monitoring systems have to become oriented outward towards the community and not upward towards the bureaucracy. For example, so far the health information we have in our country through the National Family Health Services Reports are seen at State and Central levels and hardly ever at district and below district levels. If information is to lead to action, it should be available and used at the local level. We must reorient the information system to support an accountability-structure by developing district health reports, state health reports and so on.

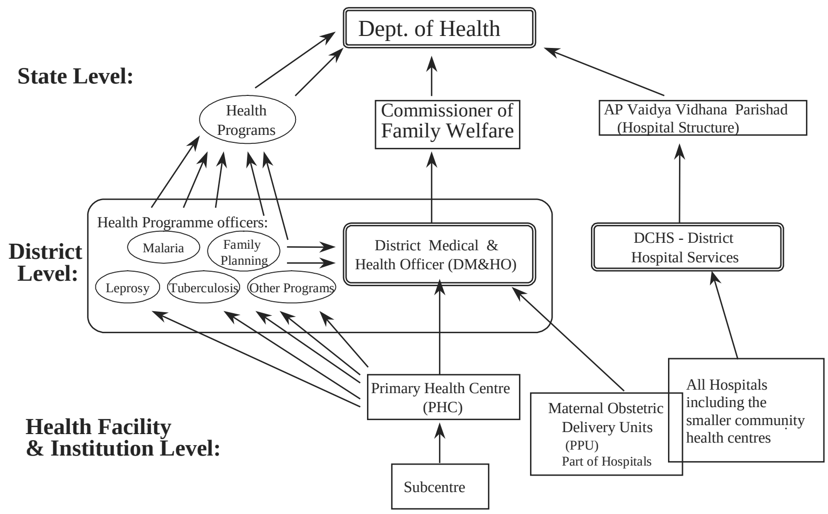
While the challenges to HIS reform are now being recognized at the policy level, the harder task is to address them on the ground. With this as the backdrop, we now shift to describing the specific research setting and the particular HIS implementation efforts.

The empirical material reported here is derived from a situational analysis of the implementation of three HIS projects in AP, a state with a population of 75.7 million (http://www.censusindia.net). Figure 1 provides a broad structure of the health-sector network and an overview of health information flows in Andhra Pradesh.

Subcentres have been established under the PHCs at a rate of one for every 5,000 people in rural areas, and one for every 3,000 people in tribal areas. One or two health assistants (auxilliary nurses and midwives) work at the SCs and are responsible for outreach programmes. They travel from one village to another to provide services such as immunisation, antenatal care, family planning, information and education activities, and school health programmes to the rural population. PHCs serve as the hub for implementing national programmes such as population control, and disease-control programmes like malaria, tuberculosis, leprosy, and gastroenteritis among others. Mainly, however, the PHCs are the cornerstone of the maternity and child health programme, now integrated into the reproductive and child health programme.

For example, one district, Chittoor, has a population of about 3.75 million served by 84 PHCs,

Figure 1. Health Information Flow (Source: Sahay & Walsham, 2005, p. 44).



each theoretically catering to a population of about 30,000. Each PHC has under it about four to seven SCs, each of which in turn is supposed to cater to a population of about 5,000. In practice, the PHC population varies from 25,000 to 40,000, and the SCs from 4,000 to 8,000. The state has about 1,386 PHCs, of which 470 PHCs are expected to provide round-the-clock services for safe deliveries. There will also be 268 hospitals, dispensaries, community health centres, and mobile units, and 10,568 SCs spread across 23 administrative districts, covering an area of about 246,793 sq km. In Andhra Pradesh, there are about 2,848 medical officers, 506 community health officers, 1,686 multipurpose health extension officers, 1,621 female multipurpose health supervisors, 614 public health nurses, about 11,000 multipurpose health assistants (also called auxiliary nurses and midwives), and 500 other administrative staff at the primary health centre level (family welfare [FW] departmental manual).

RESEARCH SETTINGS: ONGOING HIS INITIATIVES IN ANDHRA PRADESH

Over the last few years, HIS projects have been initiated in the Family Welfare Department (FWD) in AP to help provide electronic support tools to the health staff at different levels of the organisation. These HIS efforts of the state government were initiated with the backdrop of larger good-governance reform policies initiated by the former chief minister Chandrababu Naidu within the framework of SMART governance (Simple, Moral, Accountable, Reliable, and Trustworthy). As a component of this ambitious good-governance policy, the chief minister drew up vision statements for every government department. The government's IT vision states, "Andhra Pradesh will leverage Information Technology to attain a position of leadership and excellence in the information age and to transform itself into a knowledge society." The vision document also established ambitious and demanding goals for the health department to be achieved by the year 2020. These goals were, among others, to achieve infant and child mortality rates of 10 per 1,000 (live births) and 20 per 1,000, respectively, a total fertility rate of 1.5 (average number of children per woman), and population growth of 0.8% a year. Good governance and the use of ICTs in public administration were emphasized to help achieve these goals.

In January 2002, the government of Andhra Pradesh brought out a strategy paper on governance and public management that forms the basis for governance reform activities in the state. In the book *Good Governance Initiatives in Andhra Pradesh* (Mohanty, Jones, & Rao, 2004), 28 projects have been documented to disseminate examples of good practices in the field of governance and public-sector reforms. The introductory chapter notes that:

governance involves mechanisms, processes and institutions, through which people articulate their characteristics of good governance to include: rule of law, participation, transparency, responsiveness, consensus, equity and inclusiveness, economy, efficiency and effectiveness, and accountability…The government of Andhra Pradesh has adopted an approach of combining knowledge, technology and citizen-centricity as the key strategy to improve governance in the state. (p. iii)

This focus on good governance to reform public services is indeed welcome. However, there is a need to shift the focus from merely elaborating on the content of reforms to the feasibility of implementing them on the ground. Implementation of these modern and managerial models to support good governance and best practices is complex and problematic (Madon, Sahay, & Sahay, 2004). Given the excessive focus on the provision of infrastructure rather than on implementation and use, asymmetries exist between macro-level governance policies and the micro-level implementation of particular ICT projects (Madon, Sahay, & Sudan, 2005).

Toward a Computerised Health Database: Vision and Reality

In Andhra Pradesh, toward achieving Vision 2020 and the SMART objectives, the government undertook a number of IT initiatives, some of which have had direct implications for the health initiatives under study. One such initiative was the Multipurpose Household Survey (MPHS).

The MPHS was an ambitious project of the government of Andhra Pradesh to build a massive name-based database of 75.7 million citizens and 76.5 million land records to help provide a social-security identification number for every citizen of the state. The survey was originally compiled in 1995. It was later computerised and the data was hosted on a central server for public access. The citizen's database with about 130 data elements has been installed in 1,125 mandal[1] revenue offices. Although this database was created by the state revenue department for its own purposes, the government passed a directive that MPHS should be used as a standard database by all government departments and other agencies in an attempt to prevent multiplicity of databases. In order to comply with this directive, other departments started to add data from their own databases to the MPHS.

Creating a name-based database with various socioeconomic indicators for a population of 75.7 million is a challenging task. A consultant in the revenue department explained that this database was compiled in 1995 by conducting a household survey, and subsequently between 1998 and 2003 two private companies were given contracts to update these data and also to include other data such as deaths, births, migration, and so forth. Additional data (such as the number of school-going children and the number of working children in the age group 6 to 14) from another survey called the Human Development Survey (HDS), which was conducted by the planning department, was integrated with the MPHS data in January 2000. Integrating the data from the HDS into the MPHS database had inherent problems because the basic unit for the MPHS was the village and that for the HDS was the habitation.[2] In the HDS, data were collected on total numbers and not on the details of all individuals in a family. For example, details about a family included the total number of children in the age group of 0 to 5, the number of illiterates above 15 years, and the number of family members above 65 years, while in the MPHS, data were collected about every individual member. There were commissions and omissions because of the migration of people, and also because some people were reluctant to give their details for the survey. There were also quality issues raised about the HDS data as they had been collected by school children. As a result, the database was not accurate, and the quality of data was very poor.

In addition, with regard to our main concern here, the HIS, the PHC as the fundamental unit for the state health department did not match with the fundamental unit of the revenue department (the owners of the MPHS), the village. As a result, many of the health parameters necessary for health services were not included in the MPHS database, and it had to be updated with these data subsequently.

Around the time that the MPHS was being upgraded, in June 2000, Infodev (the Information for Development Programme), an agency of the World Bank, granted $250,000 in funding for the first of the projects under study here: the IHC project. The key objectives of the IHC were to reduce or eliminate the redundant entry of data prevalent in paper registers, to generate monthly reports automatically for health assistants, and to make data electronically available for further analysis and compilation at higher levels of the health-care system. The project was eventually implemented in 200 subcentres (spread across 32 PHCs). The larger objective of the initiative was to reduce infant mortality and maternal mortality by improving the quality of antenatal care and child health through improved information management.

The second project under study here, the FHIMS, was an offshoot of the IHC project. One district, Nalgonda, out of the 23 districts in the state was selected to pilot this project. It took off simultaneously in all 67 PHCs in the district in 2002. The information needs at different levels forced the family welfare department to embark on a more elaborate project to have information systems not just at the PHCs, but also in the district offices and the state office. The overall aim of the project was to "computerize the activities of the PHCs as a whole to take care of (a) Family Welfare needs (b) control of communicable diseases and (c) PHC management" (CFW, Office of the Commissioner of Family Welfare, 2003).

As already mentioned, during the initial days of the IHC project, it was decided that MPHS should form the basis for a name-based health registry, and that instead of conducting a new survey, the health staff should update the MPHS. The survey data were crucial for both the projects because they were to be the foundation of the name-based IS. This name-based system (with details such as, for example, who has malaria) was to replace the earlier one in which data were collected based on numbers (for example, how many cases of malaria in a region). In the FWD, the family or household was the basic unit of analysis, and it was expected that the database would maintain this unit of analysis, along data on births, marriages, family planning methods, antenatal and postnatal cases, immunisation of children, deaths, and diseases. It was argued by the proponents of the name-based system that it was necessary to provide targeted health care, especially in following up with antenatal cases, immunising children over a period of time, and in tracking patients. While the officially stated aim of the FHIMS was to improve patient-specific care, in practice, as one doctor said, the rationale underlying FHIMS was to try and "prevent the manipulation of figures by the health staff, since with names it is easier to track whether or not the right numbers have been filled."

Apart from these data-related issues, there were also software challenges. A consultant in the revenue department explained:

Quality problems with the software are because each department hires an agency on its own to get the necessary fields from the MPHS. That company may not know the design of the MPHS data completely. They are given a very short time, say a month or so, to deliver the product. Understanding a huge database and to know how to do it takes time. Quality suffered because of this.

Another software programme was developed to download the MPHS and print the MPHS data on habitation in the form of a household register. A software engineer of the IT company that had been involved in the development of the software for the FWD said, "Nobody knows about the accuracy of the data. The survey was initially done for the revenue department. We did not know the software codes because it was done many years back. We cannot assure the accuracy of the data."

The development of the FHIMS software was contracted to an IT company located in the state capital at a cost of about $68,500, and the requirements included 17 modules (such as family welfare activities, tuberculosis control, leprosy eradication, budget monitoring, personnel information, etc.). The software was developed on an Oracle back end using Visual Basic. Although the project itself was delayed (because the household survey, which was expected to be completed by December 2003, was still ongoing in March 2004 in many PHCs), the software was expected to be deployed toward the end of 2003. The health department mobilised about $7.3 million for the project through a consortium of funding agencies including the World Bank to buy about 1,500 computers and install it in every PHC and district office, and in the state office, and load the computers with Oracle software (with a license fee of about $200 per computer) and Microsoft software.

The idea of computerization was that the MPHS data would be loaded in the FHIMS software, and using this health survey, data could be entered (based on names). It was soon realized, however, that the MPHS data was unsuitable because of quality issues and the fact that the data did not match the needs of the health department. As already mentioned, this was because the MPHS was conducted by the revenue department and hence did not have all the data that were necessary for the health department.

This process of collecting data for such a massive population, as can be imagined, was a complicated task and involved multiple challenges. When the health staff took the MPHS survey books to their respective habitations, they found them afflicted with many errors. Many health assistants complained that names were missing or inaccurate, ages of family members were wrong, or sometimes data on entire habitations were missing. Sometimes there would be a Muslim name for the head of the household, and the family members would have Hindu names (a quite unlikely possibility!). In another instance, the same name (of a head of the family) was in all the families in a particular habitation. Also, since house numbers were not arranged in an order in many villages, the health assistants had to rely on the names of people to locate their families.

Health assistants were expected to do the survey in addition to their regular work. Some of the health assistants said they were ordered to conduct this survey at their own cost and time. However, higher officials disagreed and said doing such surveys is part of their job. The health assistants would go in the evenings after work to conduct the survey, and had to be accompanied by their husbands because they had to travel after dark. Also, as the list of names in the forms (printed from the FHIMS database) given to them did not match the people who were physically located in the habitations, they found it easier to buy new books and write the names by pen themselves. However, in the absence of any financial support, the health staff had to buy these books at their own cost. Because of time constraints, and the fact that they had to enter whether a person had (or had not) a list of about nine diseases and particulars of physical disability (six types) in code numbers, the assistants would typically simply enter "no" for all. A project coordinator of the FHIMS explained some of these difficulties experienced by the health assistants while conducting the survey during the pilot phase of the project as follows:

The other big difficulty that our staff faced is that they were given registers…per habitation, per village like that. Each book runs, depending on the size of the habitation, anywhere between 500 and 600 pages. Some habitations are really big and have 10,000 people. Our health assistants have to go to the field with these registers, not knowing on which page the family they have to survey is on because the books do not have any index. Addresses are not printed in a serial order as to how they are located nor are the names listed in an alphabetical order. So the health assistant practically had to turn all the 600 pages or 10,000 pages every time to locate a family. And she couldn't just tear all the pages and arrange it one after the other…She knows perfectly well which house comes next to which…That was also not possible because in one sheet we had two to three families. Later, in replication phase we have made sure we get only one family per page and we now have an index for every bound book…in an alphabetical order.

After the house-to-house survey, the next stage of the project was to enter the data into the computerised database, which was a time-consuming and arduous task given the already heavy workload and time constraints of the health assistants. The data entry into the household database, which would then be imported to FHIMS, was outsourced to a private party. This private agency further recruited data-processing operators (data operators; at a salary of about less than $100 per month) to do the data entry.

During the pilot phase of the FHIMS project involving data entry, there were many omissions because the work of the data operators was not well monitored. While the health assistants were instructed to be present with their registers when data were being entered by the data operators to oversee the process, the health assistants would simply leave their registers with the data operators and go away to fulfill other duties. The data operators, for their part, found it difficult to make out the handwriting of the health assistants and the variety of informal symbols they had made in the registers. A retired district officer, who was appointed as a consultant for the project, explained why this lack of oversight of the data-entry process occurred as follows:

On many occasions health assistants were not present when the data entry was done. If the health assistant sat with the data operator with her register, that would slow down the data operator because his work would be constantly double-checked. On the other hand, work would proceed faster when the health assistant was not present because the data operator would worry less about whether he was entering the records correctly or not. Remember his payment was based on the number of records he entered per day. So naturally, his interest was not whether data he entered was right or wrong, his interest was to enter as many records as possible in a day and collect his money. The health assistants were not able to sit with the data operators, oversee and clarify, because she had so many other priorities. Moreover, that sort of accountability and sense *of responsibility was not there among most of the field staff. So, in effect, the data entry was not done in a systematic manner.*

As the project expanded, data entry was moved to the respective PHCs. Now, a new set of problems caused delays in the process. Multiple software bugs, especially due to software incompatibility between the household survey module and the FHIMS database, were some of the problems encountered. The result was that certain data, for instance, data on antenatal women who did not reside in habitation or data on pregnancies leading to abortion or medical termination, could not be entered into the database. In some PHCs, the household survey module could not be linked to FHIMS database, leading to discrepancies between the population figures that were entered and what was actually displayed in FHIMS. In fact, although it was claimed that all the data had been entered, close examination revealed later that data entry was done only for the FWD module and even then for only a portion of it.

The India Health Care Project

The IHC project was started as a pilot in three health centres in the Nalgonda district where health assistants were given personal digital assistants (PDAs) and another health centre was selected to pilot the same application on a desktop computer. Subsequently, PDAs, which are mobile computing devices, were given to about 200 women health assistants. Typically, health assistants spend a considerable amount of time collecting, collating, and reporting health data to different officials. This paperwork is time consuming and cumbersome, and historic data are difficult to manage, retrieve, and use. The health assistants submit several reports of the services they have offered to the supervisors and medical officers at the weekly and monthly meetings, generally organised at the PHCs. At times, they also submit reports on different health programmes to different functionaries and agencies. The stated objectives of the project were to reduce manual paperwork and eliminate redundant data entry, and thereby free up assistants' time for health-care delivery. The application was also expected to facilitate the generation of schedules with information pertaining to immunization and antenatal services for the health assistants, which would help them to geographically identify antenatal cases to be visited and children to be immunized. The system was also supposed to help the health assistant in tracking the history of the patients and enable her to take preventive measures such as treating high-risk pregnant women or giving timely vaccinations to children. The schedules also highlighted pending cases and those who had been missed by her in previous visits. This was supposed to help the health assistant to prioritise her work so that she could attend to cases that needed immediate attention. The PDAs were expected to eventually replace paper registers. Thus, the use of these mobile computing devices was projected as providing ample opportunities for assisting health staff in their routine work. At the PHC level, it was expected that the data from PDAs would regularly be transferred and uploaded onto desktop computers located at the PHCs and support the routine HIS.

However, the use of the PDAs was largely unsuccessful for a variety of reasons. For one, the memory supplied was insufficient (16MB). There were problems with charging the batteries in the rural areas. Many health assistants said the device was slow, and since using the device while on the job was time consuming, a few health assistants would do the data entry later in the bus on their way back from work or at home. One health assistant said she would ask her children to do the data entry at home. Those health assistants with eye-sight problems complained that they found it difficult to read small letters on the black and white screen, and said that a colour screen would have been better. Also, while standing with the PDAs in the sun made it difficult for the health assistants to see the screen clearly. Since they had to deal with long lines of people waiting to get their attention, they found it troublesome to spend time trying to enter data into the PDAs. Added to the poor quality of the MPHS and the database, there were technical problems during the uploading of data into the database. The PDAs were perceived to be costly, and the health assistants were scared that they might lose them and they would have to pay for it from their pockets. Due to these and other problems, PDA use has now stopped. A health worker shared her experience:

The PDAs which were supplied were low in memory. When we started using PDAs it was taking a lot of time to search the name, or open the screen or save the entry. When we had a long queue of patients, mothers holding their small children waiting for us to give the children immunization shots, if we were operating a PDA which was slow, the patients would be frustrated. If I don't attend to them, they will go away. So, very soon I realised this PDA was not much of use when I was on the job. So I resorted back to writing in my old book. What I did was I would go home and then again enter all the day's entry into the PDA. To tell you frankly, though it has helped me in getting schedules, it has also increased my work load.

A doctor working at a health centre also shared a similar opinion on the reasons for nonuse of the devices:

The health assistants are reluctant to use PDAs because the operating system is such that if the power is discharged or if a health assistant forgets to recharge it, all the data will vanish. She has to come to the PHC and upload the data from the desktop to the PDA. In this division, the reason for reluctance of health assistants to use computers is that we got the PDAs first and desktops later. So whenever data vanished, the software engineers would take the PDAs to Hyderabad (the state capital) and would feed the data there into the PDA. The second problem was memory. Third problem is when we said the memory is not sufficient, they changed the software so only the target population is there in the PDA. There are many software problems because of this. The health assistants who are really interested even today are not able to use them because some records are not captured properly.

Another doctor who was involved in training and supervising the project said both technical and nontechnical factors contributed to the failure of the PDA project:

The processing capacity of the PDA was slow. The intention of giving PDAs was that the health worker would give immunisation and immediately enter that data in the field. But that did not happen in practice. People won't wait. So they wrote manually in their registers. Once the immunisation session was completed, the health workers would have to again enter the data, come to the health centre and then upload the data (into the desktop). Only after that could they get the schedule. In that schedule few names would be missed because some records would be rejected. So due to all these issues health assistants were not able to use the PDAs comfortably. If these technical problems could have been solved, the percentage (of health assistants using the PDAs) would have been about 70-75%, but not 100%, because you cannot make everyone use that device. Also, battery has to be replaced after every two years or so. All these issues were not taken into consideration.

The reason to do a pilot is to try out a technology on a small scale in order to identify problems and resolve them before making a large investment. However, about 200 PDAs were brought by the IT company all at once. The consultant of the IHC project said:

The irony of it is we started the project in three PHCs initially as a pilot. So they could have procured the PDAs only for those 3 PHCs. When we raised this problem of memory they could have procured higher memory PDAs later. After having gained experience during this pilot phase they could have bought the rest of the PDAs. Anybody would do that. But the software company said they have bought all the PDAs at once. For half of the district, it was 200+ PDAs.

The poor quality of the survey database, recurring software problems, insufficient memory capacity, and the absence of timely and ongoing technical support and maintenance gradually led to nonuse of these devices on the job. The health assistants who were trained and were enthusiastic initially gradually lost interest and reverted back to their earlier manual system of recording health data. The project was neither sustainable nor was it scalable because of the costs involved. Out of 200, only about 9 health assistants were using PDAs in March 2004 during a field visit by one of the authors. During the visit, the health staff took out the PDAs from locked cupboards and showed the researcher blank PDAs that had not been used for months. In some other PHCs, we saw the children of the health assistants playing with these devices.

The aim of the project was to provide support tools that would allow health assistants to reduce the time spent doing paperwork. However, in practice, it actually increased their work. The second aim was to increase the accuracy of the data flowing up from the health assistants; however, this could not be achieved because the quality of the multipurpose survey data itself was poor. The third aim was to provide a means for getting health-care data at the village level into electronic form and generate reports. This was too ambitious to achieve because shifting from a manual to an electronic form was difficult for the reasons cited above. Finally, the project aimed to provide health assistants with information that allowed them to provide more targeted and effective services to the people. Health assistants need more than just support tools, however, but also support to use those tools. In the absence of this, they were not able to use these devices effectively. As the consultant said, "it works only if the supporting staff trouble-shoot problems promptly and in a reasonable period. But this did not happen, so gradually the health assistants lost interest and developed a negative attitude towards computers." So when plans to expand the computerisation project to the whole state were being drawn, it was decided not to opt for PDAs because of these problems, and also because it was considered too expensive.

Family Health Information Management System

The FHIMS project was conceived based on the experience of the IHC project. While the IHC project covered only one aspect of the work of the health assistants (family welfare services), it was decided by the FWD that a comprehensive HIS was necessary, and that it should include the different activities of PHCs including family welfare services, various health services (disease control, for example), and administrative aspects (such as budgets and logistics support). The department in its official circular (CFW, 2003) outlined the key objectives of this project as follows:

- To provide a name-based follow-up of family welfare services including antenatal services and immunization, and the early identification and timely referral of high-risk antenatal cases
- To improve the full immunization rates and thereby contribute to the reduction of infant mortality rates and maternal mortality rates
- To facilitate the health assistants to easily get schedules of services to be rendered in each habitation during a month; that is, to

provide schedules on pregnant women and children to be visited by the health assistants in their respective subcentres
- To reduce the burden of manual record keeping by the field staff and all the higher levels
- To contain the spread of communicable diseases and blindness by tracking incidences of diseases
- To improve the functioning of PHCs by facilitating effective service delivery
- To streamline inventory and infrastructure management at the health centres
- To manage career and training issues of field personnel

The project was conceived over the following phases: the pilot at the Nalgonda district (December 2001-November 2002), and the state-wide replication (June 2003-ongoing). We discuss these two phases followed by a summary of the identified implementation challenges.

The Pilot Phase of FHIMS at the Nalgonda District

The pilot phase of the FHIMS project was started in 67 PHCs in December 2001 about a year after the IHC project in the same district. A software engineer who had been involved throughout the design and development of IHC and also FHIMS explained that elaborate consultations were held with health officials at different levels, including health assistants, during the design of the software. However, when we spoke to many doctors, they typically said that the process had been largely top down with limited involvement of PHC-level doctors.

After the software was installed in all the PHCs, the department appointed data-processing operators (data operators) on a temporary basis for a period of 9 months to do the data entry and to train the staff during this transition period. Although data operators initiated the data-entry process, they did not seem to have the capacity to take on the role of change agents and to motivate the health staff. Although the district administration had instructed every PHC to identify a competent person as a system administrator who could act as a leader and be trained by the data operators from the beginning, this did not happen. Also, the role of systems administrators was not clearly defined, and adequate guidelines were not given to them on how to manage their regular work (for example, as lab assistants) with these additional responsibilities. Additionally, no incentives were provided to concurrently carry on both this task and that of a system administrator. The system administrators also complained that the training that was provided was inadequate and focused primarily on computer awareness with hardly anything on the FHIMS and nothing on health management itself. A coordinator for FHIMS who was involved in conducting training sessions confirmed this exclusion:

Initially, all the cadres were trained in all the modules. That was not really useful. For example, a health assistant need not know about the budget module, or vehicle or personnel module. If she knows about the family welfare module and one or two national health programmes, that should be enough. Similarly, a pharmacist need not know about budget or family welfare modules. If he knows about his stores, inside out, it is more than enough. It won't take more than 2 days. He will be more than happy if he is thoroughly trained for 2 days on his module. But they were called for 5 days and were taught all the modules. By the time they came to their module, they had probably lost interest.

The doctors said they were not fully involved in the project from the beginning and were unaware about the aims of the project. A FHIMS coordinator, who is also a doctor said:

From the beginning, the doctors did not have ownership of the data… since the time the survey started. Training on updating the registers was given only to the health assistants, and the doctors did not know what the training was about. Had the doctors at the PHCs been told about the registers, all about the project, and given guidelines as to how a health assistant has to go about her work, this is where you as a doctor have to guide them…then things would have been different. However, the doctor was bypassed. The data entry was done at a central point (not at PHCs). The doctor lost track there again. That's how they drifted slowly. They could have probably fallen back on track when the software was installed, but that didn't happen due to many reasons. Hence the doctors did not own the project.

Online data transfer took place initially for sometime, but then stopped. The coordinator of FHIMS explained why:

Online data transfer was going on in 50% of the PHCs for sometime. But the thing is, the district administration is still taking only the manual reports. It is not taking the reports generated from FHIMS because all the data are not being entered into FHIMS. So, whatever report you get from that (software), it is not complete. We can generate only one report, Form B from FHIMS. But that had a few errors. And even if you could correct one or two errors with a pen, you need the other data. And where will the other data come from?

In March 2003, the authors visited 12 PHCs, one community health centre, and the district office. The health staff complained that many software- and system-related problems were identified but not addressed. In many PHCs, the health staff did not know the passwords, and did not know how to burn CDs and save the data. At the time of our visits, a majority of the health staff were not using the computerised system, were not entering and updating the health data regularly, and hence were not generating computerised schedules or reports. The projects in most PHCs had stopped when the data operators left after their temporary tenure.

STATE-WIDE EXPANSION OF FHIMS

State-wide replication of the project involved the extension of the Nalgonda project to all the other 22 districts and 1,319 PHCs. The expansion phase of the project, which started in 2003, involved the computerisation of all the PHCs, district medical and health offices (DM&HOs), and the offices of the Director of Health and the Commissioner of Family Welfare at the state level (Circular I, CFW, 2003, p. 1). The replication involved the conduct of a household survey (to replace the MPHS data as was done in Nalgonda), the installation of computers and the FHIMS, the entering of survey data, training for the health staff, and the initiating of the other activities (such as identifying private agencies for printing MPHS books or team building) necessary for starting the project.

Based on the experience of the pilot, many changes were made during the replication phase. Survey registers were printed with an index and in alphabetical order. A team of 16 people were trained in each district, who in turn trained the health staff at the district level on how to conduct the household survey. Doctors were involved and system administrators were identified. Videoconferencing facilities were used regularly by the commissioner to coordinate and give instructions to health officials in the district, and to obtain status reports on the data-entry process. A number of circulars were sent to every PHC regularly, giving detailed instructions at different stages of implementation.

The computers were introduced in September 2003 in the PHCs, and the initial plan was to start the project by December 2003. In July 2005, one of the authors and his team conducted a situational analysis of the status of the implementa-

Table 1. Summary of the implementation challenges

*	Improving data-collection approaches
*	Developing a culture of using information to support local action
*	Providing adequate training not just in using computers to generate regular reports, but also in interpreting data
*	Providing on-site training (at PHCs) for the health staff on a regular basis
*	Adopting an iterative process to improve technical features of the software
*	Managing the workload of the staff
*	Getting health staff to start doing data entry rather than data-processing operators
*	Initiating collective participation to facilitate ownership of the systems
*	Motivating health staff to change existing work practices
*	Encouraging system administrators and doctors to take on the role of change agents
*	Building the capacity of health managers for evidence-based decision making
*	Providing timely hardware and software support
*	Addressing infrastructural issues
*	Switching over to online reporting
*	Sustaining the project over time by involving officials at the state, district, and local levels
*	Managing interdependencies (for example, the building of the health database was dependent on MPHS, which was developed by another department)

tion of FHIMS in 84 PHCs in another district (called Chittoor). This evaluation report, which was provided to the collector of the district, the DM&HOs, and the state-level authorities, identified a number of ongoing problems in the implementation of FHIMS contributing to it being far behind schedule. The evaluation identified that all the PHCs had reported software bugs of varying levels of complexity. Although household data had been entered, service data remained grossly incomplete. None of the PHCs were generating any of the reports that were needed to be sent to the district. This was because, except for one report, other reports could not be generated through the FHIMS. Although in a few cases schedules for health assistants were being printed, the demand for these schedules did not come from the health assistants themselves, but the schedules had been printed by the data operators (before they had left) and given to the health assistants. The study also revealed that on average only three or four people per PHC had received training in computer awareness through the 5-day training programme. No formal training was given to the PHC staff on FHIMS. In some cases, the data operators had provided some basic training to the system administrators. In more than 50% of the PHCs, none of the staff knew the username and password to enter the FHIMS modules after the departure of the data operators.

In summary, various factors have been identified that have impeded the implementation of the HIS initiatives discussed. These complexities are sociotechnical in nature. While some of these problems are expected given the extreme complexity of the projects, some, we believe, can be addressed by providing increased care and attention to the sociotechnical issues. In Table 1, we summarize some of these key challenges, and in the following section, we discuss how they were addressed.

DISCUSSION

One of the key recommendations of WHO (2005) in *Health and the Millennium Development Goals* is to improve the information basis on which health management decisions are made. The report notes, "making available information concerning the location, functioning, and performance of health services should also improve transparency and accountability in the management of the health sector" (p. 79). Accurate information can support better management and enhance the efficiency of health-care delivery. HIS can contribute to human development by supporting health management. Jacucci, Shaw, and Braa (2005) argue that the overall sustainability of a standards-based HIS is dependent on the quality of data and the skillful use of data at the level of collection. Local sustainability of the HIS requires the information to be of relevance not just for the managers, but also at the local level. This requires training health staff not just on the mechanics of using software, but also on public health-related topics of information use, data quality, significance of health indicators, validation, analysis, and interpretation of data, and linking these with interventions on the ground.

The implementation of HIS is not just a matter of introducing technology, but is also dependent on the social, organisational, and political issues that need to be addressed on an ongoing basis. Several authors have emphasized the need to understand and analyse information systems from a sociotechnical perspective (Coakes, Willis, & Lloyd-Jones, 2000; Kling & Lamb, 2000; Walsham, 1993). For example, Kling (1987) and Kling and Scacchi (1982) utilize Web models to take into account the social relations between the set of participants, the infrastructure available for its support, and the previous history of commitments in the organisation around computer-based technologies. Walsham (1993) provides an analytical framework for understanding the mutually shaping linkages that exist between content, social context, and social processes of IS. Over the years, several authors have both applied Walsham's structural framework and also expanded it by including various facets of the economic, cultural, and political context (e.g., Avgerou & Walsham, 2000; Jarvenpaa & Leidner, 1998; Madon, 1993; Sahay, 2000; Walsham, 2001). Avgerou and Madon (2004) argue "for broadening of the perspective of situated research to make explicit the contextual origins of the meanings, emotional dispositions, and competencies actors bring to bear on the interactions that constitute these processes" (p. 176).

Even though a participatory approach was stated to have been adopted to a certain extent during the design and implementation of the FHIMS, in practice, this did not evolve into effective and sustainable partnerships resulting in end users owning the new system. As an example, the HISP team conducted a situation analysis in the 84 PHCs of the Chittoor district in August 2005 after the contracts of the data-processing operators responsible for the FHIMS implementation were discontinued. The analysis revealed that in about 50 of the 84 PHCs, the FHIMS system had not been used since the departure of the operators because the PHC staff did not even know the password to log into the system, indicative of the absence of a partnership between the implementing team and the end users, and the lack of local ownership of the system. This raises the need to adopt complementary or alternative approaches that can facilitate the creation of local capacity and ownership. This requires the trainers to play multiple roles from building technical capacity to motivating the staff, and facilitating the change processes at multiple levels (Tomasi, Facchini, & Maia, 2004). Building these capacities is a slow and gradual process, but they are crucial and fundamental for the collective ownership of the project. This requires the need to shift the focus from being primarily on technology to the various surrounding processes required to strengthen the informational basis of health management. Some ways in which these challenges are being addressed in the FHIMS case is now discussed.

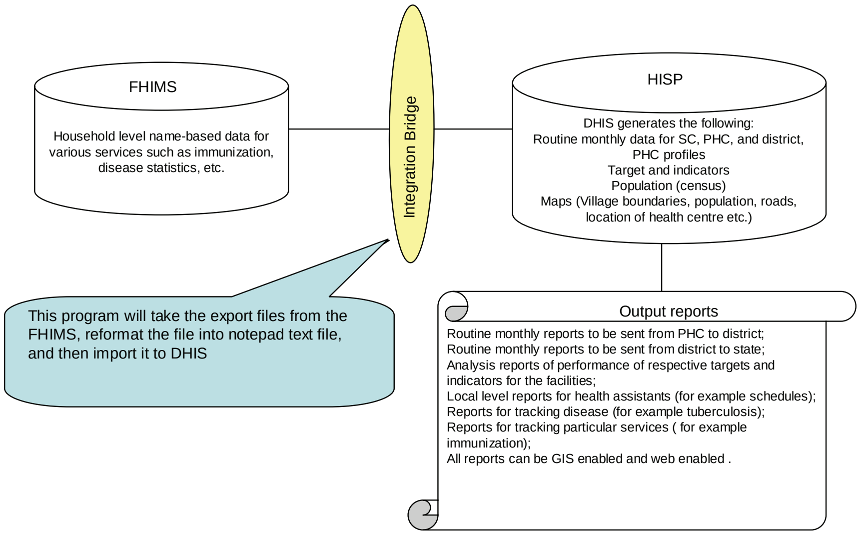
Shifting Focus from Technology to Information: Empirical Experiences from IHIMS

In order to address the various implementation challenges, the State Health Department endorsed in July 2005 a project called Integrated Health Information Management System (IHIMS) through a Memorandum of Understanding (MoU) with HISP India[3] (a not-for-profit society set up by the University of Oslo). HISP is an ongoing, large-scale action research initiative that operates as a global network across a number of developing countries. HISP was started in India in December 2000, and was operating in some limited areas parallel to the FHIMS project, which had a state mandate. Since FHIMS was the official state project, and HISP was seen primarily as a university research and development project, HISP was always vulnerable to being terminated by the state in favour of FHIMS. As a strategy to protect itself against such an event, the HISP management proposed to the state the IHIMS project, which would integrate the individual capabilities of the FHIMS and HISP projects and develop synergies to contribute to the overall health information management in the state. The integration logic argued by HISP was informational rather than technological. While FHIMS was a name-based software, the District Health Information Software (DHIS), a not-for-profit customised open-source software developed by HISP, was a facility-based system. By linking the two technological systems, it was argued that an integrated informational analysis could be carried out both at the level of individuals and health facilities. Also, since the DHIS was linked to the Web and GISs, the integrated system could provide visual representations of the information on the Web and maps, something not currently possible with FHIMS.

Technically, the integration was developed in the following way. With its focus on name-based data, FHIMS was capable of generating subcentre-based reports and schedules to help the health staff identify which households to visit and when. The name-based household data was entered into the FHIMS, through which the subcentre-based reports were generated. The data were then aggregated and imported into the DHIS by facilities (subcentres) and processed, which resulted in the generation of various analysis reports, including on the GIS and Web. The integration was thus proposed to convert some of the raw (household based) data into useful information through the use of analytical tools, and to improve the representation of the data using GIS-generated maps. In arguing for the integration, the HISP team emphasized that map-based reports would be useful for district- and state-level officials, for example, in resource-allocation decisions (such as opening a new facility or buying an ambulance). The conceptual schema of the integration proposed, and subsequently accepted, through the MoU under the IHIMS project is depicted in Figure 2.

The IHIMS implementation commenced in August 2005, and the first phase of the project was a situation analysis to try to understand the status of the earlier FHIMS implementation and the underlying challenges. A key problem identified was that the operators who had been working previously in the PHCs under the FHIMS project had operated as independent islands and had not involved the health assistants in the PHCs. As a result, there was little development of local expertise and ownership of the system. As mentioned earlier, in nearly 50 of the 84 PHCs, even

Figure 2. The FHIMS-HISP integration



the passwords to log on to the system were not known to the staff in clinics there. In contrast to the approach of the operators whose focus was on data entry, the HISP's emphasis was on system facilitation whereby the data entry and report generation was not done by the trainers, but by the health staff themselves, supported and facilitated by the trainers. Also, there was a difference in the focus of the training content, which was not limited to the mechanics of the software, but also to issues of how information can be made useful to support local action.

After the situation analysis in August, the HISP trainers were assigned to a cluster of four to five PHCs and were responsible for providing ongoing support to the staff on site in each of the PHCs. This on-site training approach was radically different from the earlier FHIMS training strategy of one large district-level 3- to 5-day training programme followed by literally no support to the staff in the course of their everyday work. From August to November, a key focus of the IHIMS project was on completing the huge amounts of backlog in data entry in the FHIMS database, and trying to change the mindsets of the health staff from FHIMS to IHIMS, which included both DHIS and FHIMS. At the end of November 2005, an evaluation of the IHIMS indicated that 68 of the 84 PHCs had completed the backlog of data entry, and they had generated the Form B, mother and child report, and also the Forms A and C, for communicable diseases and other health services. The way IHIMS worked was that data for the Form B was entered into the FHIMS, and the report accordingly generated. However, the data required to generate Forms A and C were directly entered into the DHIS database, and the reports were generated through the DHIS. The FHIMS technically was not capable of developing the Forms A and C, but through the integration of the two systems, the whole suite of reports could be generated. Another evaluation at the end of December indicated that the number of PHCs that had generated the Forms A, B, and

C was now 75, and the remaining 9 had serious structural problems (lack of staff or very large PHCs). Also, some of the PHCs, in addition to the routine reports of Forms A, B, and C, had also started to generate some analysis reports such as those for conducting inter-subcentre comparison. The HISP's training focus in the future was to be increasingly on the analysis reports and the use of information for action.

The relative success of the IHIMS programme can be attributed to three key reasons. First, by linking up the FHIMS and DHIS, the state government gave opportunities for the different people involved to pool their strengths and capacities rather than fight with each other. Second, the individual capabilities of the two systems could be innovatively combined so that the FHIMS software was used for generating subcentre-based reports while the DHIS provided the facility-specific reports and also incorporated different forms of representation including the Web and GIS. Third, alternative training approaches could be included through the integration. Unlike the FHIMS project, in which training focused primarily on the use of computers, HISP emphasized the use of information, including the generation of reports, the analysis of data, and the use of GIS to make visible spatial correlations in health data. Also, HISP tried to develop the capacity of the health assistants and other staff at the PHC in the local use of information. This effort is part of a larger and long-term process of developing an information culture where data is not just seen as a mindless effort to fulfill the needs of an uncaring bureaucracy, but a resource that can strengthen the everyday local work. This approach is compatible with Sandiford et al.'s (1992) argument:

the ability to analyse and interpret data is indispensable but insufficient to overcome the inertia of the status quo to ensure that the information is translated into decisions and action….The need to tailor the presentation of information to the intended audience must be recognised. (p. 1085)

The analysis capabilities of the DHIS could be emphasized through various preprogrammed tools that provided the user with the ability to make comparisons and present data in the form of graphs, charts, and maps. Similarly, GIS provided particular analytical tools that enabled the mapping of health status and diseases, programme planning, displaying health-care coverage, planning health infrastructure and maintenance, and comparing performance indicators across PHCs (Sauerborn & Karam, 2000).

Prior research in the domains of IS and HIS, both in developed and developing countries, has emphasized that integrating systems is no easy task as it involves aligning social, technical, and political linkages, and fostering coalitions and building synergies (Chilundo, 2004; Hanseth, Ciborra, & Braa, 2001). We agree with Chilundo, who has argued that the integration of ISs is not only a technical exercise, but involves the creation of a heterogeneous network comprised of various groups of people, technological artefacts, medical practices, and different information systems. While, at one level, integration can be seen as an attempt to reduce some of the heterogeneity in the system arising from a multiplicity of information systems, by adopting a modular approach to integration, heterogeneity can be encouraged rather than neutralized. While both the DHIS and FHIMS were capable of operating independently, in combination they were able to produce the whole set of reports required by the health department. Integration of not only the technical systems also implied bringing together different training approaches and also the pooling of financial and human resources. For example, the system administrators, a post created to support the FHIMS implementation, now were made the focal persons under the IHIMS and became responsible for the generation of the reports from both the DHIS and FHIMS systems.

In summary, we argue that the integration of systems can provide potential opportunities to address some of the implementation challenges discussed earlier. However, the philosophy underlying the integration attempts should not be to neutralize or eliminate one system, but to see how synergies can be created through drawing upon the positive aspects of both the systems and associated resources. For example, through IHIMS, the positive aspects of FHIMS, which were the name-based service schedule and summary reports, and DHIS, which was the facility-based reports, could be meaningfully incorporated toward the broader goals of health management.

CONCLUSION

The case studies presented earlier provide evidence of the importance of adopting a sociotechnical and integrated approach to the development of HIS. Too often, the introduction of ICTs is equated with development as an end in itself rather than as a means for supporting health management, which can in practice support development. These projects end up as largely computerisation projects and unsuccessful ones at that, rather than effective health-management initiatives. The introduction of computers, and the development and implementation of HIS in itself are not sufficient. Rather, the focus should be on how to develop and improve the informational base whereby information comes to be seen as a resource at all levels—at the local, district, state, and national levels—rather than just for bureaucracy.

There is typically an overemphasis on technical aspects during the implementation of HIS, and a relative neglect in addressing problems of information support that is necessary for better health management. The challenge is to move away from this pattern and to adopt a more integrated, holistic, and comprehensive approach to reforming informational and organisational practices. Introducing a HIS, particularly a name-based system, without addressing issues related to information generation, management, and use (analysis, interpretation, and evidence-based decision making) does not solve existing problems of poor quality of data, or the manipulation of figures, because health staff working in the field may have their own logic, stakes, and priorities that might not align with the goals of the promoters. That is, introducing information technology and systems does not in itself necessarily lead to better information practices or better health management. The challenge is not just to have adequate information, but to make better use of it and have a greater ability to act on it.

A grand vision to have a unified system (one-for-all purposes), as in the case of MPHS and FHIMS, may be attractive, but it brings with it numerous problems and is difficult to put effectively in practice. Instead, a more realistic approach should attempt to recognise what exists, what works well, and at what costs new systems are being built. While the name-based approach of the FHIMS is useful at the community level to track patients and provide targeted health care, HISP emphasizes the use of analytical tools for better health management and for capacity building of health staff. Initiating synergies between the FHIMS and DHIS through an integrated approach, as is being empirically attempted, provides opportunities to develop more sustainable health information systems that either of the systems by themselves cannot provide. These opportunities come through increased technical functionalities to serve different needs, a pooling of financial and human resources, and the incorporation of different and complementary training approaches.

ACKNOWLEDGMENTS

The authors would like to thank Professor Lucy Suchman for her helpful comments on previous drafts of this chapter. We would also like to thank the officials at the office of the Commissioner of Family Welfare, Andhra Pradesh, all the health staff and members of HISP India, and the FHIMS team for their support.

REFERENCES

Anderson, J. G., & Aydin, C. E. (1997). Evaluating the impact of health care information systems. *International Journal of Technology Assessment in Health Care, 13*(2), 380-393.

Avgerou, C., & Madon, S. (2004). Framing IS studies: Understanding the social context of IS innovation. In C. Avgerou, C. Ciborra, & F. Land (Eds.), *The social study of information and communication technology: Innovation, actors, and contexts* (pp. 162-182). Oxford, United Kingdom: Oxford University Press.

Avgerou, C., & Walsham, G. (Eds.). (2000). *Information technology in context: Studies from the perspective of developing countries.* Aldershot, United Kingdom: Ashgate.

Bodavala, R. (2002). ICT applications in public healthcare systems in India: A review. *ASCI Journal of Management, 31*(1-2), 56-66.

Braa, J., Heywood, A., & Shung King, M. (1997). District level information systems: Two cases from South Africa. *Methods of Information in Medicine, 36*(2), 115-121.

Braa, J., Monteiro, E., & Sahay, S. (2004). Networks of action: Sustainable health information systems across developing countries. *MIS Quarterly, 28*(3), 337-362.

Braa, J., & Nermukh, C. (2000). Health information systems in Mongolia: A difficult process of change. In C. Avgerou & G. Walsham (Eds.), *Information technology in context: Studies from the perspective of developing countries* (pp. 113-133). Aldershot, United Kingdom: Ashgate.

Braa, J., Macome, E., Mavibe, J. C., Nhampossa, J. L., da Costa, J. L., Manave, A., et al. (2001). A study of the actual and potential usage of information and communication technology at district and provincial levels in Mozambique with a focus on the health sector. *Electronic Journal in Information Systems for Developing Countries, 5*(2), 1-29.

Chilundo, B. (2004). *Integrating information systems of disease-specific health programmes in low income countries: The case study of Mozambique.* Unpublished doctoral dissertation, University of Oslo, Faculty of Medicine, Oslo, Norway.

Coakes, E., Willis, D., & Lloyd-Jones, R. (2000). Graffiti on the long wall: A sociotechnical conversation. In E. Coakes, D. Willis, & R. Lloyd-Jones (Eds.), *The new sociotech: Graffiti on the long wall* (pp. 3-12). London: Springer-Verlag.

Gupta, D. M., & Chen, C. L. (1996). Introduction. In D. M. Gupta, C. L. Chen, & T. N. Krishnan (Eds.), *Health, poverty & development in India* (pp. 1-22). Delhi, India: Oxford University Press.

Hanseth, O., Ciborra, C., & Braa, K. (2001). The control devolution: ERP and the side effects of globalization. *The Data Base for Advances in Information Systems, 32*(4), 34-46.

Heeks, R., Mundy, D., & Salazar, A. (2000). Understanding success and failure of healthcare information systems. In A. Armoni (Ed.), *Healthcare information systems, challenges of the new millennium* (pp. 96-128). Hershey, PA: Idea Group.

Helfenbein, S., et al. (1987). *Technologies for management information systems in primary health care* (issue paper). Geneva, Switzerland: World Federation of Public Health Associations.

Jacucci, E., Shaw, V., & Braa, J. (2005). *Standardization of health information systems in South Africa: The challenge of local sustainability.* Proceedings of the Eighth International Working Conference of IFIP WG 9.4: Enhancing Human Resource Development through ICT, Abuja, Nigeria.

Jarvenpaa, S. L., & Leidner, D. E. (1998). An information company in Mexico: Extending the resource-based view of the firm to a developing country context. *Information Systems Research, 9*(4), 342-361.

Kling, R. (1987). Defining the boundaries of computing across complex organisation. In R. Boland & R. Hirschheim (Eds.), *Critical issues in information systems research* (pp. 307-362). New York: Wiley.

Kling, R., & Lamb, R. (2000). IT and organizational change in digital economies: A socio-technical approach. In B. Kahin & E. Brynjolfsson (Eds.), *Understanding the digital economy: Data, tools and research* (pp. 295-324). Cambridge, MA: The MIT Press.

Kling, R., & Scacchi, W. (1982). The web of computing: Computer technology as social organisations. *Advances in Computing, 21*, 1-90.

Lippeveld, T., Sauerborn, R., & Bodart, C. (2000). *Design and implementation of health information systems.* Geneva, Switzerland: World Health Organisation.

Lippeveld, T. J., Foltz, A., & Mahouri, Y. M. (1992). *Transforming health facility-based reporting systems into management information systems: Lessons from the Chad experience* (Development Discussion Papers No. 430). Cambridge, MA: Harvard Institute of International Development.

Madon, S. (1993). Introducing administrative reform through the application of computer-based information systems: A case study in India. *Public Administration and Development, 13*(1), 37-48.

Madon, S., Sahay, S., & Sahay, J. (2004). Implementing property tax reforms in Bangalore: An actor-network perspective. *Information and Organization, 14*, 269-295.

Madon, S., Sahay, S., & Sudan, R. (2005). *E-governance and health care in Andhra Pradesh, India: Reflecting upon the mutual linkages between macro and micro.* Unpublished manuscript.

Mohanty, P. K., Jones, K., & Rao, S. J. (2004). *Good governance initiatives in Andhra Pradesh 2003.* Hyderabad, India: Centre for Good Governance.

Mosse, E., & Nielsen, P. (2004). Communication practices as functions, rituals and symbols: Challenges for computerization of paper-based information systems. *The Electronic Journal of Information Systems in Developing Countries, 18*(3), 1-17.

Mosse, E., & Sahay, S. (2003). Counter networks, communication and health information systems: A case study from Mozambique. In *Proceeding from IFIP TC8 & TC9/WG8.2 Working Conference on Information Systems Perspectives and Challenges in the Context of Globalisation*, pp. 35-51.

Nhampossa, J. L. (2004). Strategies to deal with legacy information systems: A case study from the Mozambican health sector. *Innovations through Information Technology: Proceedings of the 15th Information Resources Management Association International Conference*. Retrieved March 21, 2005, from http://folk.uio.no/leopoldo/Publications/Papers/IRMA_USA.pdf

Nhampossa, J. L., & Sahay, S. (2005). Social construction of software customisation: The case of health information systems from Mozambique and India. In *Proceedings of the Eighth International Working Conference of IFIP WG 9.4: Enhancing Human Resource Development through ICT, Abuja, Nigeri*a.

Office of the Commissioner of Family Welfare. (2003, May 13). *Various official circulars. Circular – I* (Rc No. FHIMS/9511/2003). Hyderabad, India: Author.

Opit, L. J. (1987). How should information on healthcare be generated and used? *World Health Forum, 8*, 409-438.

Piotti, B., Chilundo, B., & Sahay, S. (2005). *An institutional perspective on health sector reform* *and the process of reframing health information systems: Case studies from Mozambique* (Working Paper 9). Retrieved September 14, 2005, from http://www.ifi.uio.no/forskning/grupper/is/wp/092005.pdf

Quraishy, Z., & Gregory, J. (2005). Implications of (non) participation of users in implementation of the Health Information System Project (HISP) in Andhra Pradesh: Practical experiences. In *Proceedings of the Eighth International Working Conference of IFIP WG 9.4: Enhancing Human Resource Development through ICT, Abuja, Nigeria*.

Sahay, S. (1998). Implementing GIS technology in India: Some issues of time and space. *Accounting Management and Information Technologies, 8*, 147-188.

Sahay, S. (2000, May 24-26). *Information flows, local improvisations and work practices.* Paper presented at the IFIP WG 9.4 Conference, Cape Town.

Sahay, S. (2001). Introduction. *The Electronic Journal of Information Systems in Developing Countries, 5*, 1–6.

Sahay, S., & Walsham, G. (2005). *Scaling of health information systems in India: Challenges and approaches*. Proceedings of the Eighth International Working Conference of IFIP WG 9.4: Enhancing Human Resource Development through ICT, Abuja, Nigeria.

Sandiford, P., Annett, H., & Cibulski, R. (1992). What can information systems do for primary health care? An international perspective. *Social Studies and Medicine, 34*(10), 1077-1087.

Sauerborn, R., & Karam, M. (2000). Geographic information systems. In T. Lippeveld, R. Sauerborn, & C. Bodart (Eds.), *Design and implementation of health information systems* (pp. 213-224). Geneva, Switzerland: World Health Organisation.

Sauerborn, R., & Lippeveld, T. (2000). Introduction. In T. Lippeveld, R. Sauerborn, & C. Bodart (Eds.), *Design and implementation of health information systems.* Geneva, Switzerland: World Health Organisation.

Singh, M. (2005). *PM launches National Rural Health Mission*. Retrieved August 31, 2005, from *http://pmindia.nic.in/speech/content.asp?id=101*

Tomasi, E., Facchini, L. A., & Maia, M. F. S. (2004). Health information technology in primary health care in developing countries: A literature review. *Bulletin of the World Health Organization, 82*(11), 867- 869.

Walsham, G. (1993). *Interpreting information systems in organisations.* Cambridge: John Wiley & Sons.

Walsham, G. (2001). *Making a world of difference: IT in a global context.* Chichester, United Kingdom: John Wiley.

World Health Organisation (WHO). (1987). *Report of the interregional meeting on strengthening district health systems based on primary health care* (WHO/SHS/DHS/87.13). Unpublished document.

World Health Organisation (WHO). (2005). *Health and the millennium development goals.* Avenue Appia: Author.

ENDNOTES

1. A mandal is an administrative unit.
2. A habitation is a group of households in a village. It is also a unit of data for the census.
3. For more on HISP activities in India, see Sahay and Walsham (2005) and Braa, Monteiro, and Sahay (2004).

ANMs (auxiliary nurses and midwives) and MPHAs (multipurpose health assistants) are synonyms for health assistants.
The quotes that are presented here are not verbatim but with slight alterations to clarify language usage.

GLOSSARY

AP: State of Andhra Pradesh, India

ANMs: Auxiliary Nurses and Midwives or health assistants

CFW: Commissioner of Family Welfare

DMHO: District Medical and Health Office

DHIS: District Health Information System

FHIMS: Family Health Information Management System

FWD: Family Welfare Department

GIS: Geographic Information System

HISP: Health Information System Programme

HIS: Health Information System

IHC: India Health Care Project

IHIMS: Integrated Health Information Management System

MPHS: Multipurpose Household Survey

PDA: Personal Digital Assistant

PHC: Primary Health Centre

SC: Subcentre

This work was previously published in Information Communication Technologies and Human Development: Opportunities and Challenges, edited by M. Gascó-Hernández, F. Equiza-López, and M. Acevedo-Ruiz, pp. 266-292, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 4.15
Assessing Physician and Nurse Satisfaction with an Ambulatory Care EMR:
One Facility's Approach

Karen A. Wager
Medical University of South Carolina, USA

James S. Zoller
Medical University of South Carolina, USA

David E. Soper
Medical University of South Carolina, USA

James B. Smith
Medical University of South Carolina, USA

John L. Waller
Medical University of South Carolina, USA

Frank C. Clark
Medical University of South Carolina, USA

ABSTRACT

Evaluating clinician satisfaction with an electronic medical record (EMR) system is an important dimension to overall acceptance and use, yet project managers often lack the time and resources to formally assess user satisfaction and solicit feedback. This article describes the methods used to assess clinician satisfaction with an EMR and identify opportunities for improving its use at a 300-physician academic practice setting. We administered an online survey to physicians and *nurses; 244 (44%) responded. We compared physician and nurse mean ratings across 5 domains, and found physicians' satisfactions scores were statistically lower than nurses in several areas (p<.001). Participants identify EMR benefits and limitations, and offered specific recommendations for improving EMR use at this facility. Methods used in this study may be particularly useful to other organizations seeking a practical approach to evaluating EMR satisfaction and use.*

INTRODUCTION

The degree of interest and momentum in furthering the widespread adoption and use of electronic medical record (EMR) (or electronic health record systems) is at an all time high in the United States. Healthcare providers, purchasers, payers and suppliers are all looking to the EMR as a tool to help promote quality, enhance patient safety, and reduce costs. Despite this energy, recent estimates indicate EMR adoption rates in ambulatory care remain in the 15-20% range (Hillestad et al., 2005). Cost, lack of uniform interoperability standards, limited evidence showing use improves patient outcomes and clinician acceptance are among the barriers to widespread EMR adoption (Bates, 2005). Those who have overcome these initial hurdles and made the transition from a paper-based medical record system to an EMR often lack the time, resources and expertise to evaluate the system's impact on the organization, including clinician use and satisfaction with the system (Anderson & Aydin, 2005; Wager, Lee, & Glaser, 2005).

Use and satisfaction are two key measures of the success of any information system (DeLone & McLean, 2003), including EMR system success (Anderson & Aydin, 2005). Various researchers have assessed physician use and satisfaction with the EMR (Sittig, Kuperman, & Fiskio, 1999; Gadd & Penrod, 2001; Penrod & Gadd, 2001; Likourezos et al., 2004; Joos, Chen, Jirjis, & Johnson, 2006) and some have found that user group perspectives can differ even within the same institution (Wager, Lee, White, Ward, & Ornstein, 2000; O'Connell, Cho, Shah, Brown, & Shiffman, 2004b; Hier, Rothschild, LeMaistre, & Keeler, 2005). Assessing user reaction to the human-computer interface is also an important dimension (Sittig, Kuperman, & Fiskio, 1999; Despont-Gros, Mueller, & Lovis, 2005) in evaluating EMR satisfaction. Likewise, the timing of the evaluation study is important. Conducting formative evaluation studies during enterprise EMR implementation can be particularly useful in identifying perceived problems and making adjustments to the implementation plan or reallocating resources as needed (Friedman & Wyatt, 1997; Burkle, Ammenwerth, Prokosch, & Dudeck, 2001; Anderson & Aydin, 2005; Brender, 2006).

Our study was designed to assess physician and nurse use and satisfaction with an enterprise-wide ambulatory care EMR. We incorporated into the evaluation a means of assessing physician and nurses' reactions to the human-computer interface and also solicited their input and suggestions on how to improve the system's usefulness to our organization. This article summarizes the methods used, findings, and relevance to other organizations in the throes of implementing and evaluating EMR acceptance.

BACKGROUND

The Medical University of South Carolina (MUSC) in Charleston, South Carolina has implemented an electronic medical record (EMR) system, known as Practice Partner Patient Record® in the majority of its ambulatory clinics over the past few years. Although EMR use is not new to MUSC (family medicine has used the system since the early 1990s and internal medicine since the mid-1990s), it was not until February 2004 that we secured funding to deploy the EMR throughout the ambulatory care enterprise. When the system is fully implemented, paper medical records will have been replaced, and both primary care and specialty providers will share a *single* electronic medical record for their patients. This message was conveyed from the top down, starting with senior leadership and the dean of our medical school.

The EMR product itself has many of the attributes of a typical EMR system, including electronic health data capture, results management, decision-support, and electronic communications. However, MUSC has not yet installed enterprise-wide direct order entry nor activated the preventive care reminder functions. Disease-specific progress note templates are available for facilitating program note entry; however, direct data entry is not a requirement for using the system, and transcribed notes can also be incorporated. The system also includes a prescription writer, which allows prescriptions to be entered and tracked in the patient record while automatically checking for interactions with other medications and documented allergies. Patients receive printed prescriptions at the time of the visit; the prescriptions are not yet sent electronically to the pharmacy.

We provided formal training sessions to physicians, nurses and administrative support; however, nurses and administrative staff received at least four hours of initial training, while most physicians only had 30-45 minutes of training. Nurses and administrative staff were trained in small groups in a classroom setting with computer workstations for each individual. Physicians were offered several options for training, including small group sessions, but the far majority opted for one-on-one training with a clinical analyst. Knowing that physicians' schedules were harried, we decided to provide more intensive training to nurses and administrative staff. We felt if nurses and administrative staff were comfortable with the system, they could assist physicians in their respective departments as needed.

The purpose of the study was to gain insight into user satisfaction and determine if there were differences in user group satisfaction, and solicit formal feedback on how to improve the system's use at our organization. We defined 'user' as any MUSC employee or student who had been issued an EMR login and password. This article describes the methods used to conduct the survey and the results from our attending physician and nurse user community.

METHODS

Survey Instrument

We conducted preliminary interviews with physicians and nurses regarding their views and experiences with the EMR, but were interested in soliciting input from all EMR users at MUSC. Given the size of the user group, we decided to adopt a survey approach. We adapted the *Questionnaire for User Interaction Satisfaction (QUIS)*, developed by researchers at the University of Maryland, for use in this study after reviewing the literature, consulting with medical informatics professionals and soliciting input from EMR physician leaders at MUSC. The QUIS has been tested as a valid and reliable instrument in settings similar to MUSC (Chin, Diehl, & Norman, 1988) and was used recently by researchers affiliated with Brigham and Women's Hospital to assess physician satisfaction with its internally developed outpatient EMR (Sittig, Kuperman, & Fiskio, 1999). The QUIS has also validated for online administration (Slaughter, Harper, and Norman, 1994). The instrument assesses user satisfaction in five major domains, with four to six questions in each: (1) overall user reactions, (2) screen design and layout, (3) terminology and system messages, (4) learning, and (5) system capabilities. Recognizing that participants might respond differently depending upon whether they were assessing how long it takes to open the application or navigate within it, we separated the item "system speed" survey component into two statements—system launch speed and navigation speed. We asked participants to rate each question on a scale from 0 (the lowest) to 9 (the highest) level of satisfaction.

We included several additional items in the survey: (a) demographic data (e.g., position, age, gender), (b) use of the EMR (whether provider dictated notes or directly entered them into EMR), and (c) three open-ended questions, in which participants were asked to identify the three greatest benefits or advantages and three greatest limitations or disadvantages to using the EMR and any recommendations they have for improving the use and effectiveness of the EMR within their clinic or at MUSC. The items and scale from the QUIS were kept intact. A draft survey was sent to a pilot group of physicians, nurses and administrative staff. Suggested changes were incorporated into the final version of the survey. Institutional Review Board approval was obtained prior to conducting the study.

User Population and Survey Administration

We sent an e-mail message from one of the authors (who happened to serve as chair of the Physician Information Council) to all attending physicians and nurses listed in the EMR user database. Our sampling frame included all individuals who had been issued an EMR username and password—for a total population of attending physicians (245) and nurse EMR users (304) of 549. In the e-mail message, participants were asked to click on an embedded hyperlink to complete the online questionnaire. We gave participants the option of printing and faxing the completed survey to us (48 surveys were returned by fax). Two reminder notices were sent—one week and two weeks following the initial mailing. All participants were assured confidentiality of their responses.

Data Analysis

Survey responses were downloaded from the survey provider to an electronic database and then imported into SPSS for analysis (Windows, 2005).

Responses to open-ended questions were grouped and assigned to categories by the primary author, experienced in qualitative research. These were then tabulated within each of the categories and discussed among the participating authors.

RESULTS

Demographics of Participants and Their Use of EMR system

We received 244 completed surveys from attending physicians and nurses for an overall response rate of 44%. Forty-seven (47%) of physicians and 38% of nurses responded. 32% of participating physicians and 98% nurses are women; 85% are between the ages of 30-60, equally distributed within each decade of life. Nearly 37% of physicians and 19% of nurses reported having prior experience with other EMR systems, and 62% have used the EMR at MUSC for at least one year.

When physician participants were asked how often they dictate patient information that is eventually transcribed into the EMR, 39% reported frequently (defined as more than 50% of the time), 4% sometimes (25-50% of the time) and 58% reported rarely or never (defined as less than 25% of the time) (n=109). Almost 40% of participants reported having paper medical records pulled daily for patient visits, 14% sometimes, and 44% rarely or never (n=108). 70% reported that they access the EMR remotely on at least a weekly or daily basis.

Satisfaction with EMR

We assessed the internal consistency reliability of the five domains of the QUIS part of the survey (Table 1) and examined mean satisfaction scores for attending physicians and nurses for each item within the five domains. Overall Crohbach's alpha was 0.961.

Attending physician and nurse mean scores are shown in Table 2. Items were rated on a scale of 0 to 9, with 0 being the most negative, 9 the most positive. Both groups gave low ratings to system launch speed and clarity of error messages. Items rated most positive by both groups included system noise, ease in reading characters, consistency of terms, and the clarity of messages appearing on-screen. We calculated overall satisfaction scores for physicians and nurses using the average of the six items in the overall user reaction domain as a proxy. Overall satisfactions scores were 4.67 and 5.50 (t-test, $p < .001$) for attending physicians and nurses, respectively. We compared attending physician ratings of the other 22 individual items with those of nurses and found that the physicians' mean ratings of the EMR were statistically significantly lower in nine of the 22 items ($p<.05$). Using one-way ANOVA, we found no difference in overall satisfaction between experience groups, less than 6 months, 6-12 months, 1-2 years, and more than 2 years.

Table 1. Reliability of domains (Crohbach's Alpha)

Domain	Alpha
Overall user satisfaction	.911
Screen design and layout	.859
Terms and system information	.891
Learning	.876
System capabilities	.835
Overall	.961

We also ran a standard multivariate regression with overall satisfaction score as the dependent variable and the 22 remaining items as independent variables and found two items were significant predictors of overall satisfaction—(1) task can be performed in a straightforward manner ($p<.001$) and (2) clarity of help messages on screen ($p<.01$).

Perceived EMR Benefits, Limitations and Recommendations for Improved Use

Participants were asked three open-ended questions:

- What have you found to be the three greatest benefits or advantages to using the EMR?
- What have you found to be the three greatest limitations or disadvantages to using the EMR?
- What recommendations do you have for improving the use and effectiveness of the EMR within their clinic/area at MUSC?

Benefits/Advantages to the EMR System

Nearly 82% of physicians and 56% of nurses identified availability and accessibility of the patient's record as a major benefit or advantage to using the EMR. See Table 3. Included in this category were comments related to having access to other clinic notes and records and having remote access to the EMR. One physician commented, "*Everyone has access to the entirety of the record across all disciplines.*" Another noted, "*The EMR greatly facilitates communication among primary care providers and specialists by having all patient information in one place.*" Other frequently cited benefits included (a) quality of record/documentation (legibility, data consistency and accuracy, organization of record, coding/compliance); (b) comprehensiveness and completeness of patient

Table 2. Physician and nurse mean satisfaction scores by domain/item

Domain	Item (0-9 scale)	Physician	Nurse
Overall satisfaction	Terrible……………………………………Wonderful	5.07	5.92*
	Difficult……………………………………Easy	5.17	6.21*
	Frustrating……………………………………Satisfying	4.33	5.09*
	Inadequate Power……………………………………Adequate Power	4.45	5.02
	Dull……………………………………Stimulating	4.65	5.39*
	Rigid……………………………………Flexible	4.42	5.41*
	Overall satisfaction composite score	**4.67**	**5.50***
Screen Design and Layout	Characters on the screen: Hard to read……………………Easy to read	6.67	7.16
	Highlighting on the screen simplifies task: Not at all……………Very much	5.49	6.54*
	Organization of information on screen: Confusing……………Very clear	5.45	6.17*
	Sequence of screens: Confusing……………………Very clear	5.56	5.81
Terms and System Information	Use of terms throughout system: Inconsistent……………………Consistent	6.09	6.58
	Computer terminology is related to the task you are doing: Never…………Always	5.89	6.31
	Position of messages on screen: Inconsistent……………………Consistent	6.05	6.48
	Messages on screen which prompt user for input: Confusing…………Very clear	5.42	6.14*
	Computer keeps you informed about what it is doing: Never…………Always	4.90	5.42
	Error messages: Unhelpful……………………Helpful	3.22	4.19
Learning	Learning to operate the system: Difficult……………………Easy	5.23	5.98*
	Exploring new features by trial and error: Difficult……………………Easy	4.73	5.76*
	Remembering names and use of commands: Difficult……………………Easy	4.93	5.51
	Tasks can be performed in a straight forward manner: Never…………Always	5.06	5.74*
	Help messages on the screen: Confusing……………………Clear	4.14	5.12*
	Supplemental reference materials: Confusing……………………Clear	4.10	5.45*
System Capabilities	System speed (to open or launch program): Too Slow…………Fast Enough	2.65	2.85
	System speed (to navigate within EMR, e.g. open a note): Too slow…………Fast enough	4.41	4.13
	System reliability: Unreliable……………………Reliable	4.72	5.01
	System tends to be: Noisy……………………Quiet	7.23	7.87*
	Correcting your mistakes: Difficult……………………Easy	5.12	4.96
	Experienced and inexperienced users' needs are taken into consideration: Never……………………Always	4.68	5.31

**significant at the p>.05 level*

Table 3.Most frequently cited benefits/advantages to using MUSC's EMR system

EMR Benefit	**Physicians** n=114	**Nurses** n=130
Availability and access to patient record	81.6%	56.2%
Quality of record/documentation	36.0%	14.6%
Comprehensiveness and completeness of record	33.3%	25.4%
Positive impact on efficiency	13.2%	10.0%
Continuity of care and standards of care	13.2%	10.8%
Ease of use	12.3%	15.4%
Avoidance of perils of paper	9.6%	20.0%

record (e.g., medications, results from diagnostic tests); (c) impact on efficiency (e.g., speed, timeliness in completing notes);(d) continuity of care and standards of care (e.g., templates, continuous outpatient record); and (e) avoidance of perils inherent with paper records (e.g., no lost charts, no moving charts).

Limitations/Disadvantages to the EMR System

Limitations or disadvantages identified by at least ten participants are listed in Table 4. The two most-cited limitations or disadvantages of the EMR were speed and cumbersome user

Table 4. Most frequently cited limitations/disadvantages to using MUSC's EMR system

EMR Limitation	**Physicians** n=114	**Nurses** n=130
Cumbersome user interface	38.6%	24.6%
Speed/too slow	37.3%	32.3%
Having to access multiple clinical systems or poor integration with other systems	17.5%	4.6%
Problems with templates	11.4%	4.6%
Insufficient training and support staff	11.4%	3.8%
Downtime (including system crashes or system locking/ freezing up)	14.0%	21.5%
Time-consuming to use	13.2%	14.6%

interface. Both physicians and nurses described the system as being too slow. Approximately 20 participants mentioned that the start time to open the application was particularly problematic. System downtime, including the moments when the "system crashes", "locks up or freezes" and requires user to "reboot" were described as frustrating by 14% of physicians and 22% of nurses. One physician commented that system crashes are frustrating: "*The program froze three times today while I was editing the same note, losing the edits each time. I spent 30 minutes trying to edit a single note without success*." A nurse noted, "*When the system is down, you are put completely on hold*." Several participants also expressed frustration with the amount of typing required and the time-consuming nature of entering data into the system. Understanding error messages and correcting mistakes also emerged as points of frustration. Comments reflecting concerns including statements such as: "*too much trial and error*", "*if you don't already know how to do something, you'll never figure it out", and "some error messages don't tell you how to solve the problem*."

Suggestions for Improving EMR Use at MUSC

Participants gave numerous written suggestions for improving EMR use at MUSC—73% of the physicians and 49% of nurses wrote comments. Suggestions ran the gamut from *"provide more training", "fix bugs in system", "address speed and reliability of system*", *to "make it more user friendly*."

Several common themes emerged from the written suggestions. Some of these relate to how the EMR is structured and used at MUSC and others relate more directly to the Practice Partner application itself. General suggestions included (1) address the speed/performance issues, (2) expand training and support personnel/resources, (3) improve the user interface/templates and simplify data entry, and (4) integrate more fully with other MUSC applications (e.g., scheduling, lab, radiology).

DISCUSSION

Nearly 250 physician and nurse EMR users at MUSC participated in this study. Nurses are more satisfied overall with the EMR system than are attending physicians. The greater system acceptance by nurses may be due to the approach MUSC took in implementing the EMR enterprise-wide. Initial efforts were focused on nurses and administrative staff, because these individuals were expected to lead in the effort. We felt if nurses were comfortable with the system, they could assist physicians. The downside of this approach is it has been time-consuming and resource-intensive to roll out the EMR throughout all the ambulatory care areas. Consequently, some physician users have not been fully trained on all of the system capabilities—yet, they have been using the application for two or more years.

Additionally, 58% of participating physicians report that they rarely, if ever dictate. Earlier studies have shown that data entry can negatively impact physicians' perceptions of their time, particularly if physicians are not proficient typists or comfortable entering information in the examination room with the patient (Gadd & Penrod, 2001; Penrod & Gadd, 2001; O'Connell, Cho, Shah, Brown, & Shiffman, 2004a; Hier, Rothschild, LeMaistre, & Keeler, 2005; Scott, Rundall, Vogt, & Hsu, 2005; Linder et al., 2006).

Researchers at Partners HealthCare System conducted a time-motion study and found no differences in primary care physician time utilization before and after EMR implementation, yet the majority of their physicians still perceived that the EMR required more time than paper records to document patient information (Pizziferri et al., 2005). Nurses in an earlier study reported that the EMR enabled them to finish their work much faster than before implementation (Likourezos et al., 2004); although we did not ask this specific question, it may help explain why MUSC nurses generally viewed the system more positively than physicians.

Nurses also tended to value the "avoidance of the perils of paper" more often than the physicians did, yet they did not mention availability as a major benefit as often as physicians did. We categorize these measures of availability/accessibility and avoidance of periods with paper separately, yet they are related. Nurses are responsible for ensuring that the patient's record is available at the time of the visit, thus, they experience the "grief" when the record is not available. Physicians are not the ones searching for the paper record, so they may not describe the benefit in quite the same manner. On the other hand, physicians were more likely to mention the benefits associated with having access to clinic notes from other physicians and more likely to comment on the overall quality of the documentation. Having a more "complete" picture of the patient's care (e.g., other visit notes, ancillary test results) stood out as particularly important to physicians. The EMR benefits identified by MUSC physicians and nurses are consistent with earlier studies (Sittig, Kuperman, & Fiskio, 1999; Gadd & Penrod, 2001; Penrod & Gadd, 2001; Likourezos et al., 2004; Joos, Chen, Jirjis, & Johnson, 2006).

Physicians and nurses had similar concerns when asked about limitations of the system or opportunities for improving its use at MUSC. In fact, speed and performance-related concerns seemed to cast a cloud over the benefits. Physicians and nurses identified speed and the slowness of the system as a *major* concern—launching the system was particularly problematic. Yet, in our multivariate regression analysis, system speed did not show up as a significant predictor of overall satisfaction. Although this finding surprised us, an earlier study at Brigham and Women's also found that system "response time" was not correlated with overall user satisfaction (Sittig, Kuperman, & Fiskio, 1999). Factors associated with physicians' ability to effectively use the system were more likely to be predictors of EMR satisfaction than speed alone.

In addition to their concerns with speed and performance, nearly 40% of the physicians felt the user interface was cumbersome, took too many clicks, and made it difficult to make corrections or changes. A common complaint was the "*system requires too many steps or clicks to perform a simple task*." Other institutions using different EMR applications have found their clinicians share these same concerns (Sittig, Kuperman, & Fiskio, 1999; Wager, Lee, White, Ward, & Ornstein, 2000; Gadd & Penrod, 2001; Penrod & Gadd, 2001; Miller & Sim, 2004; Scott, Rundall, Vogt, & Hsu, 2005). Some of the difficulty or frustration at MUSC may stem from a lack of sufficient training. For example, we observed from the open-ended comments that participants are frustrated with aspects of the system they may not fully understand how to use (e.g., writing prescriptions, modifying templates). Other frustrations may stem from using an EMR whereby primary care providers and specialists share and record their notes in a *single* patient record. Historically, each set of providers had specialty-specific records that only those in the specific group could view and modify.

Interestingly, the results of this survey are remarkably similar to those reported by researchers at Brigham and Women's in their ambulatory care division, despite this group's use of an in-house-developed EMR system. Using the same QUIS instrument, researchers there discovered that attending physicians scored their EMR lowest in the areas of system speed (although they did not distinguish between launch speed and speed of navigating within the application), helpfulness of error messages, and flexibility of system (Sittig, Kuperman, & Fiskio, 1999). The pattern of responses are quite similar, although MUSC physician reactions were less positive in terms of screen design and layout ($p<.01$), terms and system information ($p<.01$) and learning ($p<.001$). Brigham and Women's may have a more advanced training environment, and because they developed the EMR in-house, they have more flexibility in customizing the screens and layout to accommodate their physicians' preferences. No differences were found among overall user reaction and system capabilities. The physicians participating in the study at Brigham and Women had two or more years EMR experience.

Our study evaluated responses from both relatively new and experienced users. One might assume that due to the learning curve in moving from a paper-based system to an EMR, more experienced EMR users would be more satisfied with the system than less experienced users as Gamm et al. suggest (Gamm, Barsukiewicz, Dansky, & Vasey). We did not find that to be the case in our study. EMR users with less than one year of experience were equally satisfied overall with the system as were those who had used the application for two or more years.

Participants in this study provided a host of important suggestions for improving the EMR's use at MUSC. Other institutions would do well to address these issues and concerns up front to avoid problems later.

Our study has two primary limitations. First, the study is limited to a single healthcare organization and one EMR product. Thus, the results may not be generalizable to other healthcare organizations or those who use a different EMR system. Second, response bias is a concern. We compared the demographics of our physician and nurse participant populations with the general physician and nurse EMR population at MUSC and found no differences, yet response rate remains a limitation.

The real value of this study to other healthcare organizations may not be the results per se found at MUSC, but rather the methods and process used. Using the QUIS instrument, we were able to identify the aspects of our EMR system that stood out as problematic to clinician users—launch speed, navigation speed, and the clarity and helpfulness of error messages. Equally, if not more importantly, we provided all clinicians with the opportunity to offer suggestions on how to improve the system, and since this survey, we have incorporated many of their suggestions into our environment. Our leadership team has taken a number of steps to address the most widespread concerns identified by this survey and validated through other avenues. We believe other institutions will find the methods used in this study helpful in assessing clinician satisfaction and in soliciting suggestions for improvement.

CONCLUSION

Results of this survey suggest that MUSC physicians and nurses recognize and value having access to a single electronic patient record that is shared across the ambulatory care enterprise. However, they view the current system as less than ideal. Speed, performance and the user interface (e.g., need to simplify data entry) are of concern. Likewise, additional training and resources are needed to more effectively support the system and its users. Our leadership team has taken a number of steps to address the most widespread concerns identified by this survey and validated elsewhere. Steps to enhance system speed/performance and reliability are being taken, and multi-modal programs of improved training and user support are being implemented. We expect to observe the results of these initiatives and report outcomes in due course.

Assessing user satisfaction with the EMR is important in providing leadership with additional insight into the issues and concerns. Conducting formal evaluation studies, however, are often not done because of lack of available expertise and resources. We believe that surveys such as this one can prove to be a useful resource not only to our healthcare organization's leadership team, but to those in other healthcare institutions interested in easily assessing EMR satisfaction across the enterprise.

REFERENCES

Anderson, J., & Aydin, C. (Eds.). (2005). *Evaluating the organizational impact of healthcare information systems*, 2nd ed. Springer.

Bates, D. (2005). Physicians and ambulatory electronic health records. *Health Affairs, 24*(5), 1180-1189.

Brender, J. (2006). *Handbook of evaluation methods for health informatics.* Elsevier.

Burkle, T., Ammenwerth, E., Prokosch, H., & Dudeck, J. (2001). Evaluation of clinical information systems. What can be evaluated and what cannot?. *Journal of Evaluation in Clinical Practice, 7*(4), 373-385.

Chin, J., Diehl, V., & Norman, K. (1988). *Development of an instrument measuring user satisfaction of the human-computer interface.* Paper presented at the CHI, New Ork.

DeLone, W., & McLean, E. (2003). The DeLone and McLean model of information systems success: A ten-year update. *Journal of Management Information Systems, 19*(4), 9-30.

Despont-Gros, C., Mueller, H., & Lovis, C. (2005). Evaluating user interactions with clinical information systems: A model based on human-computer interactions models. *Journal of Biomedical Informatics, 38*, 244-255.

Friedman, C., & Wyatt, J. (1997). *Evaluation methods in medical informatics.* New York: Springer.

Gadd, C., & Penrod, L. (2001). Assessing physician attitudes regarding use of an outpatient EMR: A longitudinal, multi-practice study. *Proceedings/ AMIA Annual Symposium*, 194-198.

Gamm, L., Barsukiewicz, C., Dansky, K., & Vasey, J. *Pre- and post-control model research on end-users' satisfaction with an electronic medical record: Preliminary results.* Paper presented at the AMIA.

Hier, D., Rothschild, A., LeMaistre, A., & Keeler, J. (2005). Differing faculty and housestaff acceptance of an electronic health record one-year after implementation. *International Journal of Medical Informatics, 74*, 657-662.

Hillestad, R., Bigelow, J., Bower, A., Girosi, F., Meili, R., Scoville, R., et al. (2005). Can electronic medical record systems transform healthcare? Potential health benefits, savings and costs. *Health Affairs, 24*(5), 1103-1117.

Joos, D., Chen, Q., Jirjis, J., & Johnson, K. (2006). *An electronic medical record in primary care: Impact on satisfaction, work efficiency and clinic processes.* Paper presented at the AMIA, Washington, DC.

Likourezos, A., Chalfin, D., Murphy, D., Sommer, B., Darcy, K., & Davidson, S. (2004). Physician and nurse satisfaction with an electronic medical record system. *Journal of Emergency Medicine, 27*(4), 419-424.

Linder, J., Schnipper, J., Tsurikova, R., Melnikas, A., Volk, L., & Middleton, B. (2006). *Barriers to electronic health record use during patient visits.* Paper presented at the AMIA, Washington, DC.

Miller, R., & Sim, I. (2004). Physicians' use of electronic medical records: Barriers and solutions. *Health Affairs, 23*(2), 116-126.

O'Connell, R., Cho, C., Shah, N., Brown, K., & Shiffman, R. (2004a). Take note: Differential EHR satisfaction with two implementations under one roof. *J Am Med Inform Assoc., 11*(1), 43-49.

O'Connell, R., Cho, C., Shah, N., Brown, K., & Shiffman, R. (2004b). Take note: Differential EHR satisfaction with two implementations under one roof. *J Am Med Inform Assoc.*, *11*(1), 43-49.

Penrod, L., & Gadd, C. (2001). Attitudes of academic-based and community-based physicians regarding EMR use during outpatient encounters. *Proceedings/AMIA Annual Symposium*, 528-532.

Pizziferri, L., Kittler, A., Volk, L., Honour, M., Gupta, S., Wang, S., et al. (2005). Primary care physician time utilization before and after implementation of an electronic medical record: A time-motion study. *Journal of Biomedical Informatics, 38*, 176-188.

Scott, J., Rundall, T., Vogt, T., & Hsu, J. (2005). Kaiser Permanente's experience of implementing an electronic medical record: A qualitative study. *BMJ, 331*, 1313-1316.

Sittig, D., Kuperman, G., & Fiskio, J. (1999). *Evaluating physician satisfaction regarding user interactions with an electronic medical record system.* Paper presented at the AMIA.

Slaughter, L.A., Harper, B.D., and Norman, K.L. (1994). Assessing the equivalence of the paper and online formats of the QUIS 5.5. In *Proceedings of Mid Atlantic Human Factors Conference*, (pp. 87-91)Washington, DC.

Wager, K., Lee, F., & Glaser, J. (2005). *Managing healthcare information systems: A practical approach for healthcare executives.* San Francisco, CA: Jossey-Bass.

Wager, K., Lee, F., White, A., Ward, D., & Ornstein, S. (2000). Impact of an electronic medical record system on community-based primary care practices. *The Journal of the American Board of Family Practice, 13*(5), 338-348.

Windows, S. (2005). Version 14.0.2005. Chicago: SPSS Inc.

This work was previously published in the International Journal of Healthcare Information Systems and Informatics, edited by J. Tan, Volume 3, Issue 1, pp. 63-74, copyright 2008 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 4.16
Computerization of Primary Care in the United States

James G. Anderson
Purdue University, USA

E. Andrew Balas
Old Dominion University, USA

ABSTRACT

The objective of this study was to assess the current level of information technology use by primary care physicians in the U.S. Primary care physicians listed by the American Medical Association were contacted by e-mail and asked to complete a Web-based questionnaire. A total of 2,145 physicians responded. Overall, between 20% and 25% of primary care physicians reported using electronic medical records, e-prescribing, point-of-care decision support tools, and electronic communication with patients. This indicates a slow rate of adoption since 2000. Differences in adoption rates suggest that future surveys need to differentiate primary care and office-based physicians by specialty. An important finding is that one-third of the physicians surveyed expressed no interest in the four IT applications. Overcoming this barrier may require efforts by medical specialty societies to educate their members in the benefits of IT in practice. The majority of physicians perceived benefits of IT, but they cited costs, vendor inability to deliver acceptable products, and concerns about privacy and confidentiality as major barriers to implementation of IT applications. Overcoming the cost barrier may require that payers and the federal government share the costs of implementing these IT applications.

INTRODUCTION

The adoption of information technology (IT) to support the delivery of healthcare is recognized increasingly in many countries as an essential tool to improve patient care (Dick & Steen, 1997; Leaning, 1993; President's Information Technology Advisory Committee, 2004). Until recently, IT products available for healthcare providers mostly were designed for large organizations, were business-oriented, complex to implement, and costly. Recent advances in technology have made IT applications more available to primary care physicians in smaller practices. Products are available that are modular; able to be integrated with different systems, and designed to fit the physician's practice pattern without substantial investments in hardware, software, and maintenance (McDonald & Metzger, 2002).

As a result, the introduction of computers and IT applications into primary care in countries with favorable government policies and financial incentives has been rapid (Kidd, 2000; Mount, Kelman, Smith, & Douglas, 2000; Purves, Sugden, Booth, & Sowerby, 1999; Thakurdas, Coster, Guirr, & Arroll, 1996). A number of English-speaking countries has experienced widespread implementation of information technology. The Harvard School of Public Health and the Commonwealth Fund's International Symposium survey of primary care physicians found the following proportions of primary care physicians in the following countries who were using electronic medical records: U.S. (17%); Canada (14%); Australia (25%); New Zealand (52%); and the U.K. (59%). The survey also found the following use of electronic prescribing by primary care physicians: U.S. (9%); Canada (8%); Australia (44%); New Zealand (52%); and the U.K. (87%) (Harris Interactive, 2001a).

The U.S. trails European countries in the use of information technology in patient care. Overall, 29% of general practitioners in the European Union use electronic medical records compared to only 11% in the U.S. Only three countries from the Organization for Economic Cooperation and Development (OECD)—Portugal, France, and Spain—lag behind the U.S. (Harris Interactive, 2002b). Despite its potential to improve efficiency and quality of care, use of information technology in healthcare lags behind other sectors of the economy in the U.S. In 2001, most of the $20 million invested in healthcare information technology was used to computerize financial systems (Goldsmith, Blumenthal, & Rishel, 2003). Less than 10% of U.S. hospitals had adopted electronic medical record systems and less than 5% had implemented computerized physician order entry by 2001.

Given the increasing public attention to the importance of health information technology, the rate of IT adoption among primary care providers is important (Hillestad, et al., 2005). Accurate estimates of the adoption rate for information technology form the basis for policy regarding how to stimulate its use by physicians. The overall aim of this study was to determine primary care physicians' use of information technology in patient care. The specific objectives included the following:

1. Estimating the proportion of primary care physicians who have adopted information technology applications in their practices.
2. Determining physician perceptions of the benefits of these IT applications.
3. Determining physician perceptions of the barriers to the adoption of IT applications in their practices.

Primary care in the U.S. is delivered by physicians who comprise several specialties; namely, family practice (FP), internal medicine (IM), pediatrics (PEDS), and obstetricians and gynecologists (OBGYN). One other group of physicians was included in the survey comprising medical specialties such as geriatrics and occupational medicine.

Four IT applications were selected for investigation. First, electronic medical records (EMRs) are promoted as more comprehensive and accessible to healthcare providers. Studies have shown that EMRs have the potential to reduce medical errors, especially when integrated with other applications such as decision support (Bates et al., 1998). Electronic prescribing involves the use of computers or hand-held devices to submit prescriptions to pharmacies electronically. E-prescribing has the potential to improve efficiency, to reduce prescription errors, and to improve compliance with managed-care formularies (Miller, Gardner, Johnson & Hripcsak, 2005; Schiff & Rucker, 1998). Third, point-of-care decision support tools can improve the quality of patient care; for example, an antibiotic decision support system (Evans et al., 1998) and automated decision support alerts for contraindicated medications (Galanter, Didomenico & Polikaitis, 2005). Fourth, patients consistently have expressed a strong desire for online communication with physicians (Harris Interactive, 2005). This may involve e-mail queries as well as online consultations. Electronic communication allows physicians to deliver better care and patients to assume greater responsibility for their own care.

METHODS

Survey Method

A Web-based survey was developed to investigate primary care physicians' use of the four IT applications described previously. These applications were selected because healthcare providers in the U.S. and the EU find them helpful and effective (Harris Interactive, 2003, 2005). Comparative data also exist from earlier surveys on the use, perceived benefits, and barriers to these applications. At the same time, earlier studies failed to differentiate primary care physicians by specialty.

We describe the design and administration of the survey. A Web-based survey method was chosen, because it permitted us to survey a national sample of primary care physicians with a reasonable budget (Eysenbach, 2005; Lazar & Preece, 1999; Wyatt 2000). Also, we wanted to sample an Internet-literate population that is most likely to be early adopters in their practices (Rogers, 1983).

Survey Design

The study was sponsored by the Quality Improvement Working Group of the American Medical Informatics Association and the School of Public Health at St. Louis University. The e-mail that was sent out inviting primary care physicians to participate in the study contained a link to the Web-based survey (see Appendix).

In order to facilitate comparisons to earlier surveys, items were adapted from other widely cited surveys; in particular, the annual Health Care Information and Management Systems Society (HIMSS) Leadership Survey (HIMSS, 2002) and the Harris Interactive polls that were conducted in the U.S. and the EU (Harris Interactive, 2002b, 2003).

The questionnaire was divided into seven sections. The first section included information about the physician's specialty and practice. The second section asked physicians to rate the priority of a number of Internet technologies. The next three sections listed specific financially focused, clinically focused, and patient-focused IT applications. The physician was asked to indicate for each IT application if he or she (1) had implemented, (2) planned to implement within one year, (3) had no plans to implement but was interested in learning more, or (4) had no interest. Physicians also could respond by indicating that they didn't know or that they chose not to answer that question. The sixth section asked physicians to rate the benefits of using IT applications on a Likert scale. Responses ranged from (1) high benefit to (4) not a benefit. The final section asked about barriers to implementing IT applications. Responses ranged from (1) not a barrier to (5) insurmountable barrier. A copy of the survey is included in the Appendix.

Factor analyses were performed on the items that measured perceived benefits of the IT application and on the perceived barriers to implementation. A single factor accounted for 63% of the variance in the benefits items. The reliability based on Chronbach's Alpha was 0.93. For the barriers items, a single factor accounted for 48% of the variance. The reliability was 0.86 based on Chronbach's Alpha.

Sample

We contracted with SK&A Information Services to broadcast an e-mail invitation to primary care physicians to participate in the study. This company maintains a comprehensive list of physicians based on the AMA Physician Masterfile. The list is updated weekly through the use of surveys, publication mailings, and the U.S. Postal Services Address Correction Services. E-mail invitations to participate in the study were sent out to 31,743 primary care physicians. Of these e-mails, 1,101 were rejected due to invalid e-mail addresses. A total of 2,145 physicians responded, representing a 7.3% response rate to the survey. The software prevented respondents from completing the survey more than one time. Questionnaires from physicians who were not currently practicing or who were not currently engaged in primary care were eliminated as were questionnaires with significant missing data. This resulted in a final sample of 1,665 that was used in the analysis.

Table 1 presents demographic data and practice information about the study sample. Sixty percent of the physicians were between 41 and 60 years of age, while 29% were younger. Three-fourths of the responding physicians were male. Almost 75% practiced family medicine, internal medicine, or pediatrics, while 15% practiced obstetrics and gynecology and 9% other medical specialties. More than 88% of the respondents were primar-

Table 1. Physician characteristics of the study sample

Characteristic	N	%
Age		
30 or less	16	1.1%
31–40	259	17.9%
41–50	537	27.0%
51–60	484	33.4%
61–70	108	7.4%
70 or above	46	3.2%
Gender		
Male	1134	74.5%
Female	388	25.5%
Specialties		
Family Practice	448	29.8%
Internal Medicine	368	24.5%
Pediatrics	324	21.5%
Obstetrics and Gynecology	225	15.0%
Other Medical Specialties	138	9.2%
Role		
Physician	1972	88.4%
Administrative	176	11.6%
Type of Organization		
Hospital	232	13.9%
Group: 10 or more	298	17.9%
Group: Less than 10	607	36.5%
Solo	327	19.6%
Other Settings	201	12.1%

ily clinicians. The other 12% held primarily administrative positions in their practices and were excluded from the final analysis.

About 14% of the respondents were hospital-based. Almost 18% of the physicians were in group practices of 10 or more; more than one-third of the respondents were in small group practices with less than 10 physicians, and 20% of the physicians were in solo practice. The remaining 12% were in integrated health delivery service organizations, managed care organizations, and so forth.

RESULTS

Use of Information Technology

Table 2 shows the extent to which physicians in each specialty have implemented each of the four IT applications. Overall, only one out of four has implemented electronic medical records and report using point-of-care decision support tools. About 23% communicate electronically with patients. Only one out of five primary care

Table 2. Use of information technology by primary care specialty (%)

Application	FP	IM	PEDS	OB/GYN	Other	Total
Electronic Medical Records						
Implemented	23.2	31.2	23.0	16.4	40.6	25.8
Plan to implement	16.9	13.9	12.5	16.0	12.8	14.4
Interested in	26.7	23.7	33.4	23.7	19.5	26.4
No interest	24.8	21.2	19.7	31.5	21.8	23.5
NA	8.4	10.0	11.5	12.3	5.3	9.5
Electronic Prescribing						
Implemented	17.7	26.4	20.4	13.3	24.0	20.1
Plan to implement	18.2	16.7	13.0	14.3	15.2	16.2
Interested in	21.5	15.5	21.8	17.6	12.0	18.6
No interest	31.1	30.2	30.6	35.2	34.4	31.3
NA	11.6	11.2	14.1	19.5	14.4	13.8
Decision Support Tools						
Implemented	27.6	25.7	24.0	15.6	30.8	25.1
Plan to implement	16.6	11.1	9.4	10.1	8.7	12.0
Interested in	11.2	11.5	9.4	15.6	11.5	12.2
No interest	33.9	35.9	35.6	43.6	35.6	35.9
NA	10.7	15.8	21.5	15.1	13.5	14.8
Electronic Communication						
Implemented	25.5	26.6	20.4	21.2	26.2	23.2
Plan to implement	11.4	7.1	8.2	10.1	1.6	8.7
Interested in	9.3	6.5	12.1	11.1	9.5	9.9
No interest	29.0	28.1	28.9	31.7	29.4	28.9
NA	24.8	31.7	30.4	26.0	33.3	29.4

NA=don't know or I choose not to answer physicians utilizes electronic prescribing. A surprisingly high number of physicians indicated no interest in all of the IT applications. Thirty-six percent indicated no interest in decision support tools, while 31.3% and 23.5% evidenced no interest in electronic prescribing and electronic medical records, respectively. Almost 30% stated that they were not interested in electronic communication with patients.

A greater proportion of internists report having implemented all four of the IT applications in practice ($p<0.05$). Thirty-one percent have implemented electronic medical records; about 26% have implemented electronic prescribing, decision support tools, and e-mail communication with patients. In general, OBGYNs are the least likely to have implemented any of the IT tools with the exception of electronic communication with patients ($p<0.05$). Less than one out of six of these physicians have implemented electronic medical records or electronic prescribing or decision support tools, and only one out of five have implemented electronic communication with patients. OBGYNs also expressed the least interest in IT applications ($p<0.05$). More than 30% indicated no interest in electronic medical records, electronic prescribing, and e-mail communication with patients. More than 40% indicated no interest in implementing decision support tools. There may be several major reasons for this low use of IT and lack of interest by OBGYNs. Most of the IT applications are general and may not meet the specific needs of this specialty. Also, there appear to be few published studies involving the use of IT by OBGYNs.

Perceived Benefits and Barriers

Overall, the majority of primary care physicians surveyed perceived benefits from implementing IT applications (see Table 3). Almost 75% indicated that these applications could reduce errors; 70% perceived IT as potentially increasing their productivity; more than 60% indicated that IT tools have the potential to reduce costs and to help patients assume more responsibility. Physi-

Table 3. Perceived benefits of implementing IT applications (%)

Benefit	High	Medium	Low	None
Patients assume responsibility for monitoring symptoms/ disease	23.6	38.7	22.1	15.6
Shorter consultations	17.0	29.1	20.9	32.9
Patients not seeking medical care when it was not needed	22.5	28.2	24.4	25.0
Patients coming in sooner for necessary treatment	33.8	29.6	18.4	18.3
Fewer unnecessary tests	29.4	27.9	16.1	26.5
Fewer unnecessary treatments	32.8	24.9	16.9	25.4
Fewer errors	53.4	21.4	10.5	14.7
Increased productivity	39.2	30.3	14.2	16.3
Reduced costs	37.5	25.5	15.4	21.6

cians are less certain about some of the other potential benefits of IT applications. About half of the physicians surveyed evidenced skepticism that IT applications would shorten consultations and reduce the number of patients who seek unnecessary healthcare. More than 40% felt that IT is unlikely to reduce unnecessary tests and treatments.

More than 80% of primary care physicians report the lack of financial support for IT applications as a major barrier to adoption. This is followed by their perceptions that vendors fail to deliver acceptable products as primary barriers to implementing these tools (79.3%) (see Table 4). In general, physicians perceive these barriers as difficult to overcome. Almost two-thirds of the physicians surveyed also cited the lack of a strategic plan for implementing applications and difficulty in recruiting experienced IT personnel as major barriers, while more than one-half cited lack of sufficient knowledge of IT as a barrier to implementation. At the same time, physicians indicated that these last three barriers easily could be overcome.

Table 4. Perceived barriers to implementing IT applications (%)

Barriers	**No Barrier**	**Easily Overcome**	**Overcome some effort**	**Overcome great effort**	**Insurmountable**
Lack of financial support	7.6	5.0	35.3	41.3	10.7
Vendors' inability to deliver acceptable products	12.4	8.3	34.8	36.3	8.2
Acceptance by staff	17.8	23.9	41.6	15.3	1.3
Difficulty proving quantifiable benefits	14.8	18.0	38.7	24.6	3.9
Lack of strategic plan for implementing	19.7	15.2	35.7	25.3	4.1
Recruiting experienced IT personnel	22.0	17.6	31.7	24.0	4.8
Retaining experienced personnel	24.6	17.9	36.6	18.1	2.8
Insufficient knowledge of IT applications	15.0	22.5	41.4	19.3	1.7
Considerable investment in IT applications	6.1	6.9	28.8	47.6	10.6

Predictors of IT Implementation

Table 5 provides the logistic regression models and predictors for implementing each of the IT applications. Demographic factors, specifically age and gender, were not associated significantly with the implementation of the four IT applications. In only one instance was there a significant difference between male and female physicians. Males were almost twice as likely to implement e-prescribing as females.

Physicians' specialties did predict whether or not they had implemented certain IT applications. Pediatricians and obstetricians and gynecologists were significantly less likely to have implemented electronic medical records. In contrast, family practitioners were almost three times more likely to have implemented point-of-care decision support tools. Specialty was not a significant predictor of implementation of electronic prescribing and communication with patients.

Table 5. Predictors of the implementation of IT applications (odds ratios)

Characteristic	EMR	E-Prescribing	Decision Support	E-Communication
Age				
30 or less	1.000	1.000	1.000	1.000
31–40	0.668	1.474	0.761	1.360
41–50	0.421	0.401	0.760	1.614
51–60	0.568	0.392	0.660	1.157
61–70	0.530	0.503	0.606	1.393
70 or above	0.503	0.706	0.499	1.393
Gender				
Male	1.175	1.942**	1.094	1.066
Female	1.000	1.000	1.000	1.000
Specialties				
FP	1.420	1.433	0.591**	0.924
IM	0.712	1.125	0.957	0.851
Pediatrics	0.513**	0.622	1.206	0.616
OBGYN	0.406**	0.957	1.180	0.586
Other	1.000	1.000	1.000	1.000
Benefits				
Fewer Errors	1.541**	1.574**	1.238*	1.086
Increased Productivity	1.023	1.282*	1.157	0.919
Reduced Costs	0.804*	0.724**	0.788*	0.868
Barriers				
Lack of Financial Support	1.591**	1.452**	1.296*	0.960
Vendors' Failure to Deliver	1.169	1.211*	1.309**	1.108
Considerable Investment	1.207	1.271*	1.221	1.278

***p<0.01* **p<0.05*

Perceived benefits and barriers appear to be consistent predictors of whether or not primary care physicians implemented three of the four IT applications. Physicians who perceived that IT can reduce medical errors were one and one-half times more likely to have implemented electronic medical record, e-prescribing, and decision support tools. In contrast, physicians who cited lack of financial support and the considerable investment required to implement these applications as significant barriers were less likely to have implemented all three of these IT applications. Physicians who perceived vendors as failing to deliver useful and acceptable products were significantly less likely to have implemented decision support tools. The decision to implement electronic communication with patients did not appear to be affected by demographic characteristics, specialty, or perceptions of benefits or barriers.

DISCUSSION

Adoption of electronic medical records has been the most widely surveyed IT application. A review of 22 studies of outpatient electronic medical record (EMR) adoption from 1998 to 2002 suggested a utilization rate of 20% to 25% at the time of the surveys (Brailer & Terasawa, 2003). However, data from the U.S. National Ambulatory Medical Care Survey (NAMCS) indicated that in 2001, only 17% of office-based physicians used electronic medical records (Burt & Hing, 2005).

These studies vary considerably in terms of how respondents were selected and their generalizability to a physician population. Many of the studies are unscientific and utilized surveys of meeting attendees. Only three of the 22 studies reviewed were rated as generalizable. Also, most of these studies do not differentiate among physicians by specialty. Consequently, there is only limited data on adoption of EMRs by specialty. The 2002 Health Care Information and Management Systems Society (HIMSS, 2002) survey administered to attendees and exhibitors at the annual conference found that 42% of internal medicine practices and 30% of family medicine practices reported using EMRs. These rates show little change from the HIMSS survey in 2001. However, since only meeting attendees were surveyed, it is impossible to extrapolate these results to the U.S. primary care physician population as a whole.

There are fewer studies of the adoption of other IT applications such as electronic prescribing and online communication between physicians and patients. The National Ambulatory Medical Care Survey (NAMCS) indicated that only 8% of office-based physicians in 2001 ordered prescriptions electronically (Burt & Hing, 2005). The Harris Interactive study that compared use of IT by U.S. general practitioners to European physicians found that 17% of physicians in primary care practices reported that they used EMRs, and 9% reported using electronic prescribing (Harris Interactive, 2002a). This survey also dates back to 2000-2001. Neither study differentiates physicians by specialty.

More recent information is needed about the extent to which primary care physicians use information technology for patient care, patterns of use, and perceived barriers to use of IT. Many of the surveys discussed earlier were undertaken before the year 2000. The NAMCS statistics on uses of computerized clinical support systems in medical settings are based on office-based physician practices rather than only on primary care physicians (Burt & Hing, 2005). The Harris Interactive study reports aggregate statistics for primary care physicians and specialists. Our survey examined IT applications that appear to offer the greatest potential to primary care physicians in providing high-quality patient care. It also differentiates primary care physicians by specialty.

This study provides evidence from a large sample of U.S. primary care physicians that there is limited implementation of clinical and patient care IT applications. Overall, only about 25% of primary care physicians have implemented electronic medical records, e-prescribing, point-of-care decision support tools, or electronic communication with patients. These results are similar to those from a Harris Interactive survey of 400 U.S. physicians conducted in 2001 and other earlier studies indicating a slow rate of adoption However, the proportion of physicians who have implemented e-prescribing has almost doubled from 11% to 20% since 2001. This may be due in part to improvements in the technology, such as the use of wireless devices.

Of concern is the finding that almost one out of three primary care physicians surveyed expressed little or no interest in the four IT applications. This may indicate that while two-thirds of primary care physicians perceive that implementation of IT can reduce costs and errors and help patients assume more responsibility for their medical conditions, a significant number of these physicians does not perceive the advantages of implementing IT technologies to provide patient care. One way of overcoming this barrier may be for medical specialty societies to offer seminars, short courses, and/or Web seminars on IT for CME credit with a focus on those features that are most useful to physicians in that specialty.

Age and gender on the whole do not appear to predict implementation of these four IT applications. However, there are significant differences in implementation among the specialties. A greater proportion of internists report having implemented all four IT applications. Pediatricians and obstetricians and gynecologists are less likely to have implemented EMRs, while family practitioners are more likely to have implemented decision support tools. OBGYNs, in particular, have been slow to adopt IT in practice. Only 16% have implemented EMRs and decision support tools. Even less, 13%, have implemented electronic prescribing. The slow adoption of IT applications by this specialty group may be due to the fact that these tools fail to address the special needs of this group of physicians. Also, OBGYNs may need to see more studies that demonstrate how these tools can help them to improve their practices.

This finding suggests that future surveys that assess adoption of IT applications by physicians need to differentiate by specialty rather than to treat primary care physicians or office-based physicians as homogeneous groups. Efforts to encourage IT adoption by physicians need to be tailored to specific specialty groups by emphasizing features of the technology that are particularly useful to that specialty.

Perceptions of benefits and barriers are significant predictors of implementation of three of the four applications. Physicians who perceive that EMRs, e-prescribing, and decision support tools can help them to reduce medical errors are significantly more likely to have implemented these technologies. At the same time, perception of barriers is a significant impediment to implementation (Anderson, 1997, 1999; Harris Interactive 2001b). Those physicians who perceived lack of financial support and high investment cost required were much less likely to have implemented these three IT applications. Also, physicians cited lack of experience and knowledge of IT as barriers. This may indicate that physicians may feel that learning to use IT applications in practice may require too much time and energy by them and their staff in order to achieve the perceived benefits. Consequently, a key to increased use of patient care IT applications by primary care physicians may be to convince them that the benefits significantly outweigh the barriers, primarily cost. Also, physicians do not perceive vendors as delivering acceptable IT products that meet their needs. More than 70% of physicians who responded to the survey perceived vendors' unresponsiveness as a barrier to implementation of IT. It may be necessary for vendors to examine more thoroughly the needs of primary care physicians and how their IT applications fit into clinical practice in order to convince physicians to adopt them.

Other studies have indicated that lack of funding and costs are the largest barriers to the adoption of EMRs. Surveys have found that 50% or more of respondents cited lack of adequate funding as the major barrier to implementation (HIMSS, 2002; Medical Group Management Association, 2001; Medical Records Institute, 2002; Miller & Sims, 2004). This perception is based on the fact that implementation of some IT applications such as EMRs requires large up-front investment and ongoing maintenance costs. A study by the California Health Care Foundation (2003) estimated that the cost of implementing a computerized physician order entry (CPOE) system in an ambulatory care practice ranges from $15,000 to $50,000 per physician with a median cost of $30,000 per physician.

Overcoming the cost barrier will be difficult and may require incentives by payers and the government. An example is New Zealand, Australia, and the U.K., which have introduced government funding programs to stimulate adoption and use of EMRs (Bates, Ebell, Gotlier, Zapp & Mullins, 2003). Professional associations also can facilitate adoption of IT. The American Academy of Family Physicians, through a nonprofit foundation, is developing low-cost, open-source EMR software that will be available to physicians with no licensing fee.

Decisions to implement electronic communication with patients appear to be independent of perceptions of benefits and barriers. Barriers to electronic communication with patients may be different than barriers to the other IT applications. Physicians generally express concerns about the legal status of these communications and concern about the security of patient information sent over the Internet.

One of the limitations of this study is the low response rate (7.3%). Low response rates are one of the major limitations of Web-based surveys in general (Eysenbach, 2005). A systematic review of 17 Internet-based surveys of health professionals found that reported response rates ranged from 9% to 94% (Braithwaite, Emery, de Lusignan & Sutton, 2003). Most of these studies utilized professional e-directories. Some used commercial organizations' e-mail directories or recruited volunteers via Web sites of electronic discussion groups. Six of the 17 studies reviewed did not report response rates. A meta-analysis of response rates in Web- and Internet-based surveys found that the mean response rate for 68 surveys was 39.6% with a standard deviation of 15.7% (Cook, Heath & Thompson, 2000). Other researchers have reported similarly low response rates of 18% for a study of physicians in Hong Kong (Leung, Johnston, Ho, Wong & Cameo, 2001).

One study of general practitioners' use of decision support for management of familial cancer sent five separate e-mail reminders and achieved a response rate of 52.4% (Braithwaite, Sutton, Smithson & Emery, 2002). In the case of our study, the high cost of sending additional reminders to physicians precluded our doing so.

Since our survey was administered online and did not include an alternative mail survey, there is a risk of over-sampling respondents who are more likely to utilize computers in their practices. Our sample was drawn from physicians with e-mail addresses listed by the American Medical Association (AMA). These physicians may be knowledgeable about IT applications and more likely to implement them in patient care. This sample design was adopted since we wanted to sample an Internet- and computer-literate population of primary care physicians. These physicians are most likely to be early adopters of IT applications in their practices. Consequently, estimates of implementation reported in this study are likely to be higher than for the entire population of primary care physicians.

At the same time, limitations on the generalizability of the results apply to many of the earlier reported studies of IT adoption by physicians

(Brailer & Terasawa, 2003). The HIMSS surveys were voluntary surveys administered to conference attendees (HIMSS, 2002). The MediNetwork 2002 Medical Group Office Management Systems Survey was voluntary and reported a 7.52% response rate. The AHA Most Wired Survey 2002 and the Medical Records Institute Survey of Electronic Health Record Trends and Usage sponsored by SNOMED were online voluntary surveys and did not report response rates. Comparative data for the U.S. and the E.U. reported by Harris Interactive did not report response rates. Data on the E.U. countries were based on the EuroBarometer 104 conducted in June/July 2001. U.S data were collected by Harris Interactive. Our study is an improvement over a number of these earlier studies in which there are serious questions about the reliability and the generalizability of results due to flawed study design or industry sponsorship (e.g., the HIMSS Leadership Survey). Also, earlier studies with few exceptions failed to differentiate primary care physicians or office-based physicians by specialty.

In this study, no attempt was made to specify specific features of each of the four IT applications. Physicians simply were asked if they had implemented or intended to implement each application. However, features of each application vary considerably from practice to practice. For example, an EMR in addition to patient problem lists, medications, allergies, tests, and personal information and medical history may be linked to an electronic prescribing system and evidence-based decision support tools.

CONCLUSION

The present study has documented the extent to which primary care physicians use IT in providing patient care. Variation among different primary care specialty groups is an important finding as is the finding that one out of three primary care physicians expressed no interest in using any of the four IT applications for patient care. Moreover, the finding that perceived that benefits and barriers are the most significant predictors of IT implementation has implications for strategies to promote implementation of IT in clinical practice. Primary care physicians will need to be convinced that the benefits of these tools outweigh their costs. Also, vendors will need to be more responsive to the needs of primary care physicians. Finally, overcoming the costs barrier will require incentives and/or cost sharing by payers and the federal government.

ACKNOWLEDGMENT

This study was supported by the Center for Education and Research in Information Assurance and Security (CERIAS). Data were collected by the Purdue University Social Research Institute. Marilyn Anderson, Don Malott, and Heather Rodriquez assisted with the data analysis and manuscript preparation.

REFERENCES

Anderson, J.G. (1997). Clearing the way for physicians' use of clinical information systems. *Communications of the ACM, 40*, 83–90.

Anderson, J.G. (1999). Increasing the acceptance of clinical information systems. *MD Computing, 16*, 62–65.

Bates, D.W., Ebell, M., Gotlier, E. Zapp, J., & Mullins, H.C. (2003). A proposal for electronic medical records in U.S. primary care. *Journal of the American Medical Informatics Association, 10*, 1–10.

Bates, D.W., Leape, L.L., Cullen, D.J., Laird, N., Petersen, L.A., Teich, J.M., et al. (1998). Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *Journal of the American Medical Association, 280*, 1311–1316.

Brailer, D.J., & Terasawa, E.L. (2003). *Use and adoption of computer-based patient records.* Oakland, CA: California Health Care Foundation.

Braithwaite, D., Emery, J., de Lusignan, S., & Sutton, S. (2003). Using the Internet to conduct surveys of health professionals: A valid alternative? *Family Practice, 19*, 587–590.

Braithwaite, D., Sutton, S., Smithson, W.H., & Emery, J. (2002). Internet-based risk assessment and decision support for the management of familial cancer in primary care: Survey of GP's attitudes and intentions. *Family Practice, 20*(5), 545–551.

Burt, C.W., & Hing, E. (2005). *Use of computerized clinical support systems in medical settings: United States 2001-03. Advanced data from vital and health statistics, No. 353.* Hyattsville, MD: National Center for Health Statistics.

California Health Care Foundation. (2003, October). *Electronic medical records: Lessons from small physician practices. ihealthreports.*

Cook, C., Heath, F., & Thompson, R.L. (2000). A meta-analysis of response rates in Web- and Internet-based surveys. *Educational and Psychological Measurement, 60*(6), 821–836.

Dick, R.S., & Steen, E.B. (1997). *The computer-based patient record: An essential technology for health care.* Washington, DC: National Academy Press, Institute of Medicine.

Evans, R.S., Pestotnik, S., Classen, D., Clemmer, T., Weaver, L., Orme, J., et al. (1998). A computer-assisted management program for antibiotics and other anti-infective agents. *New England Journal of Medicine, 338*, 232–238.

Eysenbach, G. (2005). Using the Internet for surveys and research. In J.G. Anderson, & C.E. Aydin (Eds.), *Evaluating the organizational impact of healthcare information systems* (2nd ed.) (pp. 129–143). New York: Springer.

Galanter, W.L., Didomenico, R.J., & Polikaitis, A. (2005). A trial of automated decision support alerts for contraindicated medications using computerized physician order entry. *Journal of the American Medical Informatics Association, 12*(3), 269–274.

Goldsmith, J., Blumenthal, D., & Rishel, W. (2003). Federal health information policy: A case of arrested development. *Health Affairs, 22*(4), 44–55.

Harris Interactive. (2001a). U.S. trails other English speaking countries in use of electronic medical records and electronic prescribing. *HarrisInteractive Health Care News, 1*(28), 1–3.

Harris Interactive. (2001b). The increasing impact of eHealth on physician behavior. *HarrisInteractive Health Care News, 1*(31), 1–14.

Harris Interactive. (2002a). The future use of the Internet in 4 countries in relation to prescriptions, physician communication and health information. *HarrisInteractive Health Care News, 2*(13), 1–4.

Harris Interactive. (2002b). European physicians especially in Sweden, Netherlands and Denmark lead U.S. in use of electronic medical records. *HarrisInteractive Health Care News, 2*(16), 1–3.

Harris Interactive. (2003). eHealth influence continues to grow as usage of the Internet by physicians and patients increases. *HarrisInteractive Health Care News, 3*(6), 1–7.

Harris Interactive. (2005). Many nationwide believe in the potential benefits of electronic medical records and are interested in online communications with physicians. *The Wall Street Journal Online Health Care Poll, 4*(4), 1–5.

Hillestad, R., Bigelow, J., Bower, A., Girosi, F., Meili, R, Scoville, R., et al. (2005). Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. *Health Affairs, 24*(5), 1103–1117.

HIMSS 13th annual HIMSS leadership survey. (2002). Retrieved from http://www.himss.org/2002survey

Kidd, M.R. (2000). Clinical practice guidelines and the computer on your desk. *Medical Journal of Australia, 173*, 373–375.

Lazar, J., & Preece, J. (1999). Designing and implementing Web-based surveys. *Journal of Computer Information Systems, 39*(4), 63–67.

Leaning, M. (1993). The new information management and technology strategy of the NHS. *British Medical Journal, 307*(6898), 217.

Leung, G.M., Johnston, J.M., Ho, L.M., Wong, F.K., & Cameo, S.C. (2001). Computerization of clinical practice in Hong Kong. *International Journal of Medical Informatics, 62*, 143–154.

McDonald, K., & Metzger, J. (2002, September). Achieving tangible IT benefits in small physician practices. *California HealthCare Foundation, ihealth Report.*

Medical Group Management Association. (2001). *Medical group management association survey 2001.*

Medical Records Institute. (2002). *4th annual Medical Records Institute's survey of electronic health record trends and usage, 2002.*

Miller, R.A., Gardner, R.M., Johnson, K.B., & Hripcsak, G. (2005). Clinical decision support and electronic prescribing systems: A time for responsible thought and action. *Journal of the American Medical Association, 12*(4), 403–409.

Miller, R.H., & Sims, I. (2004). Physicians' use of electronic medical records: Barriers and solutions. *Health Affairs, 23*(2), 116–126.

Mount, C.D., Kelman, C.W., Smith, L.R., & Douglas, R.M. (2000). An integrated electronic health record and information system for Australia. *Medical Journal of Australia, 172*, 25–27.

President's Information Technology Advisory Committee. (2004). *Revolutionizing health care through information technology.* Arlington, VA: National Coordination Office for Information Technology Research and Development.

Purves, I.N., Sugden, B., Booth, N., & Sowerby, M. (1999). The PRODIGY project—The iterative development of the release one model. In *Proceedings of the AMIA Annual Symposium* (pp. 359–363).

Rogers, E.M. (1983). *Diffusion of innovation* (3rd ed.). New York: Free Press.

Schiff, G.D., & Rucker, D. (1998). Computerized prescribing: Building the electronic infrastructure for better medication usage. *Journal of the American Medical Association, 280*, 516–517.

Thakurdas, P., Coster, G, Guirr, E., & Arroll, B. (1996). New Zealand general practice computerisation: Attitudes and reported behavior. *New Zealand Medical Journal, 109*, 419–422.

Wyatt, J. (2000). When to use Web-based surveys. *Journal of the American Medical Informatics Association, 7*(4), 426–430.

APPENDIX

Dr. <name>

The Quality Improvement Working Group of the American Medical Informatics Association in conjunction with the School of Public Health at St. Louis University is undertaking a survey of physician experience with information technology at the point of care. The survey is being performed under contract with the Social Research Institute at Purdue University and funded by the Center for Education and Research in Information Assurance and Security.

To participate, simply click on the link below and you will be directed to the Social Research Institute Web site at Purdue University. Please complete the short survey. Your responses will be kept strictly confidential and will be used solely for academic research purposes. We are grateful for your willingness to provide your valuable perspective on the real implementation experience of a physician using information technology at the point of care.

If you have any questions, please contact:

James G. Anderson, Ph.D.
Professor of Medical Sociology
Purdue University
Chair Elect, Quality Improvement Working Group
American Medical Informatics Association
Andersonj@soc.purdue.edu

E. Andrew Balas, M.D.
Dean and Professor
School of Public Health
St. Louis University
Chair, Quality Improvement Working Group
American Medical Informatics Association
balasea@slu.edu

CLICK HERE <Web site Address>

continued on the following pages

Which of the following best describes your role within your Organization:

1. Physician
2. Director
3. Scientist
4. President
5. Chief of Executive officer
6. Medical Director
7. Chief Medical Officer
8. Vice President of Medical Services
9. Other

a. don't know
b. I choose not to answer

Which of the following best describes the environment where you spend most of your workday:

1. Hospital
2. Medium or large group practice or clinic (10 or more physicians)
3. Small group practice or clinic (less than 10 practicing physicians)
4. Solo Practice
5. Integrated Health Delivery Service Organization
6. Long Term Care
7. Managed Care Organization (MCO)
8. Mental and Behavioral Services
9. Other

a. don't know
b. I choose not to answer

Which of the following Internet Technologies are priorities during the next year:

Upgrading Security of medical information for HIPAA compliance
1. High Priority
2. Medium Priority
3. Low Priority
4. Not a Priority

a. don't know
b. I choose not to answer

Reducing Medical Errors
1. High Priority
2. Medium Priority
3. Low Priority
4. Not a Priority

a. don't know
b. I choose not to answer

Promoting Patient Safety
1. High Priority
2. Medium Priority
3. Low Priority
4. Not a Priority

a. don't know
b. I choose not to answer

Reducing Costs
1. High Priority
2. Medium Priority
3. Low Priority
4. Not a Priority

a. don't know
b. I choose not to answer

Increasing Productivity
1. High Priority
2. Medium Priority
3. Low Priority
4. Not a Priority

a. don't know
b. I choose not to answer

Internet Tools

Which of the following financial-focused Internet Technology tools have/ do you plan to implement:

Connectivity to payers
1. Have implemented
2. Plan to implement within 1 year
3. No plans to implement but interested in learning more
4. No interest

a. don't know
b. I choose not to answer

Assistance in coding patient visits
1. Have implemented
2. Plan to implement within 1 year
3. No plans to implement but interested in learning more
4. No interest

a. don't know
b. I choose not to answer

Electronic charge capture
1. Have implemented
2. Plan to implement within 1 year
3. No plans to implement but interested in learning more
4. No interest

a. don't know
b. I choose not to answer

Which of the following clinically focused Internet tool have or do you plan to implement:

Document scanning/imaging
1. Have implemented
2. Plan to implement within 1 year
3. No plans to implement but interested in learning more
4. No interest

a. don't know
b. I choose not to answer

Transcription/voice recognition
1. Have implemented
2. Plan to implement within 1 year
3. No plans to implement but interested in learning more
4. No interest

a. don't know
b. I choose not to answer

Electronic team messaging between clinic staff
1. Have implemented
2. Plan to implement within 1 year
3. No plans to implement but interested in learning more
4. No interest

a. don't know
b. I choose not to answer

Electronic lab order entry
1. Have implemented
2. Plan to implement within 1 year
3. No plans to implement but interested in learning more
4. No interest

a. don't know
b. I choose not to answer

Electronic routing of test results
1. Have implemented
2. Plan to implement within 1 year

3. No plans to implement but interested in learning more
4. No interest

a. don't know
b. I choose not to answer

Electronic medical record
1. Have implemented
2. Plan to implement within 1 year
3. No plans to implement but interested in learning more
4. No interest

a. don't know
b. I choose not to answer

Electronic Prescribing
1. Have implemented
2. Plan to implement within 1 year
3. No plans to implement but interested in learning more
4. No interest

a. don't know
b. I choose not to answer

Point–of-Care decisions support tools
1. Have implemented
2. Plan to implement within 1 year
3. No plans to implement but interested in learning more
4. No interest

a. don't know
b. I choose not to answer

Which of the following patient-focused Internet Tools do you have or plan to implement:

Incoming telephone call management
1. Have implemented
2. Plan to implement within 1 year
3. No plans to implement but interested in learning more
4. No interest

a. don't know
b. I choose not to answer

Automated telephone appointment reminders
1. Have implemented
2. Plan to implement within 1 year
3. No plans to implement but interested in learning more
4. No interest

a. don't know
b. I choose not to answer

Automated patient notification of test results
1. Have implemented
2. Plan to implement within 1 year
3. No plans to implement but interested in learning more
4. No interest

a. don't know
b. I choose not to answer

Automated telephone patient reminders for health prevention
1. Have implemented
2. Plan to implement within 1 year
3. No plans to implement but interested in learning more
4. No interest

a. don't know
b. I choose not to answer

Electronic communication between physicians and patients
1. Have implemented
2. Plan to implement within 1 year

3. No plans to implement but interested in learning more
4. No interest

a. don't know
b. I choose not to answer

Internet site with health information links for patients
1. Have implemented
2. Plan to implement within 1 year
3. No plans to implement but interested in learning more
4. No interest

a. don't know
b. I choose not to answer

In general, what have been the benefits for the health service of your patients using IT applications?

Patients assuming more responsibility for monitoring their symptoms/disease?
1. High Benefit
2. Medium Benefit
3. Low Benefit
4. Not a Benefit

a. don't know
b. I choose not to answer

Shorter consultations
1. High Benefit
2. Medium Benefit
3. Low Benefit
4. Not a Benefit

a. don't know
b. I choose not to answer

Patients not seeking medical help when it was not needed
1. High Benefit
2. Medium Benefit
3. Low Benefit
4. Not a Benefit

a. don't know
b. I choose not to answer

Patients coming in sooner for necessary treatment
1. High Benefit
2. Medium Benefit
3. Low Benefit
4. Not a Benefit

a. don't know
b. I choose not to answer

Fewer unnecessary tests
1. High Benefit
2. Medium Benefit
3. Low Benefit
4. Not a Benefit

a. don't know
b. I choose not to answer

Fewer unnecessary treatments
1. High Benefit
2. Medium Benefit
3. Low Benefit
4. Not a Benefit

a. don't know
b. I choose not to answer

Fewer errors
1. High Benefit
2. Medium Benefit
3. Low Benefit
4. Not a Benefit

a. don't know
b. I choose not to answer

Increased productivity
1. High Benefit
2. Medium Benefit
3. Low Benefit
4. Not a Benefit

a. don't know
b. I choose not to answer

Reduced costs
1. High Benefit
2. Medium Benefit
3. Low Benefit
4. Not a Benefit

a. don't know
b. I choose not to answer

Barriers to Implementation

To what extent are the following barriers to implementing IT applications:

Lack of Financial Support
1. Not a barrier
2. Barrier easily overcome
3. Barrier overcome with some effort
4. Barrier overcome with great effort
5. Insurmountable barrier

a. don't know
b. I choose not to answer

Vendors inability to effectively deliver an acceptable product
1. Not a barrier
2. Barrier easily overcome
3. Barrier overcome with some effort
4. Barrier overcome with great effort
5. Insurmountable barrier

a. don't know
b. I choose not to answer

Acceptance by the staff
1. Not a barrier
2. Barrier easily overcome
3. Barrier overcome with some effort
4. Barrier overcome with great effort
5. Insurmountable barrier

a. don't know
b. I choose not to answer

Difficulty proving quantifiable benefits
1. Not a barrier
2. Barrier easily overcome
3. Barrier overcome with some effort
4. Barrier overcome with great effort
5. Insurmountable barrier

a. don't know
b. I choose not to answer

Lack of a strategic plan for introducing application
1. Not a barrier
2. Barrier easily overcome
3. Barrier overcome with some effort
4. Barrier overcome with great effort
5. Insurmountable barrier

a. don't know
b. I choose not to answer

Recruiting experience IT personnel
1. Not a barrier
2. Barrier easily overcome
3. Barrier overcome with some effort
4. Barrier overcome with great effort
5. Insurmountable barrier

a. don't know
b. I choose not to answer

Retaining experience personnel
1. Not a barrier
2. Barrier easily overcome

3. Barrier overcome with some effort
4. Barrier overcome with great effort
5. Insurmountable barrier

a. don't know
b. I choose not to answer

Insufficient knowledge of IT applications
1. Not a barrier
2. Barrier easily overcome
3. Barrier overcome with some effort
4. Barrier overcome with great effort
5. Insurmountable barrier

a. don't know
b. I choose not to answer

Requirement of a considerable investment in IT applications
1. Not a barrier
2. Barrier easily overcome
3. Barrier overcome with some effort
4. Barrier overcome with great effort
5. Insurmountable barrier

a. don't know
b. I choose not to answer

This work was previously published in the International Journal of Healthcare Information Systems and Informatics, edited by J. Tan, Volume 1, Issue 3, pp. 1-23, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 4.17
Envisioning a National e-Medicine Network Architecture in a Developing Country: A Case Study

Fikreyohannes Lemma
Addis Ababa University, Ethiopia

Mieso K. Denko
University of Guelph, Canada

Joseph K. Tan
Wayne State University, USA

Samuel Kinde Kassegne
San Diego State University, USA

ABSTRACT

Poor infrastructures in developing countries such as Ethiopia and much of Sub-Saharan Africa have caused these nations to suffer from lack of efficient and effective delivery of basic and extended medical and healthcare services. Often, such limitation is further accompanied by low patient-doctor ratios, resulting in unwarranted rationing of services. Apparently, e-medicine awareness among both governmental policy makers and private health professionals is motivating the gradual adoption of technological innovations in these countries. It is argued, however, that there still is a gap between current e-medicine efforts in developing countries and the existing connectivity infrastructure leading to faulty, inefficient and expensive designs. The particular case of

Ethiopia, one such developing country where e-medicine continues to carry significant promises, is investigated and reported in this article.

INTRODUCTION

Healthcare consumers in general tend to seek access to affordable health services that will meet their needs. From an ethical standpoint, healthcare has to be available when and where consumers need it; physical separation between consumers and healthcare facilities must not pose severe limitations on the delivery of efficient healthcare, even if patients are located in remote areas. In this sense, information and communications technology (ICT) has been demonstrated to offer a competitive choice for accessing affordable and effective health services, especially when access is difficult and limited (Horsch & Balbach, 1999; Kirigia, Seddoh, Gatwiri, Muthuri, & Seddoh, 2005; Tan, Kifle, Mbarika, & Okoli, 2005). More recently, with the continued maturity of network such as Integrated Services Digital Network (ISDN) and Asynchronous Transfer Mode (ATM) networks and related technologies (Perednia & Allen, 1995; Tan, 2001), e-medicine implementation has entered a stage where both the health providers and consumers can now benefit significantly.

IT-based horizontal and vertical communications among the healthcare facilities within the organizational structure of the healthcare system are essential. Such communications facilitate efficient information exchange and help the delivery of essential health services to underserved rural areas. These communications can be supported through a nationwide e-medicine network that is based on affordable telecommunications infrastructure. The network should connect all regional clinics to urban area hospitals. The benefits of such a network include: (a) establishing reliable horizontal-vertical communications and information sharing among facilities, thereby driving up quality, improving efficiency, and enhancing cost-effectiveness of services; (b) achieving e-health commitments and bringing healthcare closer to underserved and un-served rural areas; (c) strengthening collaboration among hospitals within a multi-provider care management context; (d) minimizing long distance travels among rural people in need of proper medical care to urban areas or the capital city; and (e) providing medical information to clinical practitioners that will help them keep abreast of clinical breakthroughs as well as new technological advances.

For Ethiopia, a lesser developing country with significant challenges in meeting basic healthcare needs, it is argued that e-medicine development is emerging and can be fruitfully cultivated over the coming years if a vision and long-term strategy for this technology can be used to help increase the number of citizens receiving care and decrease the subsequent healthcare costs. This article lays out such a vision and strategy for a nationwide e-medicine infrastructure to be designed. It is organized as follows. First, the background and various design considerations for a nationwide e-medicine network is presented. Next is an overview of the requirements for the network design followed by a more in-depth description of the local and wide area network (LAN/WAN) architecture envisioned. The focus of the discussion will then shift to the existing Broadband Multimedia Network (BMN) and Very Small Aperture Terminal (VSAT) infrastructures and how these networks may be integrated into the nationwide e-medicine infrastructure. Finally, the article concludes with insights into potential future work in e-medicine for developing countries.

E-MEDICINE NETWORK DESIGN CONSIDERATIONS

E-medicine refers to the electronic delivery of healthcare and sharing of medical knowledge over a distance employing ICT. A national e-medicine network allows sharing and exchanging of clinical data among physicians, administrators, even patients or other participating health professionals regardless of physical distance separation or geographical terrain of the whereabouts of these network participants within the national boundaries. The network also facilitates communications among physicians and academics across diverse cultures, affiliated healthcare organizations, and publicly or privately funded research institutions. Since there is lack of transportation and communication infrastructure in developing countries, medical and clinical data exchanges can be further secured and facilitated through an existing e-medicine network.

In developed nations, e-medicine services (Wright, 1998) can benefit remote locations that may not be easily accessed due to unpredictable or harsh weather conditions found during certain times of the year, for example, parts of North America and Scandinavian countries are often heavily affected by snow and other natural hazards such as avalanches, falling boulders and closure of highways due to multi-vehicle accidents or other calamities. Mountainous terrain in certain parts of North American regions such as Alaska, British Columbia, Alberta and New Territories implies the need for viable distance healthcare delivery solutions. E-medicine allows health professionals around the world to establish faster communication and exchange information with clients and regional authorities irrespective of geographical locations. It may also support rural dwellers to get healthcare services delivery similar to their urban counterparts. A mobile e-medicine system, for instance, provides a convenient platform for acquisition, transmission and delivery of health-related data to healthcare providers through 2G/3G-based wireless networks (Wootton, 2001). Recognizing these benefits, the International Telecommunications Union (ITU) has set a global agenda to promote e-medicine applications in developing countries. Ethiopia, one of the beneficiaries of such an initiative, has commissioned some ICT projects such as SchoolNet, WoredaNet and BMN to enable fully-fledged connectivity to make better use of the ICT in the health and education sectors.

Network Architecture

The design of suitable national e-medicine network architecture in a lesser developing country such as Ethiopia requires several key components to be integrated into a dynamic and enterprising communication infrastructure. Major architectural components include: (1) LAN architecture for local networking and sharing of health-related information; here, communications may be established using wireless cellular or ordinary fixed telephone lines; (2) WAN architecture for national networking and sharing of health-related data; this will allow communications among local and national physicians, healthcare workers and clients covering urban and rural communities; and (3) designing a suitable back-end database and front-end user interface applications that integrate seamlessly with the (prototype) implementation of the proposed architecture. The overall goal of the nationwide e-medicine network architecture is then to provide an affordable and a low-cost system that facilitates uninterrupted communications among physicians and health professionals across the country. The system bolsters connectivity among rural clinics and urban area hospitals to support primarily clinical e-consultation and maintenance of stored patient records.

This network should also be cost-effective, expandable, and secure. It must support a state-of-the-art ICT access schema and connectivity to rural area clinics. Existing ICT infrastructure will be given priority to minimize the cost of implementing the nationwide network. In the Ethiopia design, expandability is a concern. First, few hospitals are built in the country while more clinics are being added every year. Moreover, there is a chance to incorporate private hospitals in the nationwide e-medicine network as and when necessary, which will further increases the number of future connected sites. As well, the area of e-medicine applications will not be limited to just some specific diseases, but will be expected to increase in type and number over the long haul. In fact, the network should also support advanced applications, which require real-time connectivity such as videoconferencing capabilities for future use.

During e-consultation or patient referral, most of the data exchanged over the network are sensitive patient information. Confidentiality of patient information must therefore be respected. For secure communications, protocols such as Secure Socket Layer (SSL) could be used. SSL ensures secured communications over web-based applications and provides the ability to safely exchange patient information across the network (Elmasri, 2000). When doctors exchange patient information, they could adhere to medical protocol that defines the rules to be followed during this process. In addition, the network and accompanying servers could be protected by firewall against hacking from external parties. Firewalls are software or hardware for the sole purpose of keeping digital pests such as viruses, worms, and hackers out of the network (http://www.cisco.com, Tanenbaum, 2004).

NETWORK DESIGN REQUIREMENTS

As cost must be one of the driving factors for choosing among existing or emerging ICT infrastructures in the country, implementing nationwide e-medicine network infrastructure may seem at first to be more expensive than building clinics or supplying existing regional clinics with medical personnel and equipment. Yet, a cost-benefit analysis comparing various IT investment approaches will provide best directions to achieving a lower cost solution to the problem of delivering adequate and proper healthcare and disseminating confidential health information to and from various connecting points throughout the country. With today's oil prices at a premium, network connectivity among the healthcare facilities, both in the urban and rural areas over an existing ICT infrastructure is now considered a cost-effective solution. Of course, set-up costs depend on the type of WAN to be used—to ensure low installation cost, it is proposed that the network design will incorporate an existing WAN provided by the Ethiopia Telecommunication Corporation (ETC).

In Ethiopia, most of the inter-hospital communications are traditionally dependent on telephone and hand-delivered referral messages. During referrals patients have to travel afar to one of the urban referral hospitals, carrying the written messages of the referring physician. Clinics located in the telephone coverage areas communicate using telephone to exchange information about availability of specialist(s) or bed in another hospital. Yet, the communication needs of hospitals have grown over the years ahead of its technological capabilities. Geographically dispersed clinics lack modern telecommunication technology access. Among them are instantaneous access to patient information, access to electronic medical records, and access to the Internet. These and other communication needs of health providers also require the development of e-medicine application software backed by electronic patient record systems. Design of such communication networks will also require the understanding of organizational structure of the clinics involved in the network.

Since the government/public clinics are owned and organized under their respective regions, the WAN design should follow the organizational structure of the administrative regions in the country. A detailed study about the inclusion of various clinics, their locations relative to the nearest access point to existing ICT infrastructure, traffic load and its characteristics, security, LAN/WAN protocol, topology and bandwidth requirements and utilization, and allocation of bandwidth, among other issues, have to be considered while trying to design a nationwide e-medicine network architecture. For example, issues of communicating patient information electronically may further raise question on medical ethics, the need for developing standard medical protocols, and detailing policies for use in routine activity via the e-medicine network.

LAN Architecture

To design the LAN for each hospital, we consider the central site, *Tikur Anbassa Specialized Hospital* located in the capital Addis Ababa, as a model. The hospital is organized into 16 departments with each department further divided into smaller units as necessary. For instance, the Internal Medicine department has several units such as the Renal Unit, the Cardiology Unit, the Neurology Unit, and other units. Physicians in these departments and units need to communicate whenever a patient visits more than one of the units. It is proposed that the LAN follows the hierarchical structure of the hospital.

The decision to make the selection between various LAN technologies was based on: (a) expected application to run on the network and their traffic patterns; (b) physical locations of the offices and users to be connected in campus; (c) the rate of network growth; (d) the abundance of the network technology in the market; and (e) simplicity of installation and maintenance. Each of these criteria will now be explored in more depth.

Expected Application to Run on the Network and their Traffic Patterns

Currently we expect a Web-based e-medicine application to run on the network. The application will use a central database server where all the user and patient information will be stored. The type of data to be transmitted on the network should accommodate both text and image formats. Since all communications are to be channeled through the server, the traffic pattern around the center is expected to be heavy. Higher speed devices should be installed at the center of the LAN where servers will be located.

Physical Locations of Offices and Users to be Connected on Campus

The sample hospital (*Tikur Anbassa Specialized Hospital*) is housed in a series of five buildings (Blocks A-E). These blocks are not physically separated. Even though precise figures were unavailable, these five buildings are built on roughly 8,000 to 10,000 square meters. While the main offices and departments in the hospital are located in one of the respective blocks, most of these offices are in either of the first two stories of the block they belong. Having routers switches in each of the departments is ideal to design a high-speed and expandable LAN, but it will also make the design expensive to install, support and maintain. A more cost-effective approach is to put switches per building and then get the departments to be connected into various groups by using Virtual LAN technology.

The Rate of Network Growth

The rate of the hospital LAN growth depends on the level of computerization in the hospital. Currently in this central hospital site, there is a LAN that connects a few offices and a computer room. The network employs star topology, using a centrally located hub and Unshielded Twisted Pair (UTP) cables forming a peer-to-peer LAN. The purpose of this LAN was to enable offices to share printer and students to get access to research documents. In this design, it is anticipated that as the use of Web-based applications becomes commonplace, there will be opportunities to add more applications and connect more computers and offices to the LAN. The switches-routers selected in this design should therefore have many free ports to help cascade the growing number of anticipated future connections.

The Abundance of Network Technology in the Market

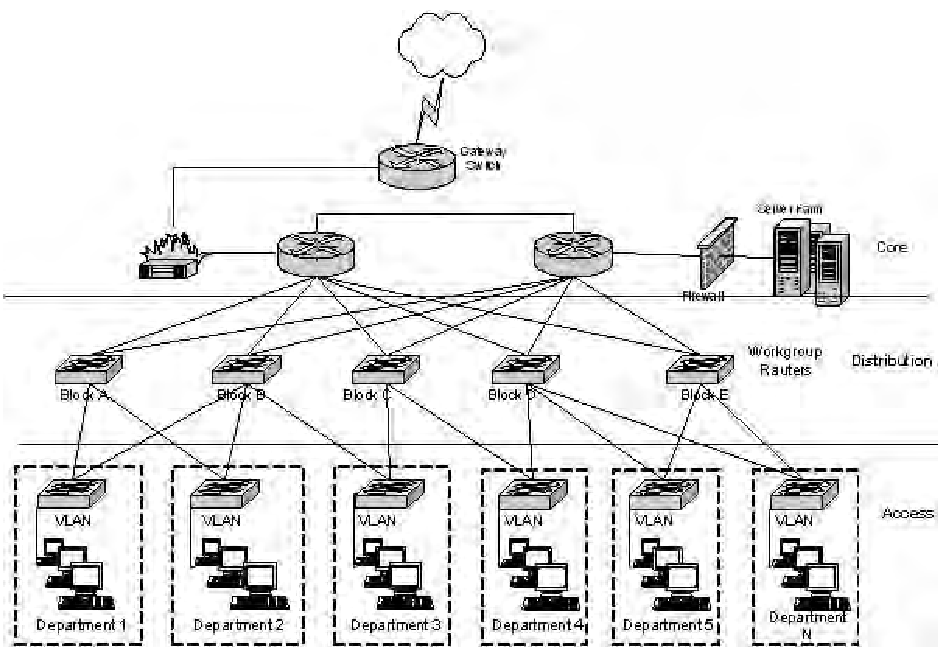
Capitalizing on the abundance of emerging network technology in the Ethiopia market, we gathered data from existing network technology vendors and organizations that implement computer networks in the capital city, Addis Ababa. Ethernet technology is common in organizations that implement computer networks, such as Addis Ababa University (AAU Net). AAU Net is a network backed by triangular shape fiber optic cable connecting the three main campuses. The topology is an extended star topology that fastens together fiber optic cables for vertical cabling (backbone cabling) between buildings that house various faculties and departments. These backbone cabling provide interconnection between wiring closets and Point of Present (POP). The zones that fall within a departmental area are served by internetworking devices such as hubs and UTP cables.

Interestingly, what is described so far appears to be the dominant design of the small number of networks existing in the capital Addis Ababa. As such, it is not surprising to find that suppliers of network technology devices and support in Ethiopia are restricted to only a limited number of vendors.

Simplicity of Installation and Maintenance

To design the LAN architecture we have therefore selected the hierarchical model. This enables us to design and arrange the inter-network devices in layers. Figure 1 depicts the hospital-based LAN architecture.

Figure 1. Hospital LAN design



It is a model preferred by most of network design experts for its ease of understanding, expandability and improved fault isolation characteristics (http://www.cisco.com). The model encompasses the following three layers: (a) first layer (core layer); (b) second layer (distribution layer); and (c) third layer (access layer).

Core Layer

Core layer high performance switches, capable of switching packets as quickly as possible, are to be deployed. Essentially, this layer connects the LAN backbone media as well as connects to the outside world via a firewall through WAN. In this design, the devices in the core layer will be placed at a central location in the hospital. The devices will then be connected with high-speed cables such as fiber optics, or fast Ethernet cables. The servers will also be connected to switches, shielded by a firewall.

Distribution Layer

Distribution layer will contain switches and routers capable of Virtual LAN (VLAN) switching and allow defining departmental workgroups and multicast domains. The devices should also support connectivity of different LAN technologies since they also serve as the demarcation point between the backbone connections in the core layer and the access layer. In this hospital-based LAN design, the distribution layer represents switches/routers at each building connected to the core layer on the one end and to the access layer on the other. Redundant links will be used for maximum availability and the departments could be grouped forming their own Virtual LAN.

Access Layer

Access Layer is where the end-users are allowed into the network. This layer contains switches/hubs from which PCs in each department gain access to the hospital-based LAN. Each department will have at least one switch/hub, which will in turn have redundant links to more than two of the switches in the distribution layer.

WAN Architecture

Designing the WAN architecture for a nationwide e-medicine network raises the issue of WAN service provider. Unlike LAN, WAN connectivity depends on the availability of WAN infrastructure in the country. The sole WAN service provider is the Ethiopian Telecommunications Corporation (ETC).

ETC provides a number of services (http://www.telecom.net.et) from which the WAN infrastructure suitable for the e-medicine network may be derived. Existing WAN services include: (a) Internet Services, or, providing basic Internet services over dial-up or leased lines; (b) Digital Data Network (DDN), supporting dedicated Internet, ISDN and frame relay services; (c) SchoolNet VSAT, covering services for secondary schools and institutes of higher learning; (d) WoredaNet VSAT, covering services for districts (Woreda) administrations; and (e) Broadband Multimedia Network (BMN), offering high-speed optical communications to major cities.

To choose among these possible infrastructures for nationwide e-medicine network, the parameters to be considered include the geographical coverage, bandwidth, mode of communication, rental cost of WAN connection and capacity to add more LANs. Table 1 summarizes the data comparing among the available ICT infrastructures in Ethiopia.

Based on the data, it appears that WoredaNet is best suited to the national e-medicine network, as long as the existing infrastructures are functioning efficiently and effectively. However, as noted in Table 1, there may be a tradeoff between coverage and capacity, that is, when the coverage is acceptable the capacity may be somewhat limited. For example, BMN coverage is ideal as it represents

Table 1. Summarized comparison of existing ICT infrastructure

	Internet	DDN	SchoolNet	WoredaNet	BMN
Coverage	Telephone coverage areas only	The capital and regional Urban areas only	About 500 schools covered. There are Woredas that do not have schools	571 Woredas out of 594 are covered	The Capital city and 13 regional towns.
Bandwidth	Maximum of 56k dialup and 1Mbps in Leased line	Maximum of 1Mbps	Can be upgraded to 384k upstream	Downstream/ up-stream 45Mbps/ 256k downlink	ADSL Services: Variable bandwidth Downstream/ upstream 512k/128k and 1024k/256k
Interactivity	Two-way	Two-way	One-way broad-casting	Two-way	Two-way
Cost	0.11 birr/min dialup 1000 birr/ month leased line		Free for schools	Free For Woredas	Not yet determined, under devel-opment
Capacity to scale	Not scalable	Not scalable enough		Will have more than 10 ports free at each Woreda	Can be expanded

state-of-the-art service and higher bandwidth. However, it is centered primarily in the urban areas. It is also under development and we have thus considered it as a potential option to be used when integrated with the VSAT-based networks to enhance nationwide e-medicine network. Finally, the SchoolNet needs to be upgraded to support two-way interactivity.

Thus, one alternative approach is using a combination of VSAT networks and terrestrial BMN. VSAT-based connectivity is believed to be cost-effective and in the case of WoredaNet and SchoolNet, it enables connectivity to the public, even in the rural areas. In addition to serving the rural areas, it also covers urban areas. Together, this will provide modern, convenient as well as economical connectivity to hospitals. For improving state-of-the-art applications such as videoconferencing, connectivity via the emerging BMN is proposed to connect urban area hospitals in the capital city and in the regions where the network can be easily accessible.

Figure 2a shows the e-medicine network as a first alternative. Note that this approach requires that urban hospitals maintain two WAN connections. Having more than one WAN connection, however, may become expensive in the long run. If the two WAN infrastructures could be integrated, an improved WAN design will result with only one WAN connection to the urban hospitals through which the hospitals will be connected to BMN and the rural area clinics through the WoredaNet. Figure 2b depicts the second alternative solution of the WAN design.

CURRENT ETC INITIATIVES

The recent development in the ETC in providing multimedia network infrastructure is the integration of the VSAT-based networks (SchoolNet and WoredaNet) with the BMN (Tiruneh, 2006). In other words, these VSAT-based networks can now be used as a point of access to the BMN.

Figure 2a. Logical WAN design based on BMN and VSAT networks (1st Alternative)

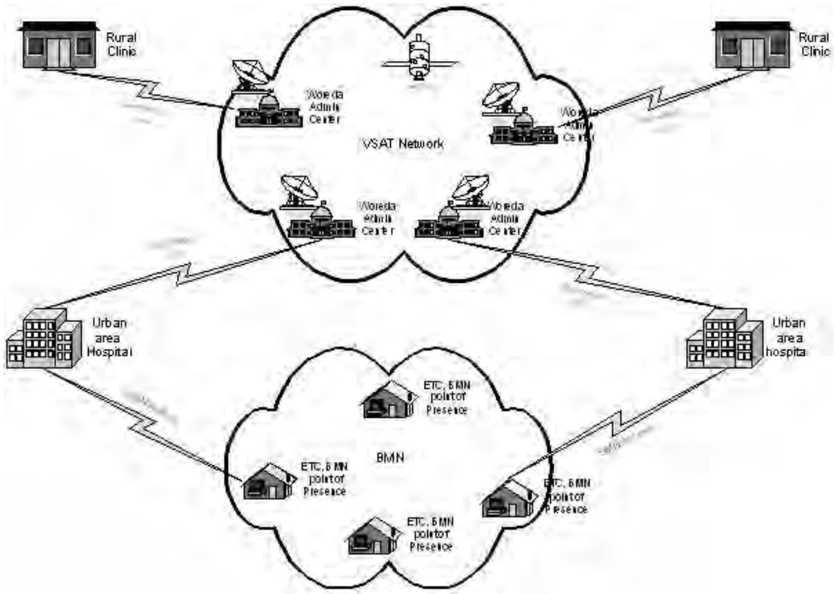
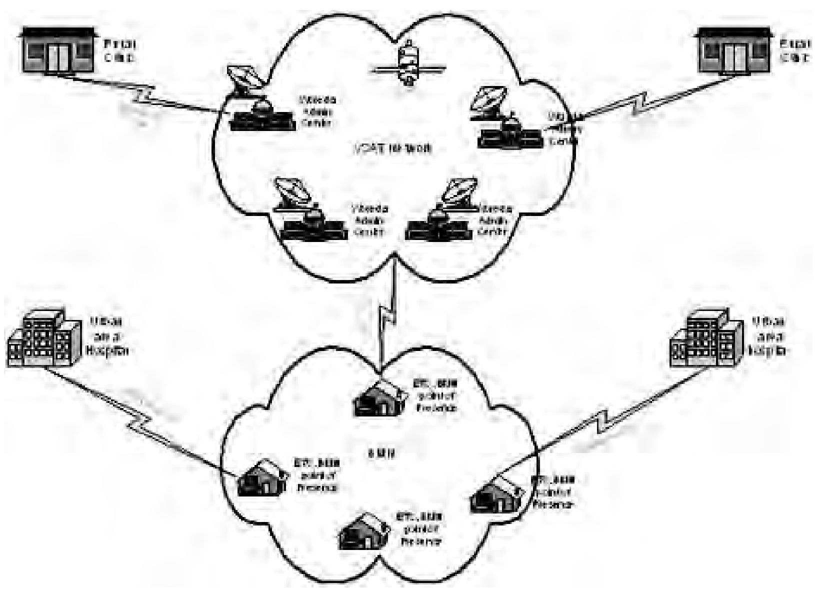


Figure 2b. Logical WAN design based on BMN and VSAT networks (2nd Alternative)



As part of a longer term vision and mission of broadband initiatives for socioeconomic development in Ethiopia, ETC has also planned an e-health setting that tries to cover rural areas, schools, clinics, hospitals, prisons, and nursing homes, including assisted living with several requirements: (a) high quality patient data, video and images to be exchanged between different medical institutions; (b) ICT infrastructure to connect geographically dispersed institutions, nationwide or worldwide; (c) infrastructure that supports data, video and voice/audio (multimedia) services; (d) high quality, secured and fast delivery of medical information; and (e) high speed (BW) connectivity or the deployment of broadband infrastructure.

Key challenges faced are the need to encompass multiple locations, to use multiple access technologies, to deliver multiple services and to address multiple user markets.

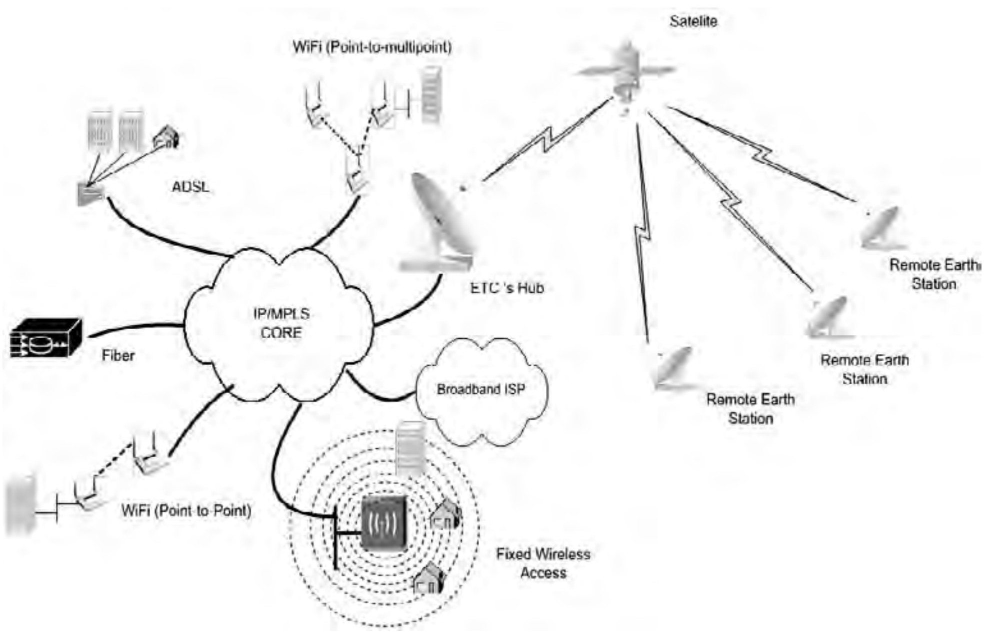
Current BMN Development

ETC has already completed building the Core Terrestrial Broadband Infrastructure, which is capable of providing data, video, and voice services with 24 Points of Presence. This infrastructure services key business sites (urban areas) and supports Multiple Broadband Access via ADSL, FWA, WIFI, and fiber networks.

Current VSAT Development

In Ethiopia, a Broadband VSAT Network platform, which supports integrated services such as video, data, Internet, and voice on a single infrastructure, is currently in place. It has countrywide coverage (450+ schools and 550+ woredas) as part of SchoolNet and WoredaNet deployments. It is integrated with the core multimedia network, also serving as broadband access means. Figure 3 shows the two recent developments.

Figure 3. ETC's broadband-integrated infrastructure for e-health applications



Apparently, the new developments in WAN infrastructure support the 2nd alternative e-medicine WAN architecture discussed here as depicted in Figure 3.

THE PROTOTYPE

Based on the specified network requirements and architecture, a working prototype for the national e-medicine network is now presented and its operations highlighted.

The prototype is a basic e-Medicine service (BEMS), which provides a Web-based graphical user interface (GUI) for health providers. BEMS facilitates the information exchange between remotely located health providers for the purpose of e-consultation, as well as for maintaining electronic patient information. The traditional paper-based forms and patient cards used in the hospitals will be digitized and reproduced electronically. Web-based technology is chosen for its ubiquity. Using Web-based technology constitutes not only a network that can be used universally, but the technology also supports system-independent platforms, thus providing access to many different computer systems at client sites (http://java.sun.com/products/jsp). Key requirements in these client sites are simply the availability of Web browser software and network connectivity.

For a secured network, password protected system ensures that user login is needed to access capabilities of the system. In addition, user types are defined so that there will be a role-based access to database and system functions in BEMS. To ensure compatibility with most legacy systems, a relational database is advocated for storing user and patient information. Beyond e-mails, this approach allows the users to mobilize structured information exchange among the communicating health providers.

Major Features of BEMS

Basically the BEMS prototype may be conceptualized as a database-driven Web site with the following main features and functions: (a) providing user management services where administrator can register users, assigning username and password, and defining user type, as well as searching and editing user information; (b) providing patient management services where health providers can register patients, search patients and view patient information on a Rolodex-like interface, as and when necessary; (c) providing, on the one hand, referral systems where physicians can write referral messages to a particular department and hospital, and, on the other, a system whereby a physician can retrieve and study the list of referrals forwarded to the department s/he is working in and allowing the physician to write feedback instantaneously after examining the referral message and patient information; (d) providing a system by which physicians can request and schedule lab test at any hospital laboratories so that patients can get tested in the clinic/hospital they are being treated; and (e) providing a list of lab test requests to laboratory technicians and allowing them to input lab test results.

BEMS Architecture

The BEMS architecture is built on three-tiered, client-server architecture. The first layer is where the client machines run Web-browser software. This layer is used to display the user interface (Web pages) of the system and send secure HTTP request to the Web server in the second layer. Along with the Web server, application server resides in the second layer. This application server manages the clinical business logic. The bottom layer contains the persistent data of the system. All data of patients, physicians and other communicated messages will be stored and maintained in this third layer. This layer runs the database management system (DBMS) software. Put simply, BEMS functions as a Web-based application connected to a Web server to provide all the interfaces of the system and that of a database server to manage all the knowledge and information elements stored in the system. Figure 4 charts the BEMS system architecture.

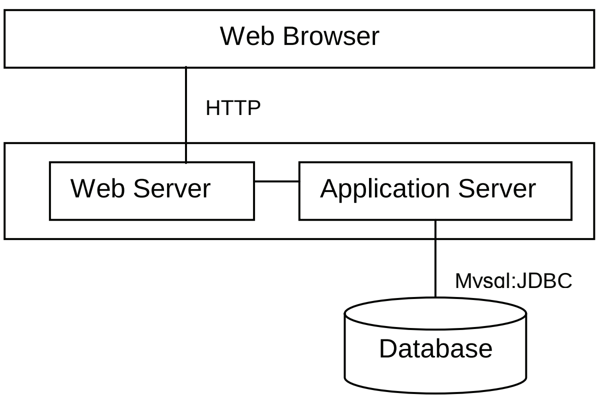
The BEMS prototype is constructed with a combination of open source products and freely available software components. The Web server suggested is the Apache Jakarta's Tomcat Web server (http://jakarta.apache.org) with the functionalities as well as the mandated business rules programmed in Java (Haile-Mariam, 2002). Java Server Pages (JSP) is used to capture the user interface and the text of Web pages (http://java.sun.com/products/jsp; http://www.coreservlets.com). Some scripting is included on the Web pages in JavaScript. JSP has a capability to import java classes and run them from the Web pages when the pages are downloaded to the client machine (http://www.coreservlets.com). Unlike other server side languages such as Active Server Pages (ASP), JSP makes the system platform independent. It also allows users to take advantage of the full power of java programming language which overcome some of the limitation of other scripting languages such as PHP (http://www.coreservlets.com).

The database conceptualized is the open source MySql to back up the database driven application. MySql works on many different operating system platforms and is known for its speed of data retrieval (http://www.mysql.com). It provides Application Program Interfaces (API) for many programming languages including Java. Passwords are secure because all password traffic is encrypted when connecting to the MySql server. For database connectivity, we use mm.mysql driver, which is a Java Database Connectivity (JDBC) driver, from MySQL AB, implemented in 100% native Java (http://www.mysql.com/products/connector-j).

Database Design Issues

BEMS needs to keep track of information about patient and related medical records, user's information, and messages for both medical referral requests and feedbacks. A well-designed minimal database is needed to manage this information. A relational database model is selected to store the persistent data of the system, as it could be easier to manage, and provides better management for complex query of such data (Amenssisa & Dabi,

Figure 4. BEMS prototype architecture



2003). This database is expected to maintain and manipulate basic entities such as users, patients, and medical records.

Each component of the medical record of a patient is an aggregation of different types of data, which are stored in the database. In the traditional paper-based system, the medical record of a patient is identified by an Out Patient Card (OPCard) Number, which is usually called patient record number. OPCard is a four-page hard-paper card, which contains patient's generic information, such as name, sex, age, address on the first page and a table of two columns for date and clinical note so as to record chronologically the compilation of health providers' notes. All other components such as laboratory test results and x-ray reports, among other pieces of information, are stored inside the hard-paper card referenced by the card number or name of the patient. The lab test results may contain zero or more test request forms along with the results for Urine, Parasitology, Blood Chemistry, Hematology, Serology, Bacteriology, Fine Needle Aspiration Cytology and Biopsy.

When a patient is admitted to the hospital, admission and social services information is stored. The admission data include identification information and name and address of next-of-kin, marital status, and number of siblings (children) information, besides occupational information and other demographics, as and when provided by the patient. Subsequently, follow-up data such as vital sign measurements, fluid balance information and other measures will be collected and recorded. Order sheet, which contains a list of treatments to be ordered following admission, is also part of the inpatient medical record. In addition to these, information about the hospitals, departments and laboratories are also stored and captured in respective entities.

To minimize connectivity cost and increase system performance, a distributed database is recommended. Horizontal partitioning that splits tables along rows, based on the location of patient and healthcare facility, is seen to be an ideal choice in the e-medicine application that tries to create nationwide connectivity. Finally, to use the database, transparent data access schemes must be defined for applications that run over the network.

BEMS Interfaces

In this part of the discussion, the design of various BEMS interfaces is presented. BEMS is accessed when opening the initial Web page where user authentication is first performed. The initial page contains a typical login screen for specifying username with authenticated password. There is no need of menu or different buttons to be submitted based on the user types. Since the user types are defined in the database when the user registered, the page corresponding to the specific user type will automatically be opened upon successful login. Currently, administrators, physicians, and lab technician user types are defined and all of these user types will have their own main pages as described below.

Administrator's Main Page

The administrator's main page is used for managing users. The functionalities accessible from this page include: *register new user* and *search user* by a combination of name, father's name, and user name. Figure 5a shows the administrator's main page whereas Figure 5b depicts the user registration page.

The other function provided to administrators is the *search user* function. It is possible to search users by keying in any combinations of name, including father's name and/or user name. Note that username is treated as "unique" in the user table with the search result being quickly displayed. Although not shown here, the full name of the search result is captured as a link. This link then leads to a page containing the relevant user information from which the administrator can edit a particular user.

Physician's Main Page

The physician's main page contains a button to open the patient manager page and has the capability of retrieving a list of referrals forwarded to the department where the physician is located. To open a new page, the physician is free to click on the *manage patients* button or link to one of the referrals. Figure 6a illustrates that the patient, Ato Andualem Lemma, is one of Dr. Aman's referrals. If Dr. Aman chooses to treat this patient in the hospital where he is privileged, he can simply open the patient manager page by clicking on the

Figure 5a. Administrator's main page within BEMS

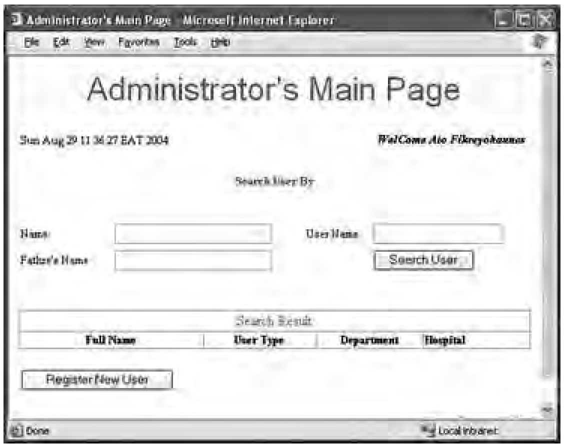


Figure 5b. User registration page

manage patients button.

Figure 6b shows that the patient manager page has two options, that is, physicians can "register new patient" or "search patient" for those who were previously registered. When selecting "register new patient," a patient registration form, similar to the user registration page, will be opened. In contrast, if the physician wants to look for a patient, s/he can input one of the search criteria such as *name* or *record number* of the patient, resulting in the display of a matching record number or name and the hospital where the patient was first registered. The patient full name is a link that leads to the patient information page similar to

Figure 6a. Physician's main page within BEMS

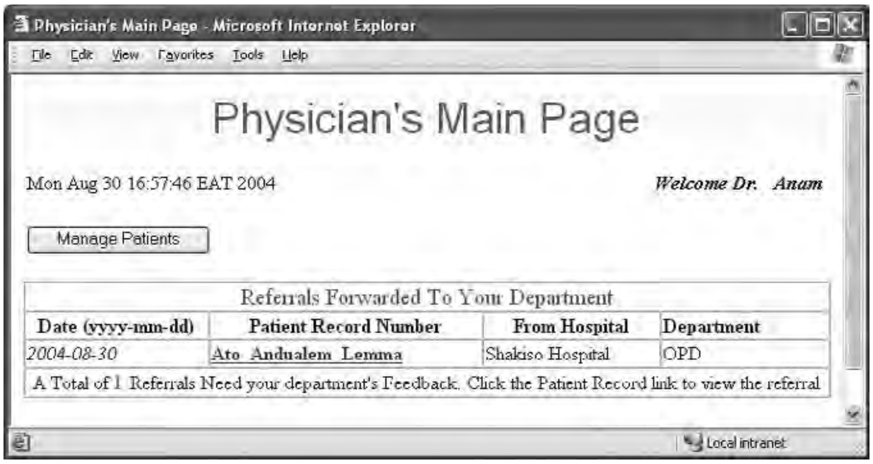


Figure 6b. Patient manager page within BEMS

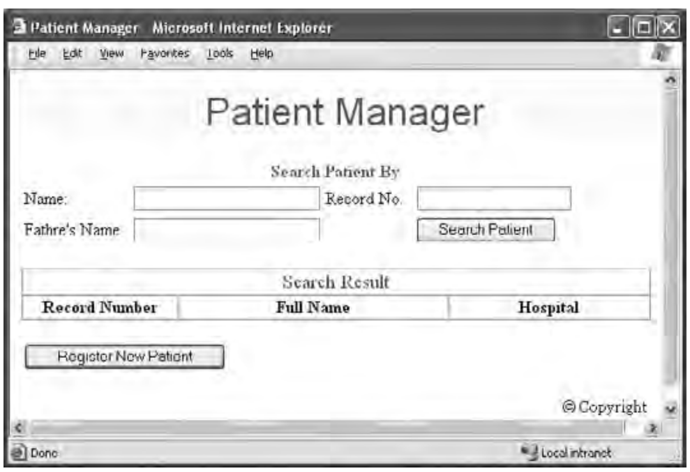
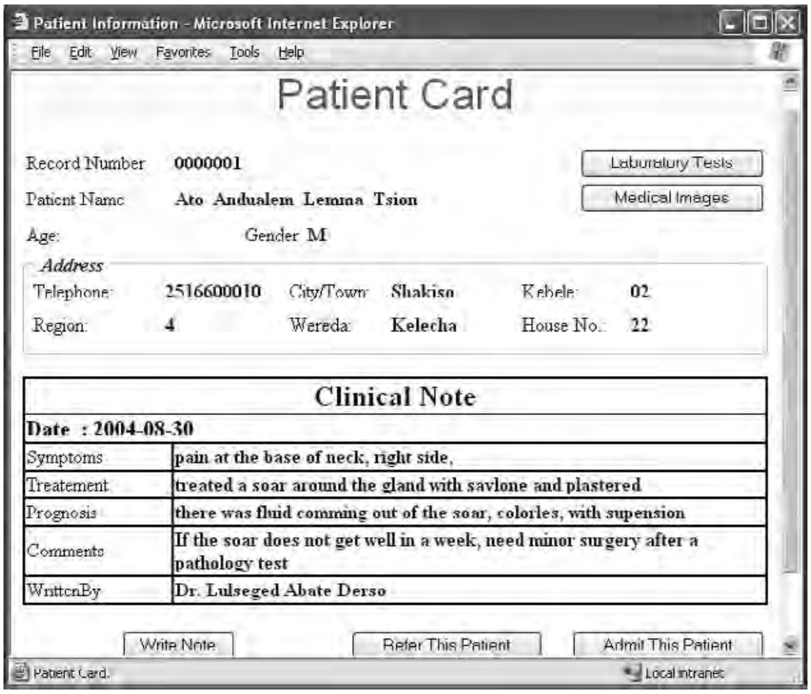


Figure 6c. Patient card page within BEMS



the traditional patient card used in the hospitals. An example of the patient card, which opens up when the full name link in the previous interface is selected, can be seen in Figure 6c.

The patient card contains patient's general information, address information and clinical notes that are ordered in descending order. In addition to the information displayed on the patient card, laboratory test results and medical images related to the patient are accessible by clicking corresponding buttons from the patient card interface. The physician can add clinical notes, refer, and/or admit the patient. From the physician's main page, the other option available to the physician is to see referrals forwarded to his or her department. This is possible by clicking the link that opens a referral page corresponding to the patient. The patient referral page contains the referral messages and buttons that will lead the user to patient information as well as a button that can lead to the feedback input page.

If the physician user wants to view the patient information, the *view patient card* button will support such a function. Otherwise, if the physician would like to give feedback to the referral using the feedback slip, the *open feedback slip* button will serve this purpose. Basically, the feedback slip is represented as an input form similar to a traditional form for capturing feedback information related to the current referral. Put simply, the idea here is to mimic within a virtual environment what the physicians have already become accustomed to routinely when working within a paper-based environment, which will only hasten the processing and matching of the stored computerized information for them.

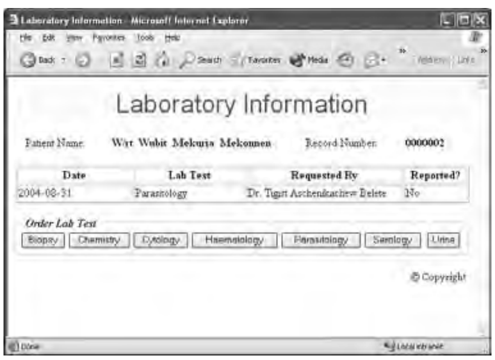
Graphical User Interface

Apparently, human computer interface (HCI) design issues are critical in determining the successful deployment and continuing use of the BEMS prototype. As indicated, the current approach attempts to optimize the interface design in mimicking more or less the traditional forms, documentation formats and paper-based patient record system that the physicians have grown accustomed to using over the years. In other words, this ensures that the BEMS supports the habits of the physician users. It will also serve to preserve physician user satisfaction and promote a high rate of acceptance among physician users with the new system implementation on the one hand while reducing disruption to the care processes on the other.

Nonetheless, new system development such as the BEMS typically provides new opportunities for revisiting the care processes that have been put in place over the years. Elimination of redundant processes as well as the need to streamline certain administrative and clinical processes may be warranted to improve quality, cost, efficiency and effectiveness of the care provided. Online requests of patient and referral information and querying of databases are expectedly translatable into more efficient, effective, appropriate and quality care. Use of GUI also permits substantial amount of data to be viewed together, improving the communication, exchange and sharing of patient data. Feedback from physician users should ultimately be channeled to an even more enhanced user interface design. In this environment, clinical test results and specialist reports can also be captured quickly and shared collaboratively among all relevant health providers.

As an example, the physician can be empowered to request laboratory test results in BEMS by viewing the patient laboratory information page, which is accessible from the patient card page by a button called "Laboratory Tests." Figure 7 provides illustrative screenshots for the laboratory information page and a parasitological test request pages. When a physician wants to request lab tests, instead of writing a prescription to the patient and having the patient wait for further scheduling information from the nursing clerk, all that is needed now is a mere click of the button corresponding to the type of "test" required from

Figure 7. Laboratory information and parasitological test request page of BEMS



the laboratory information page. The specific lab test request page can be designed such as to provide the physician user with a dropdown list from which the appropriate lab test can be immediately selected and performed as scheduled. This was found to be important in order to forward the lab test request on a real-time basis to the other user types called, the Lab technicians.

Lab Technician's Main Page

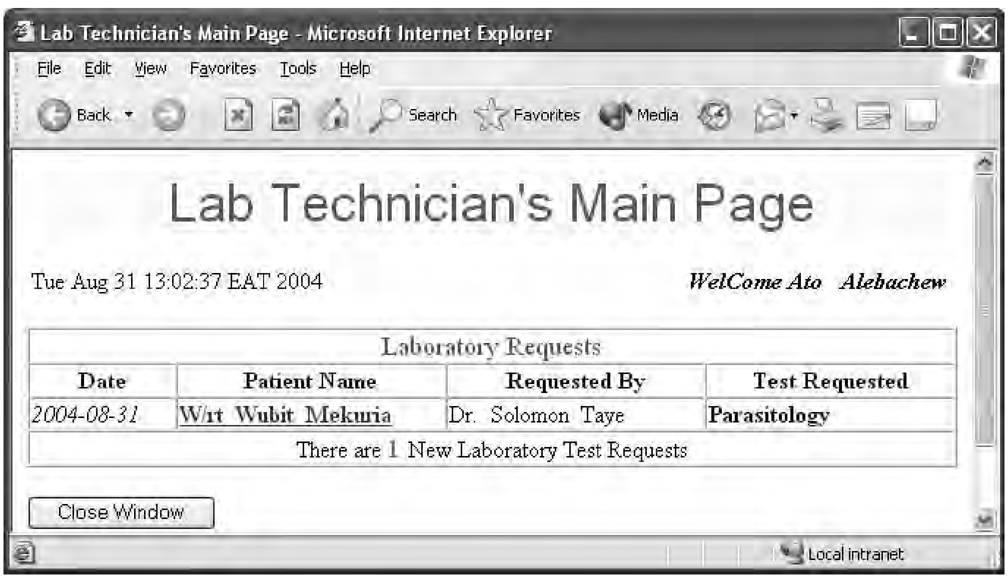
The third type of user, laboratory technician, sees a list of laboratory requests to the department s/he is practicing, on the lab technician's main page. The list contains a link to open lab test result input form where the lab technician can enter his or her report following the test as shown in Figure 8.

CONCLUSION

The key contribution of this effort lies in envisioning the planning and detailing of a Web-interface for a hierarchical model-based LAN architecture that enables the integration of the inter-network devices in layers and a WAN architecture, both fine-tuned for a developing country such as Ethiopia.

The hierarchical model adopted for the LAN is a preferred model due to its ease of expandability and improved fault isolation characteristics. The WAN design considers the existing VSAT-based WAN infrastructure in the country, the WoredaNet. Even if urban areas are relatively better equipped with adequate ICT technologies such as Internet access and digital telephone networks, the communication infrastructure is not well developed in many rural areas. These regions have to be equipped with an access to urban areas. In this context, the newly emerging state-owned, low-cost VSAT networks such as SchoolNet and

Figure 8. Lab technician's main page of BEMS



WoredaNet provide the rural areas with suitable means of communication with urban areas and beyond.

VSAT, an earthbound station used in satellite communications of data, voice and video signals, excluding broadcast television, comprises two parts: (a) a transceiver that is placed outdoors in direct line of sight to the satellite; and (b) a device that is placed indoors to interface the transceiver with the end user's communications device, such as a PC. The transceiver receives or sends a signal to a satellite transponder in the sky. The satellite sends and receives signals from a ground station computer that acts as a hub for the system. Each end user is interconnected with the hub station via the satellite, forming a star topology. The hub controls the entire network operation. For one end user to communicate with another, each transmission has to first go to the hub station that then gets retransmitted via the satellite to the other end user's VSAT. VSAT can handle up to 56 Kbps.

More importantly, the BEMS architecture discussed here is designed to integrate with a large part of the existing LAN and WAN infrastructure designs. The system can then be used to facilitate both intra- and inter-hospital communications and for all forms of information exchange. The alternative design selected will not only improve quality of healthcare services while protecting the privacy, confidentiality and integrity of sensitive patient information, but its interfaces have been set up to mimic the physician routines working in a paper-based environment. Moreover, this will also yield opportunities for further review of the paper flow and work processes to cut down on redundancies and errors while simultaneously boosting both administrative and clinical efficiencies and effectiveness of care.

As future work in the area of developing nationwide e-medicine networks, we recommend the following considerations: (a) the intended network should support real-time e-consultations via video and audio conferencing, advocate doctor-to-patient interactions, and facilitate remote training for health professionals; (b) it should also support a distributed database structure, where individual hospitals should keep their own databases, which can be further treated as one "huge" database; (c) the definition of standards is essential to facilitate information exchange among private and government hospitals as well as overseas; (d) the integration of expert systems such as case-based system where doctors can query the database to get experience from previously stored similar cases should also be considered—such a system will aid future physicians and residents working anywhere in the country to learn from past successes and/or failures of the attending specialist(s), especially for non-trivial and complex patient cases; and (e) the infrastructure should be independent of chosen platform and operating systems (e.g., Windows vs. Apple) and be able to support physicians needing to remotely monitor their patients over heterogeneous networks, including handheld devices in 2G/3G mobile networks and wirelessly.

Beyond the design of a nationwide e-medicine infrastructure, there will be a host of potential e-medicine applications that may be supported, including, but not limited to a series of healthy lifestyle promotion programs such as e-consultation and tracking of participation in smoking cessation, weight reduction, dental health, stress reduction, exercise and nutritional programs and many more. In this regard, one of the contributing authors is actively and precisely engaged with a growing network of researchers in generating such a series of educational modules intended for seniors and other population groups that will eventually be serviced on a network such as BEMS discussed in this article (Tan, 2005).

REFERENCES

Amenssisa, J. & Dabi, S. (2003). District-based telemedicine project in Ethiopia. *Ministry of Health*. Addis Ababa, Ethiopia.

Apache Jakarta Project. (2007). Retrieved from http://jakarta.apache.org

Cisco Documentation. (2007). Retrieved from www.cisco.com

Elmasri, R. (2000). *Fundamentals of database systems*, 3rd edition. Addison Wesley.

Ethiopian Telecommunications Corporation. (2007). www.telecom.net.et

Haile-Mariam, A. (2002). Renaissance: Strategies for ICT Development in Ethiopia. M.Sc Thesis, School of Engineering Postgraduate Engineering Program.

Hall, M. (2007). Core servlet and Java server pages. *Sun Microsystems Press*, Retrieved from http://www.coreservlets.com.

Horsch, A. & Balbach, T. (1999). Telemedicine information systems. *IEEE Trans. Inform., Technol. Biomed., 3*, 166-175.

Java Server Pages Documentation. (2007). Retrieved from http://java.sun.com/products/jsp.

Kirigia, J., Seddoh, A., Gatwiri, D., Muthuri, L., & Seddoh, J. (2005). E-health: Determinants, opportunities, challenges and the way forward for countries in the WHO African region. *BMC Public Health*, *5*, 137.

MySQL. (2007). http://www.mysql.com.

MySQL Connector/J. (2007). Retrieved from http://www.mysql.com/products/connector-j/.

Perednia, D. & Allen, A. (1995). E-medicine technology and clinical applications. *Journal of the American Medical Association*, *273*(6), 483-488.

Tan, J. (2001). *Health management information systems: Methods and practical applications*, 2nd edition. Gaithersburg, MD: Aspen Publishers, Inc.

Tan, J., Kifle, M., Mbarika, V., Okoli, C. (2005). E-Medicine in developed and developing countries. In J. Tan (Ed.), *E-healthcare information systems*. Jossey-Bass (A Wiley Imprint).

Tanenbaum, A. (2004). *Computer networks*, 4th edition. Prentice Hall, Inc.

Tiruneh, M. (2006). ETC, broadband network infrastructure for e-health. *ICT-H-2006 Workshop*, Addis Ababa, Ethiopia.

Wootton, R. (2001). Recent Advances: Telemedicine. *British Medical Journal*, *323*, 557-560

Wright, D. (1998). Telemedicine and developing countries. A report of study group 2 of the ITU development sector. *J Telemedicine Telecare*.

This work was previously published in the International Journal of Healthcare Information Systems and Informatics, edited by J. Tan, Volume 3, Issue 1, pp. 44-62, copyright 2008 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 4.18
Kaitiakitanga and Health Informatics:
Introducing Useful Indigenous Concepts of Governance in the Health Sector

Robyn Kamira
Paua Interface Ltd. & Rangatiratanga Canvases Ltd., New Zealand

ABSTRACT

Indigenous contributions to governance in health informatics can be drawn from cultural concepts such as Kaitiakitanga, which implies guardianship, stewardship, governance and responsibility roles. This chapter explores Kaitiakitanga, its potential implementation in the Aotearoa (New Zealand) health sector, and its contributions to our thinking. After decades of unsuccessful attempts to positively shift the status of health for Maori, we must ask whether more control by Maori over information about Maori will make a difference. Kaitiakitanga enables us to explore Maori perspectives and insights about health and information and calls for stronger inclusion of Maori in decisions. It acts as a guideline to address ongoing and complex issues such as collective ownership, the responsible publication of data and whether benefits in health for Maori can be explicitly declared and met.

INTRODUCTION

This chapter discusses the potential of indigenous concepts of governance within a contemporary health informatics setting. Health informatics is an evolving socio-technical and scientific discipline. It deals with the collection, storage, retrieval, communication and optimal use of health data, information and knowledge (hereafter referred to simply as "information").[1] The discipline attempts to assure quality healthcare for the community it serves.

Governance can be situated alongside health informatics, especially when considering the ethics, values and quality issues that impact on the care of people. Central to governance is decision-making and the process through which a group with delegated decision-making authority will direct their collective efforts. Governance involves multiple stakeholders to whom decision-makers are accountable: Governance in New Zealand's

(Aotearoa's) health sector deals with relationships between the Crown (Ministries, Government agencies, delegated authorities, etc.), communities, Maori and individuals. Governance also applies to guardianship over information in the best interests of stakeholders, and involves achieving both "desired results and achieving them in the right way" (IOG). Since the *right way* is largely shaped by the cultural norms and values of the stakeholders, there can be no universal template for good governance. Information technology is significantly redefining governance by providing enhanced opportunities to collaborate with and influence policy makers. Similarly, information technology changes accountabilities by opening new possibilities for the dissemination of information about the performance of government, district health boards and providers.

Indigenous contributions to governance in health informatics can be drawn from cultural concepts such as Kaitiakitanga, which implies guardianship, stewardship, governance and responsibility roles. This chapter explores Kaitiakitanga, its potential implementation in the Aotearoa health sector and its contributions to our thinking. After decades of unsuccessful attempts to positively shift the status of health for Maori, the indigenous peoples of Aotearoa, we must ask whether more control by Maori over information about Maori will make a difference. Kaitiakitanga can introduce a stronger position on ethics, values and quality when managing health data and optimize the benefits for both Maori and non-Maori alike.

Further, the inclusion of Kaitiakitanga concepts in governance structures, processes and roles in New Zealand's health sector in recent years points to a growing acceptance of indigenous input. For example, the establishment of the Cervical Screening's Kaitiaki Group as a legislated body (MOH, 2002a), and the Northern Region Hepatitis Consortium's Treaty Relationship Company model (C. Bullen, personal communication, 2003) are examples where indigenous ideas are perceived by their supporters to add value to governance.

The chapter specifically addresses collective ownership, collective privacy, responsible publishing and benefit. Kaitiakitanga implies that Maori will participate in *decisions* about health informatics and information technology, and influence policies and laws that support concepts of traditional protection, ownership and benefit that go beyond current laws and policies.

A MAORI PERSPECTIVE ON INFORMATION TECHNOLOGY

Definitions of information technology need not be limited to those found in academic or information technology industry journals. Potentially, any means of storing, analyzing and disseminating information can be included — *even our minds* (Kamira, 2002, p. 4). Maori concepts such as *Matauranga*, or intelligence, and *hinengaro*, or the mind, offer broader definitions and enhance what is generally understood about information technology. Matauranga refers to education and intuitive intelligence, and is linked to the divine. Hinengaro is the mind, the thinking, knowing, perceiving, remembering, recognizing, feeling, abstracting, generalizing, sensing, responding and reacting (Pere, 1991, p. 32). They are both vessels for knowledge.

Indeed, the broader Maori perspectives, such as those above, inform us why concepts of information technology as the information technology *industry* sees it are not only within the reach of Maori, but are also too simplistic since they do not include wider concepts of knowledge and understanding (Kamira, 2002, p. 5). It also explains why information technology is of great importance to Maori since the ancestor Tane-nui-a-rangi retrieved the *baskets of knowledge* from a celestial abode while coping with many dangers along the way (Barlow, 1991, p. 156), and that the dissemination of knowledge is thus a matter of great ritual and responsibility.

COLONIZATION, TIRITI (TREATY OF WATANGI) AND RELEVANCE FOR HEALTH INFORMATICS

It was the Maori version of the Treaty of Waitangi, "Te Tiriti O Waitangi," that was signed by chiefs at Waitangi on February 6, 1840 and subsequently in other locations. It is New Zealand's founding document. While debate is ongoing regarding its status and implementation, the legal status of the Treaty is not enforceable as there is an absence of statutory incorporation (TPK, 2001b, p. 16).[2] Therefore, governance models that are based on the Treaty are dependent on the application of moral obligations of the Crown and the recognition of principles.

The Hunn report (1961) first highlighted the failings of the Crown to meet its Treaty obligations to Maori. The government has attempted unsuccessfully since then to rectify the situation by focusing on socio-economic improvements for Maori. Yet, Maori continue to feature disproportionately in almost all of the negative statistics including unemployment, education, health, housing, domestic abuse and crime (Hunn, 1961; Williamson, 2001, p. 1; TPK, 2003). Particularly, the poor state of Maori health is well documented (MOH, 2002b, p. 2) and has sustained its negative status over many decades. Colonization clearly undermined the economic and social base of Maori society and resulted in mass dislocation and loss of land, language and spirituality: This has led to urgent calls for reclamation and protection of Maori "assets," including land, natural resources, language, belief systems, processes, etc. Active reclamation and protection of those assets are facilitated by movements such as Te Kohanga Reo (language nests), the Waitangi Tribunal and Maori authorities, etc. This call also applies to health and wider concepts of well being. As a result, Maori health provider groups aspire to become key providers for Maori and often coin the phrase "by Maori, for Maori" in the belief that other health services are not focused or responsive to the needs of Maori (Kamira, 1999-2000, p. 2).

Health informatics captures the need for technological capacity amongst all health providers, but access to that capacity for Maori providers is yet to be addressed fully. Anecdotal evidence suggests that the skills and equipment required for Maori providers to fully participate in health informatics is lacking (Kamira, 1999-2000, p. 14). Recent reports indicate that Maori are not participating in information technologies to the extent of other New Zealanders (TPK, 2001a; Infometrics, 2001).

However, there is some evidence that Maori are beginning to enter the information technology industry (Infometrics, 2001, TPK, 2001a), and that they are forming professional and interest groups.[3] These shifts towards "mastery" of information technology will eventually enable Maori to make well-informed decisions in health informatics. However, while it is an aspiration to build capacity in information technology in the broadest sense (Korowai Groups, 2002), this is some time away.

This leads us to the following questions:

1. Does the health sector enable Maori to make decisions about their health, or does it prefer to "look after" Maori?
2. What technological capacity would need to be available to Maori providers to assist aspirations to provide effective health care to Maori?
3. In the absence of technological capacity, what ways can Maori influence decisions about technology and health for their benefit?

WHAT IS KAITIAKITANGA?

Data — anonymous or not — has enormous spiritual and cultural significance for Maori, so it may require more attention and protection than generally given (MOH, 2001, p. 3). One way to provide this is to exercise the customary practices of Kaitiakitanga.

Kaitiakitanga (and the person or group who performs the Kaitiakitanga role – Kaitiaki), implies guardianship, protection, care and vigilance of data about Maori that is collected, stored and accessed. It introduces the idea of an inter-generational responsibility and obligation to protect, and enables the use of mechanisms such as *tapu*, the setting apart or restriction of knowledge or things, and *rahui*, the necessity to conserve, protect or restrict (Kamira, 2002, p. 22).

As governance decisions impact Maori and since the "right way" to govern is largely shaped by cultural norms and values of stakeholders, it is critical to extend current governance ideas if Maori are to successfully implement their Kaitiakitanga responsibilities. However, indigenous governance issues are complicated. For example (adapted from IOG):

1. What form(s) of indigenous governance are appropriate to the 21st century and suitable to the needs of indigenous peoples?
2. What is the appropriate balance between contemporary and traditional forms of governance for indigenous peoples?
3. What are appropriate strategies for creating indigenous capacity to successfully manage their governance responsibilities today?

It is important that members of governance or kaitiaki groups have an understanding of the historical, cultural and social complexities in which Kaitiakitanga perspectives are grounded.

Kaitiakitanga and Health Informatics Issues

The advantages of obtaining statistical information about health for Maori are clearly to profile groups — that will assist in developing effective policies and assist Maori to better manage their own health (MOH, 2002b, p. 23). However, decades of negative statistics indicate that gathering this information does not measurably achieve health gains for Maori.

1. Is data and information being gathered without producing *knowledge* that will generate benefits?
2. How would concepts of Kaitiakitanga in health informatics make any positive difference?

The following sub-headings provide some insights.

Data and Statistics

The publication of negative statistics over many decades undermines Maori and their health and produces few benefits. Negative statistics invoke the concept of *Takahia*, the act of trampling, often used to describe being belittled. Government databases collect abundant data about Maori: The data is analyzed and published, and Maori are profiled through statistical findings that continue to reinforce the most negative stereotypes. Currently, health informatics and statistics that are generated from databases symbolize disadvantage for Maori, who are busily curbing continuous socio-economic decline. The Maori experience of information technology that is in the control of others is the repeated reinforced perception of failure (Kamira, 2002, p. 22).

This is not to say that ethnicity data or data which identify particular Maori groups should not be collected. However, the accuracy through to the eventual publication of such data should be, ideally, constructive. Kaitiakitanga through the idea of *mana* — power or influence — can refocus the way that statistics are generated and published and demand a more productive and benefit-focused model.

Intellectual Property, Collective Ownership and Privacy

Kaitiakitanga introduces the idea of *tiaki*. Tiaki is to look after and guard and is a responsibility or an obligation rather than a right due to ownership. It enables a less exploitive relationship to exist where data about Maori are for the purpose of improvement and benefit first and foremost. Maori see the issues of intellectual property as a subset of these broader rights of ownership and include concepts of collective ownership. Collective ownership in health informatics can imply that grouped data about a collective such as a *hapu*, or an extended family tribal group, are owned by that collective. In turn, this implies their rights to make decisions about that data and benefit from that data.

In contrast, Western law defines intellectual property as outcomes of ideas or processes that are the result of human intervention — that is, knowledge created from the mind (Mead, 1997). Intellectual property laws both here and internationally tend to focus on commercial *ownership* and are inadequate as a way to protect indigenous collective knowledge. Until the fundamental ownership issues raised by Maori under the Treaty of Waitangi are mirrored in legislation, the best that can be achieved is interim recognition of Maori values and rights to participate in decision-making within the limits of the existing system (Putahi Associates, 1999).

Similarly, the Maori concept of privacy can encompass both the individual and the collective. An individual can have their privacy protected via the Privacy Act 1993. However, a *hapu*, or extended family tribal group, may feel they have a right to collective privacy that is not currently supported by legislation. Collective privacy is a means to protect data and information about groups of people rather than individuals, and is a key issue when data collection occurs on identifiable groups that wish to manage or control data about themselves.

First Beneficiaries

Kaitiakitanga introduces the idea of *awhina* — to assist or benefit. The ability to give what is truly needed without an expectation of reward means a clearer focus on more beneficial activities and responsible allocation of resources. The premise is that if Maori provide data then they should benefit from that data.

Some situations may prevent Maori from becoming the first beneficiaries. For example, limited participation by Maori providers in health informatics due to insufficient computer equipment or a lack of technological skills may be a barrier. Some Maori providers may need to rely on others to provide information, or they may only have access to information that has been collected for other purposes and does not focus on Maori, nor contain the detail required to initiate effective action (Kamira, 1999-2000, p. 14).

Mechanisms to promote Maori as first beneficiaries may include:

1. Protocols around grouped data with the input of Maori stakeholders that *require* benefits to be identified as criteria for collection and publication.
2. The development of initiative(s) when grouped data about Maori are identified to develop standards on the collection and use of data, and standards regarding real benefits that are more actively promoted and delivered.

THE STRATEGY

Mastery

While there is little *evidence* that information technology has positive and long-term impacts on the socio-economic status of people, the perception throughout the world is that information technology is a key driver for improving the world's socio-economic conditions (Riley, 1999), and that to ignore it will perpetuate or lead to even further disparities amongst the world's poorest and richest nations (UN, 2002). Maori will not be in a position to find out whether this is a *truth* for them unless they move from a passive role to mastery (Kamira, 2002, p. 17).

Apirana Ngata, a scholar and the first Maori university graduate in 1894, wrote in his grand-daughter's autograph book (Huta, 2001):

E tipu e rea mo nga ra o tou ao
To ringa ki nga rakau a te pakeha
Hei oranga mo to tinana
To ngakau ki nga taonga a o tipuna Maori
Hei tikitiki mo tou mahunga
To Wairua ki te Atua
Nana nei nga mea katoa

Grow up o tender youth in days of your life
Your hands grasp hold of the tools of the Pakeha
For your material well being
Your heart to the treasures of your Maori ancestors
As a plume for your head
Your spirit to God
The creator of all things

This well known passage captures the desire and the ability of Maori to acquire the knowledge of other cultures (of the Pakeha, or non-Maori) and is an important strategy for operating in a contemporary world, and for the uptake of information technology that would potentially improve health for Maori.

Using the Treaty / Tiriti Productively

The Ministry of Health's WAVE report (MOH, 2001, p. 3) states that the Treaty of Waitangi established a Crown obligation for Maori to enjoy a health status at least as good as that enjoyed by non-Maori. Further, it states that the government is committed to fulfilling its obligations to Maori to support self-determination for whanau (extended families) and Maori organizations. Kaitiakitanga can assist the Crown to achieve this obligation.

While Te Tiriti O Waitangi was signed and subsequently breached, attempts to bring it into a contemporary context have seen some progress. Concepts such as equity, partnership, collective ownership and protection (Treaty Articles II and III) can be expressed in relation to governance. Specifically, Maori are partners in the treaty and as such:

1. Article II guarantees Maori control and enjoyment of their valued possessions — tangible and intangible. This includes their health.
2. Article III affords Maori the attainment of equal human and social rights and privileges. The treaty implies that the *right* to good health can be exercised as per Article III regarding equity.

Validating and Promoting Maori Concepts

Health informatic projects are an opportunity to discover what gains can be made if Maori are actively involved in decisions through Kaitiakitanga. Aspirations of *tino rangatiratanga* — the ability to make decisions and control one's direction — are supported by Plumptre and Graham (1999), who conclude that three factors determine why some Native American tribes develop and some do not. They are:

1. Having the power to make decisions about their own future;
2. Exercising that power through effective institutions; and
3. Choosing the appropriate economic policies and projects.

Identifying the possible contributions from Maori through Kaitiakitanga and then incorporating them into health informatics will help to promote, protect and validate the assets (knowledge and skills) that Maori have and, importantly, may change the current negative health status that has become the norm.

HOW KAITIAKITANGA WOULD WORK TODAY

Treaty as a Framework

The Treaty of Waitangi can be implemented in current day activities as a localized and living model for Maori to apply tino rangatiratanga (decision making and control). It can act as a framework by which the appropriateness of decisions made during the development and implementation of health informatics projects can be assessed against Articles II and III of the Treaty.

Recognizing Existing Structures

Existing Maori structures also perform governance and Kaitiakitanga roles through Runanga (Maori authorities, trusts or similar organizations). These are legitimate structures often established by mandate and should not be undermined, but instead utilized where appropriate for health informatics projects. Maori governance roles will usually extend to a holistic range of areas along a continuum of well-being that extends far beyond the narrow constructs of health and all of its sub-categories.

Kaitiaki Groups

Kaitiaki groups would have overview roles to look after data, information and knowledge sourced from, or about, Maori. These groups would set ethical, value and quality guidelines. Ideally, participation of a Kaitiaki group would begin at the initiation stage of an IT project through to implementation and post-implementation.

The following points regarding establishing such a group are useful:

1. The membership and appointment of Kaitiaki groups is significant and Maori stakeholders would determine the appropriate type of Maori representation.
2. The position of Kaitiaki groups as a part of, or aside from, broader governance groups is significant as accountabilities and relationships are determined.
3. Kaitiaki group members should have decision-making powers and be informed on all issues, not just perceived Maori ones. This ensures Maori are defining what is of interest and prevents inappropriate filtering of information.
4. The issue of accountability is linked to risk. Kaitiaki groups would take into account the different levels of risk, the shared-risk and the authority it would accept or have delegated to it.

Kaitiaki Members

Kaitiaki members would represent stakeholder groups that would in turn expect them to pursue aspects of tino rangatiratanga (decision making and control). Kaitiaki members would be responsible for, and would ensure a focus on:

1. The treaty;
2. Safety and protection of Maori individuals and collectives; and
3. Benefit to Maori individuals and collectives.

CONCLUSION

Health informatics will need to include concepts of governance and Kaitiakitanga alongside its medical, technology and social discourses. Modern concepts of Kaitiakitanga can enable:

1. Maori to be explicitly identified as first beneficiaries of health informatics where relevant;
2. Recognition of collective ownership, use, access, analysis and interpretation of data; and
3. Recognition of collective privacy as a valid form of control for grouped data. (Kamira, 2002, p. 23)

Kaitiakitanga contributes useful indigenous ideas that add value to governance and potentially may result in a health gain for Maori in the long term as they make decisions about information that will be expected to return benefits. Maori can, through Kaitiakitanga, actively and effectively influence policies and laws that support the protection and ownership of data and information towards improving health benefits.

REFERENCES

Barlow, C. (1991). *Tikanga whakaaro: Key concepts in Maori culture*. Oxford, New York and Melbourne: Oxford University Press.

Bellinger, G., Castro, D., & Mills, A. (n.d.). Data, information, knowledge, and wisdom. Retrieved May 30, 2004, from http://www.outsights.com/systems/dikw/dikw.htm

Duffy, A. P., Barrett, D. K., & Duggan, M. A. (2001). *Report of the ministerial inquiry into the under-reporting of cervical smear abnormalities in the Gisborne region*. Retrieved September 14, 2005, from http://www.csi.org.nz/report/download.htm

Hunn, J. K. (1961). *Report on the Department of Maori Affairs*. Wellington, NZ: Government Printer.

Huta, J. (2001). A Maori perspective. *CTC Bulletin*, *17*(2), April. Retrieved December 28, 2004, from http://www.daga.org/cca/ctc/ctc01-04/ctc0104h.htm

Infometrics. (2001). *The digital divide and Maori*. Retrieved December 28, 2004 from, www.nzmis.org.nz

IOG (Institute on Governance). (n.d.) Retrieved November 3, 2005, from http://www.iog.ca

Kamira, R. (1999-2000). (Combined Report). *The kidZnet project: Guiding cultural and ethical issues for Iwi*, 1999; *Report for Iwi cultural and ethical issues in kidZnet,* 2000. Retrieved December 28, 2004, from www.pauainterface.com

Kamira, R. (2002). *Te Mata o te Tai — The edge of the tide: Rising capacity in information technology of Maori in Aotearoa*. Paper presented at the *Fourth International Information Technology in Regional Areas Conference*, Rockhampton, Australia. Retrieved May 28, 2004, from www.pauainterface.com

Korowai Groups (The). (2002). *Maori information and communication technologies strategic plan*. Aotearoa.

Mead, A. T. P. (1997). Cultural and intellectual property rights of indigenous peoples of the Pacific. In *Cultural and intellectual property rights: Economic, politics & colonisation* (vol. 2, pp. 22-30).

MOH (Ministry of Health). (2001). *The WAVE project, from strategy to reality*, Wellington, NZ.

Retrieved December, 28, 2004, from http://www.moh.govt.nz

MOH (Ministry of Health). (2002a). *Review of the health (cervical screening (Kaitiaki)) regulations 1995*, Retrieved March 25, 2004, from http://www.moh.govt.nz

MOH (Ministry of Health). (2002b). *He Korowai Oranga: Maori health strategy*. Wellington, NZ: Ministry of Health.

Orange, C. (1987). *The Treaty of Waitangi*, Wellington, NZ: Allen & Unwin.

Pere, R. T. (1991). *Te Wheke: A celebration of infinite wisdom*. Gisborne, NZ: Ao Ako Global Learning New Zealand Limited.

Plumptre, T., & Graham, J. (1999). *Governance and good governance: International and aboriginal perspectives*. Canada: Institute on Governance. Retrieved December 25, 2004, from http://www.iog.ca

Putahi Associates Ltd. (1999). *Maori and the patenting of life form inventions: An information paper produced by the patenting of life forms focus group for the Ministry of Commerce*, Wellington, NZ: Patenting of Life Forms Focus Group.

Riley, L. A., Nassersharif, B. N., & Mullen, J. (1999). *Assessment of technology infrastructure in native communities*. Washington, DC: Report for Economic Development Administration, U.S. Department of Commerce.

TPK (Te Puni Kokiri). (2001a). *Maori access to information technology*. Wellington, NZ.

TPK (Te Puni Kokiri). (2001b). *He Tirohanga o Kawa ki Te Tiriti O Waitangi: A Guide to the principles of the Treaty of Waitangi as expressed by the Courts and the Waitangi Tribunal*. Wellington, NZ.

TPK (Te Puni Kokiri). (2003). *Maori in New Zealand*. Wellington, NZ: Ministry of Maori Development. Retrieved December 28, 2004, from http://www.tpk.govt.nz/maori/default.asp

UN (United Nations). (2002). CSTD Panel on *Indicators of technology development*. Room XXVI, Palais des Nations, Geneva, Switzerland. Retrieved June 7, 2004 from, http://www.unctad.org/stdev/un/p1-02.html

Williamson, A. (2001). Emancipatory learning via the Internet: A model for reducing Maori socio-economic exclusion in Aotearoa/New Zealand. In W. Kim, T. W. Ling, Y. J. Lee, & S. S. Park (Eds.), *The human society and the Internet: Internet related socio-economic issues*. Berlin, Germany: Springer.

ENDNOTES

[1] While there are a multitude of definitions for data, information and knowledge, this is a favourite of the author. Data are raw and have little significance beyond itself. Information is data that have been given meaning by way of relational connection; it provides answers to "who," "what," "where" and "when" questions. Knowledge is the appropriate collection of information, such that its intent is to be useful; it answers "how" questions (Bellinger, Castro, & Mills, n.d.).

[2] A political analysis of the Treaty is not provided, as there are many appropriate and more in-depth sources of information available such as TPK (2001b) and Orange (1987).

[3] E.g., Te Waka Wahine Wa-hangarau: Society for professional Maori women in information technology. Note, there are at least three Maori IT groups in the country.

This work was previously published in Information Technology and Indigenous People, edited by L. E. Dyson, pp. 30-51, copyright 2007 by Information Science Publishing (an imprint of IGI Global).

Chapter 4.19
Effects of Knowledge Management Implementations in Hospitals:
An Exploratory Study in Taiwan

Wen-Jang (Kenny) Jih
Middle Tennessee State University, USA

Cheng-Hsui Chen
National Yunlin University of Science & Technology, Taiwan

Ying-Hsiou Chen
National Cheng-Kung University, Taiwan

ABSTRACT

From the knowledge management point of view, the fundamental mission of hospital management is the delivery of quality medical services, utilizing highly specialized knowledge to solve healthcare problems within various resource constraints. Similar to other knowledge-intensive industries operating in highly challenging business environments, hospitals of all sizes must view the creation, organization, distribution, and application of knowledge as a critical aspect of their management activities. Knowledge management represents a viable strategy as hospitals strive simultaneously to provide quality medical services, improve operational efficiency, and conform to the government's documentation and reporting regulations. This study examines the correlation as well as the causal relationships among knowledge characteristics, knowledge acquisition strategy, implementation measures, and performance of knowledge management implementations in the context of hospital management. Using primary data collected in Taiwan, our analyses show that the characteristics of knowledge affect the ways in which knowledge management is implemented, and the implementation measures, in turn, have a significant impact on the results of knowledge management implementations.

INTRODUCTION

Hospitals of all sizes currently are faced with a multitude of management pressures, including industry competition, customer satisfaction, shortage of specialized personnel, compliance with government regulations, cost reduction, and the ever-increasing demand for more effective cures (Camilleri & O'Callaghan, 1998; Porter & Teisberg, 2004). In coping with these challenges, hospitals actively are experimenting with various management initiatives and programs, such as total quality management and knowledge management, with varying performance results. Emerging as a new multidisciplinary management field, knowledge management (KM) promises to enhance competitive advantage in the highly dynamic knowledge economy by treating valuable and scarce knowledge as a critical organizational asset and by managing it in a systematic manner (Sharkie, 2003; Ulrich & Smallwood, 2004). From the knowledge management point of view, many hospital services involve knowledge-intensive processes that are carried out to solve patient health-related problems. Because of the knowledge-intensive nature of healthcare services, much of a hospital's success depends on effective and efficient creation, organization, validation, dissemination, and application of its highly specialized medical knowledge.

Traditional knowledge management mechanisms in most hospitals typically include morning meetings, apprenticeships, internships, professional seminars, research partnerships with outside research institutions, and other forms of human interaction. Sophisticated information technologies also are being deployed in some hospitals in order to manage medical images and to capture scarce expertise (e.g., medical expert systems) (Davenport & Glaser, 2002). The addition of Internet technologies to the portfolio of information processing technologies offers a new set of powerful tools in order for hospitals to implement knowledge management programs.

In light of the strategic value of professional knowledge, hospitals increasingly recognize a need to manage more actively their intellectual capitals. The field of knowledge management provides the frameworks and techniques that are required to transform a hospital into a learning organization (Adams & Lamont, 2003; Awad & Ghaziri, 2004; Becera-Fernandes, Gonzales & Sabherwal, 2004; Gupta & Govindarajan, 2000; Hansen et al., 1999; Leonard-Barton, 1995). These frameworks and techniques emerge from the inquiries conducted and the experiences acquired in a variety of contexts, including manufacturing (Kim, Hwang & Suh, 2003), customer relationship management (Gebert, Geib, Kolbe & Brenner, 2003), consulting (Sarvary, 1999), retail chain (Tsai, Yu & Lee, 2005), and health care (Powers, 2004). Much of the literature, however, has been either case studies or conceptual discussions. Empirical studies based on the primary data collected in the field, however, are important for advancing the field of knowledge management toward maturity.

Motivated by the dearth of empirical inquiries in knowledge management that address issues in hospital management, we conduct this study to identify the relationship between some factors that play a significant role in successful knowledge management implementations in the healthcare environment. Our purpose is to understand how knowledge management is practiced and the result of implementation in this knowledge-intensive sector. We also seek to contribute to hospital management by offering empirical evidence for the value of knowledge management in coping with the multi-faceted management challenges faced by today's hospitals.

The remainder of the article is structured as follows. The next section describes an interview process conducted with hospital executives and medical doctors for the purpose of selecting the constructs for our research model. We review some of the existing knowledge management literature that relates to our research constructs. We also discuss our research hypotheses that are formulated as a result of field interviews and literature reviews. The section on research methodology describes the research framework, data collection approaches, and data analysis techniques. This is followed by the results of data analysis. The last section discusses some implications of the research findings for hospital knowledge management in particular and knowledge management in general. Suggestions for future research in knowledge management also are presented.

Identification of Research Constructs

Knowledge management increasingly has received attention as an important multidisciplinary field in both the academic and corporate arenas over the past years (Adams & Lamont, 2003; Becera-Fernandez et al., 2004; Davenport, De Long & Beers, 1998; Gloet & Berrell, 2003; Sharkie, 2003). A common denominator among knowledge management researchers is the belief that in the knowledge economy, which is characterized by rapid change and fierce competition, knowledge should be viewed as a valuable and manageable resource, just like tangible assets such as capital and factory facility (Davenport & Prusak, 2000; Leonard-Barton, 1995; Liebowitz & Wilcox, 1997; Soo, Devinney, Midgley & Deering, 2002). Various conceptual frameworks for effective implementation of knowledge management programs or projects have been proposed in order to facilitate knowledge sharing and to stimulate continuous innovation both within and across organizational boundaries (Awad & Ghaziri, 2004; Davenport & Glaser, 2002; Lee & Hong, 2002; Soo et al., 2002; Tiwana, 2000; Wang, 2002; Wiig, 1994). Due to the knowledge-intensive nature of healthcare delivery processes, there has been a call for implementation of knowledge management in the context of hospital management in the pursuit of sustainable competitive advantage (Davenport & Glaser, 2002; Powers, 2004; Van Beveren, 2003).

In order to better understand the unique features of hospital management as related to knowledge management, we conducted hour-long interviews with two hospital executives and three medical doctors in a regional medical center. A structured questionnaire was used in the interviews, supplemented by open-ended questions, to help us to identify the research constructs for our inquiry. Several points stood out to provide a direction for the study. First of all, medical services (esp. clinical care) are very context-sensitive. A specific clinical care problem may be attributed to the patient's personal condition, disease specifics, disease history, family background, treatment history, medication history, and so forth. Effective treatment collectively must consider a wide variety of factors. The complexity of clinical care knowledge, in other words, often comes from a great number of factors and the subtle relationships among these factors that need to be considered in the treatment decision. The ability to distinguish between the relevant and irrelevant factors and to identify the relationships among these factors is a core competence of a true expert. Also, although theoretical knowledge distributed by such published avenues as books and articles is valuable in developing a medical expertise, ultimately, it is the practical experience in dealing with specific cases that truly builds up a medical doctor's professional expertise and the hospital's knowledge repository. In addition, most medical knowledge accumulated over the years is practical knowledge that is not preserved in any printed media. These findings lead to the selection of knowledge characteristics as a research construct.

We also found from the field interviews that specific implementation measures, such as the availability of information infrastructure that facilitates KM activities, incentive programs, and other people-related factors, play a key role in the success (or failure) of the KM implementation program. Information technologies are used to digitize and to store knowledge content. Once kept in the systems, valuable knowledge content easily can be disseminated, integrated, and deployed. Each specialization unit in a hospital usually represents an organizational silo separated from other specialization units. A well-designed information system enables the sharing of related information and knowledge by *baking* the system capabilities in the daily business processes (Davenport & Glaser, 2002). Furthermore, the interviews identified two preferred indicators for assessing the performance of the KM implementation: improvement of internal process and overall organizational performance. Two research constructs—KM implementation measures and KM performance—are identified as a result. Finally, the research construct Knowledge Acquisition Strategy is included in the research model in order to evaluate the strategic role of knowledge source in hospital management. Based on the findings of our preliminary inquiry, the scope of this study is limited to four research constructs: the characteristics of medical knowledge, the strategy used to acquire valuable knowledge, how KM concepts are implemented, and the result of KM implementations. The following section reviews the literature relating to these four constructs.

Knowledge Characteristics

As a service-oriented and knowledge-intensive organization, a hospital typically deals with knowledge in a variety of categories: customer (patient) knowledge, service (treatment) knowledge, process knowledge, and account management, to name just a few. In generic terms, knowledge can be characterized in many ways: shallow vs. deep, procedural vs. declarative, explicit vs. tacit, domain-independent vs. domain-specific, common sense vs. professional, static vs. dynamic, and proprietary vs. nonproprietary, for example (Awad & Ghaziri, 2004; Davenport & Glaser, 2002; Gloet & Berrell, 2003; Gupta & Govindarajan, 2000; Howells, 1996; Nonaka & Takeuchi, 1995; Polanyi, 1996; Ulrich & Smallwood, 2004). A recent list of knowledge attributes proposed by Holsapple (2003) consists of 23 items. According to the healthcare practitioners that we interviewed, for the medical and healthcare service industry, four knowledge characteristics are deemed most relevant: knowledge mode (explicit vs. implicit), knowledge complexity, strength knowledge appropriability (ease of replication and transferring), and knowledge volatility (dynamic vs. static). These four characteristics were identified using a rating questions list, augmented by open-ended elaborative discussions.

Explicit knowledge is represented in the form of recorded products, such as printed documents, formulas, software, system manuals, and hardware equipment, while implicit knowledge primarily is undocumented and retained as memory (Howells, 1996; Nonaka & Takeuchi, 1995; Polanyi, 1996; Zack, 1999). Hospitals vary in their perceptions of and relative emphasis on the explicitness—implicitness continuum of their valuable healthcare knowledge. Different perceptions and preferences may lead to different strategies being adopted for knowledge acquisition and other important aspects of knowledge management implementation measures. Explicit knowledge is more amenable to technologically oriented solutions, such as a document base and a knowledge map, whereas implicit knowledge is handled primarily through social networks.

Complex knowledge is difficult for organizations to acquire. Once acquired, however, it can become a valuable source of competitive advantage (Holsapple, 2003; Teece, 1998). Knowledge complexity is determined by the abstract nature of knowledge, the number of knowledge components, and the interaction of these components (Soo et al, 2002). In the healthcare domain, both the great amount and the intricate interactive effect of professional medical knowledge pose a substantial challenge for hospitals that are striving to provide quality healthcare services.

Knowledge that is nonproprietary in nature easily can be transferred across organizational boundaries. The proprietary nature of knowledge usually is determined by the extent to which the knowledge is tightly tied to the specific organization (Teece, 2003; Soo et al., 2002). From the system development point of view, proprietary knowledge usually is domain-specific. Both domain-specific and domain-independent knowledge are important in solving complex problems. However, past successes of expert system technology applications demonstrate that domain-specific knowledge usually contributes more than domain-independent knowledge in solving difficult problems. This notion of organizational specificity associated with knowledge management is illustrated well by Gupta and Govindarajan (2000) in a case study of the Nucor Steel Corporation. The consistently superior performance of Nucor Steel demonstrates that, once knowledge creation and sharing are embedded in the management practice and the daily operational routines, the resultant proprietary knowledge can establish a solid foundation for a truly sustainable competitive advantage (Adams & Lamont, 2003). The value of proprietary knowledge is obvious in hospital management, especially in dealing with challenging healthcare problems.

Currency of knowledge can be an important issue at times when new knowledge renders old knowledge useless. In these cases, knowledge must be subject to frequent updating in order to stay valuable. The optimal updating frequency is determined by the dynamic (or static) nature of the knowledge. The issue of knowledge updating has been addressed in expert systems development (Liebowitz & Wilcox, 1997). In fact, one of the criteria in the selection of knowledge-based expert system application domains is that the domain knowledge must be relatively static. The complexity involved in knowledge updating and validation poses a significant challenge for keeping a knowledge base current and for having the system accessible at the same time. In the broader context of medical care service delivery, updating of the knowledge repository may affect the strategy for knowledge acquisition as well as knowledge management implementation measures.

Knowledge Strategy

A knowledge strategy, as defined by Zack (1999), "describes the overall approach an organization intends to take to align its knowledge resources and capabilities to the intellectual requirements of its strategy" (p. 135). Zack's (1999) knowledge strategy framework consists of two dimensions: exploitation vs. exploration and internal vs. external. While the exploitation vs. exploration dimension distinguishes a creator from a user of knowledge, the internal vs. external dimension describes the organization's primary sources of knowledge. Internal knowledge is characterized as being "resident within people's heads; embedded in behaviors, procedures, software and equipment; recorded in various documents; or stored in databases and online repositories" (Zack, 1999, p. 138). External sources of knowledge include publications, university alliances, government agencies, professional associations, personal relations, consultants, vendors, knowledge brokers, and interorganizational alliances. Using Nucor Steel's experience, Gupta and Govindarajan (2000) argued that internally created knowledge tended to contribute more to an organization's competitive advantage than did external approaches.

Another useful KM strategy framework is represented by a codification vs. personalization dichotomy. According to Hansen, et al. (1999), the codification-oriented KM strategy is suitable for explicit, recordable, formal, and replicable knowledge, and the personalization-oriented strategy works better for implicit knowledge.

Whereas information technology plays a central role within the codification strategy, it primarily provides tool support for the personalization strategy. The choice of primary knowledge acquisition strategy usually is determined by a variety of factors, among which is the characteristics of knowledge (Awad, 2004; Davenport & Prusak, 2000; Tiwana, 2000; Wiig, 1994).

Knowledge Management Implementation Measures

Knowledge management generally is viewed as a collection of management practices consisting of knowledge accumulation, knowledge organization, knowledge dissemination, and knowledge application (Awad & Ghaziri, 2004; Davenport & Prusak, 2000). The implementation of organizational KM projects typically involves technical as well as non-technical measures. A flexible and efficient information technology infrastructure is an essential requirement. The subsequent distribution and application of the organizational knowledge depends on digital representation, computerized storage, dissemination of the knowledge content, and the application context. Additionally, properly designed incentive programs must be in place in order to discourage knowledge hoarding and to promote knowledge sharing (Davenport & Prusak, 2000; Soo et al., 2002). Sufficient resources must be committed to encourage active learning and perpetual updating of the professional staff's knowledge base.

Knowledge Management Performance Measurement Issues

Measuring the results of KM projects is a challenging task. The subjective nature of the benefit measurement and the lengthy lead time required in order for the benefit to become quantitatively measurable usually are cited as the main sources of difficulty (Abeysekera, 2003; Stone & Warsono, 2003). Although the organizational benefit resulting from KM project implementation eventually must be expressible in financial terms, a significant portion of the benefit is qualitative and only can be measured subjectively. Most literature, therefore, suggests a portfolio approach to measuring the result of KM implementation, such as innovative capability, which would include both financial and non-financial data (Chourides, Longbottom & Murphy, 2002; Darroch, 2003; Soo et al., 2002). Furthermore, it is recognized increasingly that in order for KM to achieve its greatest success, KM functions must be integrated tightly with major business processes. For example, Davenport and Glaser (2002) indicate that, based on their experience at a major medical center in Boston, medication knowledge must be embedded into a doctor's prescription writing process in order to make knowledge application a natural part of daily work practice. The tight integration of knowledge management process and other business processes contributes at least partially to a straightforward measurement of the result of KM implementation (Darroch, 2003).

The input from the practitioners also suggests the following relationships to be examined in the study:

1. The primary knowledge acquisition strategy is affected significantly by knowledge characteristics (knowledge characteristics $\rightarrow$ knowledge acquisition strategy).

   Rationale: Knowledge either is created from within the organization or collected from external sources. Each knowledge acquisition strategy has its pros and cons, depending on the nature of business, the available resource, and other factors. Knowledge that is created internally tends to fit the organizational needs better and to and contributes more to the competitiveness, for example. Although both internal and external sources can be used, an organization usually has a primary knowledge acquisition strategy regarding how to build up its knowledge stock. Everything else being equal, the perceived knowledge characteristics may determine the primary knowledge acquisition strategy that is adopted by the organization.

2. Knowledge management implementation measures are affected significantly by knowledge characteristics (knowledge characteristics → KM implementation measures).

   Rationale: When valuable knowledge is perceived to be primarily tacit, it is less likely that a significant amount of resource will be invested in building sophisticated information systems to support knowledge management activities. Rather, most of the KM implementation measures will be people-oriented. Further, in an effort to cope with knowledge complexity and volatility and to encourage the development of proprietary knowledge, an organization may pay more attention to its incentive program, expertise development, and human resource planning. Therefore, it is reasonable to assume that the way a hospital chooses to implement KM programs is influenced significantly by its perception of knowledge characteristics.

3. Knowledge implementation measures are affected significantly by the primary knowledge acquisition strategy (primary knowledge acquisition strategy → KM implementation measures).

   Rationale: Creating knowledge internally often requires knowledge workers to share knowledge with the rest of the organization. Knowledge sharing usually is not a natural component of organizational culture in most Taiwanese hospitals, where high work pressure and departmental silos tend to encourage knowledge hoarding. In order to encourage internal creation of knowledge through knowledge sharing, hospitals must implement a portfolio of facilitating measures, such as incentive programs, expertise development, human resource planning, and technological infrastructure. Therefore, we postulate that KM implementation measures are affected by the primary knowledge acquisition strategy.

4. KM performance is affected significantly by the primary knowledge acquisition strategy (primary knowledge acquisition strategy → KM implementation performance).

5. KM performance is affected significantly by KM implementation measures (KM implementation measures → KM performance).

   Rationale: O'Dell, Elliott, and Hubert (2003) identified three general categories of value propositions for KM program implementation—customer intimacy, product-to-market excellence, and operational excellence—as a result of APQC's first benchmarking study on best practices in KM. Operational excellence is the most relevant to the context of hospital management. The value proposition defines the goal for KM implementation. Since the degree to which the goal is achieved can be affected significantly by the approach (strategy) and the specific actions (implementation measures) taken to pursue the goal, we set up these two hypotheses to validate our presumptions.

RESEARCH DESIGN

Research Framework, Variables, and Hypotheses

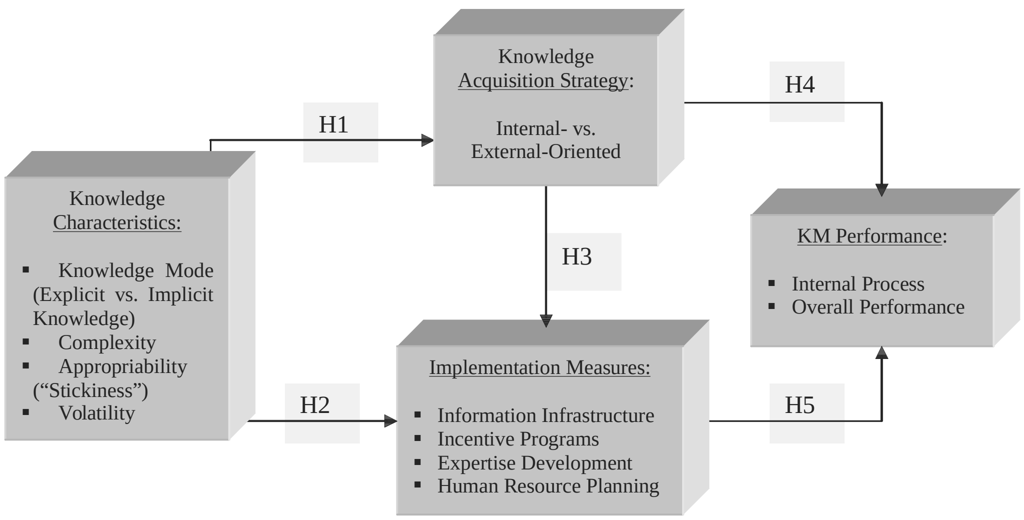
For the purpose of this research, KM was defined as a management function responsible for the acquisition, organization, evaluation, sharing, distribution, and application of both tacit and explicit knowledge for an organization. In order to gather data to analyze how KM was practiced in Taiwanese hospitals, we postulated that the hospital management professionals' perceptions of knowledge characteristics would affect the primary strategy adopted to acquire knowledge and also would affect the hospitals' KM implementation measures. We further assumed that these latter two sets of variables, in turn, would affect the result of KM implementation. These hypothesized relationships between the research variables were derived from the result of literature and our field interviews with healthcare professionals in Taiwanese hospitals. The diagrammatic depiction of our research model is shown in Figure 1. The five research hypotheses established to validate our presumption also are indicated in the diagram.

Operational Definition of Research Variables

The characteristics of knowledge were represented by four important dimensions of medical and hospital management knowledge that we identified in our preliminary interview with medical professionals: knowledge mode (explicit vs. tacit nature of knowledge), complexity, strength of knowledge appropriability, and knowledge volatility.

The primary strategy adopted to acquire knowledge in this study was dichotomized into internally oriented acquisition and externally oriented acquisition. Internal sources included apprenticeships, intranet, internal documentation, morning meetings, internal medical databases, and department meetings. External sources included collaboration with universities, external consultants, internships in other hospitals, seminars, and professional conferences.

Figure 1. Research framework



The knowledge management implementation measures were represented by activities in four areas: information infrastructure, incentive programs for knowledge sharing, expertise development, and human resource planning. The first three items corresponded to three basic entities involved in knowledge management programs: people (incentive programs), knowledge (expertise development), and technological tools (information infrastructure). The fourth item (human resource planning) corresponded to both people and knowledge.

Two categories of variables were used to represent the performance measure of knowledge management implementation in hospitals: internal process improvement and overall organizational performance enhancement. Internal process improvements consisted of communications and efficiency improvement measures, such as problem-solving time, employee participation, decision-making cycle time, and employee interaction. Overall organizational performance measures covered such items as service quality, customer focus, absenteeism, and customer (patient) satisfaction. With these definitions of the research variables, the research hypotheses were formulated as follows:

Hypothesis 1 (H1): The primary knowledge acquisition strategy is affected significantly by knowledge characteristics.

Hypothesis 2 (H2): KM implementation measures are affected significantly by knowledge characteristics.

Hypothesis 3 (H3): KM implementation measures are affected significantly by the primary strategy adopted for knowledge acquisition.

Hypothesis 4 (H4): KM implementation performance is affected significantly by the primary knowledge acquisition strategy.

Hypothesis 5 (H5): KM implementation measures affected significantly KM performance.

Questionnaire Design and Data Collection

All research variables except knowledge acquisition strategy were represented by multiple questions using five-point Likert scales, with 1 indicating very poor or highly disagree and 5 indicating very good or highly agree. A checklist was devised for knowledge acquisition strategies. A questionnaire was sent to 20 head physicians in four medical centers. This first version of the questionnaire was tested using results from the 12 respondents. Based on feedback from the pilot test, the questionnaire was revised by removing the questions with low reliability coefficients and by modifying the ones lacking semantic clarity. The revised version consisted of 44 Likert scale questions. There were 15 questions for knowledge characteristics, 18 questions for KM implementation measures, and 11 questions for performance measures. A copy of the questionnaire was mailed to the president of each of the 126 hospitals on the list that was compiled by the Department of Health of Taiwan. Hospital presidents were targeted on the assumption that they were familiar with the issues under study.

Due to the exploratory and empirical nature of the study, the questionnaire was limited in the criterion validity and the construct validity. Both the content validity and the discriminant validity were assumed to be proper, since the questions were based on the literature and the input from the practicing professionals. The reliability measures, as represented by Cronbach, if all research variables are above 0.70, an indication of acceptable reliability (Nunnally, 1978). Table 1 summarizes the reliability measures of all research variables. Note that two questions were removed from the original set of 46 questions as a result of this analysis.

Table 1. Reliability of research variables

Variables	Sub-dimensions	Question Items	Cronbach α
Knowledge Characteristics	Explicitness	1, 2, 3, 4	0.8239
	Complexity	5, 6, 7, 8	0.7605
	Appropriability	9, 10, 11, 12*	0.7187
	Volatility	13, 14, 15, 16	0.7905
Implementation Measures	Information Infrastructure	17, 18, 19, 20, 21	0.8827
	Incentive Program	22, 23, 24, 25	0.8191
	Expertise Development	26, 27, 28, 29, 30*, 31	0.8063
	Human Resource Planning	32, 33, 34, 35	0.7674
Performance Measures	Internal Process	36, 37, 38, 39, 40, 41	0.9115
	Overall Preformance	42, 43, 44, 45, 46	0.8374

*: *Questions are subsequently removed due to insufficient reliability measures.*

Data Analysis Method

In order to test the previous hypotheses, several statistical analysis techniques were employed to analyze the data. In particular, t-test was conducted to evaluate the effect of knowledge characteristics on the primary acquisition strategy and the effect of the knowledge acquisition strategy on the KM implementation measures. Canonical correlation analysis was conducted to determine the correlation relationship between knowledge characteristics and implementation measures, both of which are multiple variables. Finally, stepwise regression analysis was used to determine the impact of the KM implementation measures on both of the performance measures.

RESEARCH FINDINGS

The questionnaire was mailed to 126 hospitals in Taiwan, from which 50 questionnaires were returned. The questionnaire was addressed to the person who was most familiar with knowledge management practices in the hospital. Three questionnaires were removed due to incomplete answers. The remaining 47 accounted for an effective response rate of 37.3%. Ten questionnaires were filled out by hospital presidents, seven by vice presidents, one by a physician, and 29 by hospital management staff. The Kolmogorov-Smirnov statistics for knowledge characteristics, implementation measures, and performance measures all exhibited normal distribution at the 0.001 significance level.

Hypothesis 1 stated that the primary knowledge acquisition strategy is affected significantly by knowledge characteristics. The frequency distribution of knowledge acquisition strategies summarized in Table 2 showed that the most common internal source of knowledge was morning meetings, and the most widely used external knowledge source was outside meetings. Table 3 showed the means and standard deviations of each of the four dimensions of knowledge characteristics. The measures of knowledge characteristics were dichotomized, using averages as the thresholds, into two distinct levels: high and low. A t-test was performed to determine the correlation relationship between knowledge acquisition strategy and knowledge characteristics. Table 4 shows that the adoption of internal-oriented

Table 2. Distribution of primary knowledge acquisition strategy

Types of KA Strategy		Freq.	%	Freq. Total	%
Internal-Oriented	Manual and Instruction Management	1	2.1	20	42.6
	Morning Meeting	12	25.5		
	Group Discussion within Department	6	12.8		
	Hospital-wide Medical Database	1	2.1		
External-Oriented	Collaboration with University	3	6.4	27	57.4
	Consultant	2	4.3		
	Internship with other Hospitals	2	4.3		
	Outside Meeting	20	42.6		

Table 3. Means (μ) and standard deviations (σ) of knowledge characteristics

Knowledge Characteristics	ì	ó
Knowledge Mode	3.7872	0.5874
Complexity	2.9043	0.4056
Appropriability	3.4752	0.5462
Volatility	3.0691	0.6567

Table 4. Relationship between knowledge characteristics and KA strategy

Knowledge Characteristics	Knowledge Acquisition Strategy			
	Internal-Oriented	External-Oriented	t-value	p-value
Explicitness	3.8241	3.4825	1.152	0.012*
Complexity	3.2167	2.7621	1.224	0.000**
Appropriability	3.5667	3.4074	0.998	0.328
Volatility	3.1375	3.0185	0.610	0.545

strategy was affected significantly by the level of knowledge explicitness (t = 1.152, p-value = 0.012), and the use of external-oriented knowledge was affected significantly by knowledge complexity (t = 1.224, p-value = 0.000). In other words, the internal-oriented strategy was used more to acquire knowledge with high explicit levels, whereas the external-oriented strategy was used more for knowledge with high complexity levels. These were indications that, in general, hospitals in Taiwan tended to rely on external sources to update and upgrade their knowledge bases. The lack of resources and the competitive pressure from the environment were keeping most of them from actively investing in internal research and development in order to create valuable knowledge from within the organization. With a few exceptions (e.g., Chang-Hua Christian Hospital), only the medical centers that are affiliated with research universities had the capabilities and resources with which to carry out knowledge creation activities in a systematic manner.

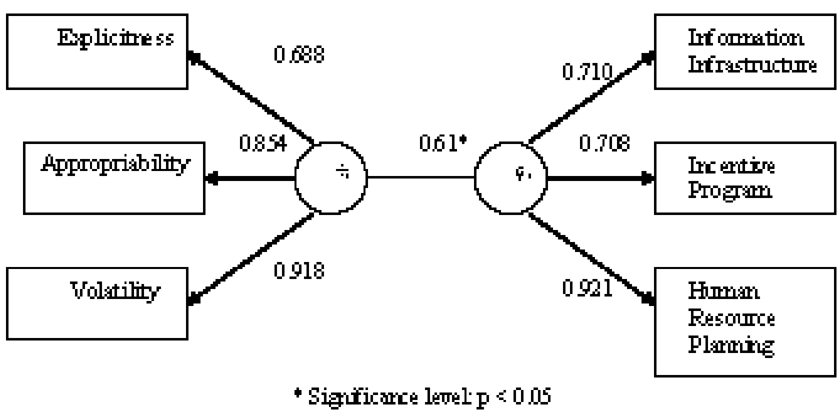
Hypothesis 2 states that KM implementation measures are affected significantly by knowledge characteristics. Since both knowledge characteristics and implementation measures consist of four variables, canonical correlation analysis was employed in order to determine the correlation relationship. As summarized in Table 5 and as graphically presented in Figure 2, the result

Table 5. Canonical correlation analysis between knowledge characteristics and implementation measures

Knowledge Characteristics	Canonical Variate	Implementation Measures	Canonical Variate
	÷		ç
Knowledge Explicitness	0.688*	Information Infrastructure	0.71*
Complexity	-0.005	Incentive Program	0.708*
Appropriability	0.854*	Expertise Development	-0.128
Volatility	0.918*	Human Resource Planning	0.921*
Extracted Variance	51.16%	Extracted Variance	46.76%
Wilks E	0.465	Redundancy	0.174
F-value	2.13	Canonical Correlation ($\hat{r}$)	0.61
p-value	0.011	$\hat{r}^2$	0.37

**:Absolute value of canonical loading > 0.6*

Figure 2. Path diagram of canonical correlation analysis between knowledge characteristics and implementation measures



of canonical analysis identified that the linear composition of three knowledge characteristics—knowledge explicitness, appropriability, and volatility—were correlated significantly with the linear composition of three implementation measures: information infrastructure, incentive program, and human resource planning. The correlation coefficient (ρ) was 0.61 at the significance level 0.05.

Hypothesis 3 states that KM implementation measures are affected significantly by the primary strategy adopted for knowledge acquisition. The result of the t-test indicated that two implementation measures were affected significantly by the knowledge acquisition strategies. Internal-oriented strategy adopters scored higher in information infrastructure than external-oriented strategy adopters, whereas external-oriented adopters scored higher in incentive measures.

Hypothesis 4 states that the performance of KM implementation is affected significantly by the primary knowledge acquisition strategy. The result of the t-test (Table 6) showed that different knowledge acquisition strategies did not cause significant differences in either dimension (internal process and overall organizational performance) of the performance of KM implementation. Hypothesis 4 was rejected. Since internally created knowledge usually is considered more likely to contribute to an organization's competitive advantage (Gupta & Govindarajan, 2000), this finding indicated that the hospitals in Taiwan still lacked a proper understanding of the strategic value of locally created knowledge.

Hypothesis 5 states that knowledge management performance is affected significantly by implementation measures. Three stepwise regression analyses were conducted to test this hypothesis. The first test was a simple regression analysis, in which both the performance measures and the implementation measures were treated as single, composite variables. The second test treated the performance measures as one variable and the four dimensions of the implementation measures as four variables. The third test treated each of the two dimensions of the performance measures as a dependent variable. The first test resulted in a regression equation:

KM Performance = 1.086 + 0.684 * KM Implementation Measures
F-value = 15.517, (p < 0.1 for the intercept and p < 0.001 for the regression coefficient), Adjusted R^2 = 0.24

This result showed that the implementation measures did exhibit significant influence on the performance measures. The second and the third regression analyses were conducted to further investigate this relationship.

In the second regression analysis, each of the four dimensions of the KM implementation measures was treated as an independent variable, with the performance measures as a single composite variable. The third regression analysis related the four dimensions of the KM implementation measures with the two dimensions of

Table 6. Effect of knowledge acquisition strategy on knowledge management performance

KM Performance	KA Strategy (i)		t-value	p-value
	Internal-Oriented	External-Oriented		
Internal Process	3.2583	3.5123	-1.028	0.310
Overall Organizational Performance	3.4500	3.5852	-0.645	0.522

the performance measures. Table 7 shows that only human resource planning and the incentive measures significantly affected the performance of KM efforts. Between the two dimensions of the performance measures, human resource planning and incentive measures contributed significantly to both the internal process and the overall performance, as shown in Table 8. The second and third regression analyses essentially confirmed the result of the first regression analysis and also provided a greater understanding of the importance of people factors in effective KM.

CONCLUSION AND DISCUSSION

This study investigated the practice and the effect of KM in hospital management. It examined the effect of knowledge characteristics on knowledge acquisition strategy and KM implementation mea-

Table 7. Impact of implementation measures on KM performance (stepwise regression analysis)

Implementation Measures	Standardized Regression Coefficient	Correlation Coefficient (R)	R^2	F-value	p-value
Human Resource Planning	0.506	0.638	0.407	30.851	0.000*
Incentive Program	0.305	0.695	0.482	20.510	0.000*
Standardize Regression Equation: KM Performance = 0.506x Human Resource Planning + 0.305x Incentive Program					

**: Level of Significance p < 0.001*

Table 8. Impact of implementation measures on KM performance

	Internal Process		Overall Organizational Performance	
	Regression Coefficient	p-value	Regression Coefficient	p-value
Intercept	-0.225	0.731	0.615	0.333
Information Infrastructure	-0.231	0.170	0.083	0.602
Incentive Program	0.380	0.019*	0.260	0.016*
Expertise Development	0.046	0.530	0.051	0.469
Human Resource Planning	0.850	0.000**	0.437	0.015*
F-value	11.487		6.258	
p-value	0.000		0.000	
Adjusted R^2	0.522		0.373	

**: p < 0.05; **: p < 0.001*

sures, the effect of knowledge acquisition strategy on implementation approaches and performance measures, and the effect of implementation approaches on KM performance. Using primary data gathered in Taiwanese hospitals, the study found that knowledge characteristics significantly affected implementation measures, and implementation measures, in turn, had a significant effect on the result of KM implementation. More specifically, the level of explicitness of knowledge (knowledge mode) was found to have a significant effect on the adoption of the internal-oriented knowledge acquisition strategy. The complexity of knowledge was related significantly to the external-oriented knowledge acquisition strategy. In general, people factors, such as incentive programs and human resource planning, appeared to have more impact on the result of KM than technology factors, confirming a popular belief regarding the importance of non-technological factors in KM, as reported in the existing literature.

The correlation relationship between knowledge characteristics and implementation measures is worth noting. In addition to incentive programs and human resource planning, information infrastructure was found to be affected significantly by three knowledge characteristics: explicitness, proprietary nature, and variation. The enabling capability of information technology for knowledge storage, dissemination, and application allows for integration of specialized knowledge with routine business processes, such as disease diagnosis and medicine prescription in the healthcare domain (Davenport & Glaser, 2002). Increasingly, hospitals are relying on technological measures as a strategic vehicle to cope with both professional and managerial challenges. The emphasis of information technology applications can be attributed to the collaboration between major hospitals and universities. Information technology development and applications have been a major thrust in Taiwan. Many universities set up research centers to function as an interface between information system researchers and the regional hospitals. Some technologies resulting from these collaborations either are adopted by or are shared with smaller hospitals. The maximal benefit of information technology deployment has yet to be realized, however. The value of technological tools is maximized when they become integral parts of truly integrated business processes.

This study contributed to the field of KM in several ways. For practicing KM professionals and business functional managers, the study demonstrates the importance of matching KM implementation measures with knowledge characteristics. Understanding characteristics of valuable knowledge in an organization is a prerequisite for effective management of its organizational knowledge. Similarly, knowledge-driven organizations consciously must evaluate and adopt their implementation measures. When properly implemented, KM programs may produce visible benefits, including those associated with operationally oriented goals and those associated with long-term financial outcomes.

For KM research, this study empirically revealed the correlation between knowledge characteristics and KM implementation measures. The performance impact of implementation measures, especially those associated with people elements, also was confirmed. The lack of the knowledge acquisition strategy's influence on performance measures suggested that hospitals may have acknowledged the value of both internal and external sources of knowledge and indiscriminately adopted both strategies in order to enrich their knowledge repositories.

Because of several limitations of the study, however, due caution must be exercised in interpreting and applying the results of this study. For example, since the data were gathered in the hospitals in Taiwan, cultural and other environmental differences may limit one's ability to generalize the research findings. Similar researches must be conducted in a variety of cultural and societal settings in order to establish the external validity in a more general fashion.

Another research limitation was that some questionnaires were filled out by the targeted respondents' delegates, including medical doctors and information technology managers. Although the delegates possessed proper understanding of knowledge characteristics or implementation measures, they might not be the best people to answer the questions about overall organizational performance resulting from KM implementation. Additionally, the subjective nature of the questionnaire responses based on the Likert scale constituted a limitation that might discount the validity of the research results.

Healthcare organizations in the United States use information technology extensively to automate important processes and support other aspects of operations. An industry expert estimated that spending in healthcare information technology will increase about 9% per year for the next three years and will reach $30 million in 2006 (Powers, 2004). In light of the important role played by healthcare organizations in our daily lives, more research is needed to provide knowledge about how to improve the quality of hospital management. Based on the findings of this research, two directions may be suggested for future research in this particular domain. First, in-depth case studies may be conducted to verify the validity of our research findings in specific contexts and to assess how contextual factors impact the linkages among knowledge characteristics, KM implementation measures, knowledge acquisition strategy, and KM performance measures. Second, our research model may be replicated in different societies to allow for cross-cultural comparison. Such comparison is necessary in order for the research model to become better established in the KM literature.

REFERENCES

Abeysekera, I. (2003). Intellectual accounting scorecard—Measuring and reporting intellectual capital. *Journal of American Academy of Business, 3*(1/2), 422–440.

Adams, G.., & Lamont, B.T. (2003). Knowledge management systems and developing sustainable competitive advantage. *Journal of Knowledge Management, 7*(2), 142–154.

Awad, E.M., & Ghaziri, H.M. (2004). *Knowledge management.* Upper Saddle River, NJ: Pearson Prentice Hall.

Becera-Fernandez, I., Gonzales, A., & Sabherwal, R. (2004). *Knowledge management: Challenges, solutions, and technologies.* Upper Saddle River, NJ: Pearson Prentice Hall.

Camilleri, D., & O'Callaghan, M. (1998). Comparing public and private hospital care service quality. *International Journal of Health Care Quality Assurance, 11*(4), 127.

Chourides, P., Longbottom, D.L., & Murphy, W. (2002). Excellence in knowledge management: An empirical study to identify critical factors and performance measures. *Measuring Business Excellence, 7*(2), 29–46.

Darroch, J. (2003). Developing a measure of knowledge management behaviors and practices. *Journal of Knowledge Management, 7*(5), 41–54.

Davenport, T.H., & Glaser, J. (2002). Just-in-time delivery comes to knowledge management. *Harvard Business Review, 80*(7), 107–111.

Davenport, T.H., De long, D.W., & Beers, M.C. (1998). Successful knowledge management projects. *Sloan Management Review,* 443–57.

Davenport, T.H., & Prusak, L. (2000). *Working knowledge: How organizations know what they know.* Harvard Business School Press.

Gebert, H., Geib, M., Kolbe, L., & Brenner, W. (2003). Knowledge-enabled customer relationship management: Integrating customer relationship management and knowledge management concepts. *Journal of Knowledge Management, 7*(5), 107–123.

Gloet, M., & Berrell, M. (2003). The dual paradign nature of knowledge management: Implications for achieving quality outcomes in human resource management. *Journal of Knowledge Management, 7*(1), 78–89.

Gupta, A.K., & Govindarajan, V. (2000). Knowledge management's social dimension: Lessons from Nucor Steel. *Sloan Management Review, 42*(1), 71–80.

Hansen, M.T., et al. (1999). What's your strategy for managing knowledge? *Harvard Business Review, 77*(2), 106–116.

Holsapple, C.W. (2003). Knowledge and its attributes. In C.W. Holsapple (Ed.), *Handbook on knowledge management* (Vol. 1) (pp. 166–188). Springer.

Howells, J. (1996). Tacit knowledge, innovation and technology transfer. *Technology Analysis & Strategic Management, 8*(2), 106–116.

Kim, S., Hwang, H., & Suh, E. (2003). A process-based approach to knowledge flow analysis: A case study of a manufacturing firm. *Knowledge and Process Management, 10*(4), 260–276.

Lee, S.M., & Hong, S. (2002). An enterprise-wide knowledge management system infrastructure. *Industrial Management + Data Systems, 102*(1/2), 17–25.

Leonard-Barton, D. (1995). Wellsprings of knowledge. *Harvard Business Review, 71*(2), 75–84.

Liebowitz, J., & Wilcox, L.C. (1997). *Knowledge management*. Ft. Lauderdale, FL: CRC Press.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. London: Oxford University Press.

Nunnally, J.C. (1978). *Psychometric theory,* (2nd ed.). McGraw-Hill.

O'Dell, C., Elliott, S., & Hubert, C. (2003). Achieving knowledge management outcomes. In C.W. Holsapple (Ed.), *Handbook on knowledge management* (pp. 253–287).

Polanyi, M. (1996). *The tacit dimension*. London: Routledge & Kegan Paul.

Porter, M., & Teisberg, E.O. (2004). Redefining competition in health care. *Harvard Business Review, 82*(6), 65–72.

Powers, V. (2004). Improving IT wellness—Health-care organizations adopt knowledge-enabled technology. *KM World*, 10, 11, 26.

Sarvary, M. (1999). Knowledge management and competition in the consulting industry. *California Management Review, 41*(2), 95–107.

Sharkie, R. (2003). Knowledge creation and its place in the development of sustainable competitive advantage. *Journal of Knowledge Management, 7*(1), 20–31.

Soo, C., Devinney, T., Midgley, D., & Deering, A. (2002). KM: Philosophy, process, and pitfall. *California Management Review, 44*(4), 129–150.

Stone, D.N., & Warsono, S. (2003). Does accounting account for knowledge? In C.W. Holsapple (Ed.), *Handbook on knowledge management* (Vol. 1). Springer.

Teece, D.J. (1998). Research directions for knowledge management. *California Management Review, 40*(3), 289–292.

Teece, D.J. (2003). Knowledge and competence as strategic assets. In C.W. Holsapple (Ed.), *Handbook on knowledge management* (Vol. 1) (pp. 129–188). Springer.

Tiwana, A. (2000). *The knowledge management toolkit*. Upper Saddle River, NJ: Prentice-Hall.

Tsai, M.-T., Yu, M.-C., & Lee, K.-W. (2005). Developing e-business systems based on KM process perspective—A case study of Seven-Eleven Japan. *Journal of American Academy of Business*, *6*(1), 285–289.

Ulrich, D., & Smallwood, N. (2004). Capitalizing on capabilities. *Harvard Business Review, 82*(6), 119–128.

Van Beveren, J. (2003). Does health care for knowledge management? *Journal of Knowledge Management*, *7*(1) 90–95.

Wang, S. (2002). Knowledge maps for managing Web-based business. *Industrial Management + Data Systems*, *102*(7), 357–364.

Wiig, K.M. (1994). *Knowledge management: The central management focus for intelligent-acting organizations.* Arlington, TX: Schema Press.

Zack, H.M. (1999). Developing a knowledge strategy. *California Management Review, 41*(3), 125–145.

APPENDIX I: SUMMARY OF HYPOTHESIS TESTING RESULTS

H1	The primary knowledge acquisition strategy is significantly affected by knowledge characteristics.	Partially Accepted
H1.1	The primary knowledge acquisition strategy is significantly affected by knowledge mode.	Accepted
H1.2	The primary knowledge acquisition strategy is significantly affected by knowledge complexity.	Accepted
H1.3	The primary knowledge acquisition strategy is significantly affected by the proprietary nature of knowledge.	Rejected
H1.4	The primary knowledge acquisition strategy is significantly affected by knowledge volatility.	Rejected
H2	Knowledge management implementation measures are significantly affected by knowledge characteristics.	Accepted
H3	Knowledge management implementation measures in the hospitals are significantly affected by the primary strategy adopted for knowledge acquisition.	Partially Accepted
H4	Knowledge management performance is significantly affected by the primary knowledge acquisition strategy.	Rejected
H5	Knowledge management performance is significantly affected by knowledge management implementation measures.	Accepted
H5.1	Internal processes are significantly improved by information infrastructure	Rejected
H5.2	Internal processes are significantly improved by incentive measures for KM implementations.	Accepted
H5.3	Internal processes are significantly improved by expertise development measures.	Rejected
H5.4	Internal processes are significantly improved by human resource planning.	Accepted
H5.5	Overall performance is significantly improved by information infrastructure.	Rejected
H5.6	Overall performance is significantly affected by incentive measures for KM implementations.	Accepted
H5.7	Overall performance is significantly affected by expertise development.	Rejected
H5.8	Overall performance is significantly affected by human resource planning.	Accepted

APPENDIX II: SURVEY QUESTIONNAIRE

(Measurement Scale: 1 = Highly Disagree, 5 = Highly Agree)

Questions About Knowledge Characteristics

1. Documentation has been created for all medical expertise.
2. Medical service delivery processes have formal specifications.
3. Doctor's practice experience may be documented in writing.
4. Doctors can share their professional expertise without any obstacles.
5. Major surgical operations are accomplished by task force teams.
6. Significant differences of expertise exist among the doctors in the same specialty.
7. Outside experts often are called upon to assist with major surgery operations.
8. Mutual support among doctors within the same specialty usually is difficult to come by.
9. Medical expertise is integrated tightly with hospital management and organizational culture.
10. Outsourcing often is used due to inadequacy of medical expertise.
11. Innovations of medical practices are difficult to be obtained by competition.
12. Doctors are expected to remain up-to-date with data in their expertises.
13. The frequency of rare case treatment experience is higher than the competition.
14. New and innovative medical knowledge and technology are adopted faster than competition.
15. Knowledge used around here advances fast.

Knowledge Management Implementation Measures:

16. Information systems are developed aggressively to enable organization, dissemination, and application of knowledge.
17. Doctors are strongly encouraged to access document bases and to systematically construct medical databases.
18. Operation automations through information technology are pursued actively to support doctors' work.
19. Substantial amounts of financial resources are invested in information technology.
20. Doctors are encouraged to use the Internet to enhance medical expertise exchange and diffusion.
21. Knowledge sharing is an important criterion in performance evaluation.
22. Proposals for creative ideas are rewarded, even when the ideas prove to be wrong.
23. Knowledge creation and sharing often are rewarded with salary increases and bonuses.
24. Knowledge creation and sharing are rewarded with promotions.
25. Doctors always are willing to accept training and work assignments that are tougher than the competition.
26. The hospital is not hesitant to increase head counts of supporting technical specialists.
27. Doctors are willing to accept the challenges to enhance their professional expertises.
28. Doctors often explicitly reject the idea of being evaluated by personnel from other fields.
29. Competition among doctors in the same field often hinders knowledge sharing.
30. Doctors are strongly encouraged to learn and to innovate.
31. Open and smooth channels of communication exist in the hospital.

32. Doctors frequently are encouraged to engage themselves in experience and expertise exchange.
33. One-on-one mentor and apprentice-style training of resident doctors is common here.

Knowledge Management Performance

(5 = Much Better, 3 = About the Same, and 1 = Much Worse)

34. Doctors' expertise and experience exchange
35. Handling of doctors' suggestions with regard to medical operations
36. Doctors' sense of participation
37. Decision-making speed
38. Proposal preparation cycle time
39. Overall efficiency improvement
40. Overall quality of service
41. Patient satisfaction
42. Decrease of number of administrative personnel
43. Reduction of impact caused by turnover
44. Handling of medical service improvement projects

Primary Strategy for Knowledge Acquisition

(Check only one of the following)

___ University
___ Mentor and Apprentice System
___ Consultant
___ Internship with other hospitals
___ Morning Meetings
___ Intranet
___ Departmental meetings
___ Outside Seminars and Conferences
___ Medical Databases in the Hospital
___ Documentation Management (including operational manuals and instructions)
___ Others

This work was previously published in the International Journal of Knowledge Management, edited by M. E. Jennex, Volume 2, Issue 3, pp. 1-20, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Section V
Organizational and Social Implications

This section includes a wide range of research pertaining to the social and organizational impact of medical informatics around the world. Chapters introducing this section analyze organizational factors in health informatics and patient centric health information systems, while later contributions offer an extensive analysis of the Internet's role in revolutionizing accessibility to medical-related information. The inquiries and methods presented in this section offer insight into the implications of medical informatics at both a personal and organizational level, while also emphasizing potential areas of study within the discipline.

Chapter 5.1
Organizational Factors in Health Informatics

Michelle Brear
University of New South Wales, Australia

ABSTRACT

There is a general recognition that numerous organizational factors will influence the success of an informatics intervention. This is supported by a body of evidence from multi-disciplinary and health-specific research. Organizational factors are highly interrelated and the exact nature and contribution of each to the success of an intervention is not clear. A health-specific understanding and recognition of these factors is necessary if informatics applications are to reach their potential in healthcare settings.

INTRODUCTION

The influence of organizational factors on the success of informatics interventions in healthcare has been clearly demonstrated. This health-specific research, informed by a larger body of evidence emerging from interdisciplinary organizational, psychological, and sociological research, has confirmed the view that organizational factors can be the decisive factor in the success of an intervention (Lorenzi, Riley, Blythe, Southon, & Dixon, 1997).

However, it remains rare for organizational factors to be explicitly addressed in the implementation process. As such, their contribution to the success or failure of informatics applications is not properly understood. This has implications for future interventions. Applications that were not utilized or did not perform adequately in a particular setting may be dismissed, while other less appropriate systems may be adopted because organizational factors influenced their success. Explicit study of the role of organizational factors on the implementation of health-informatics interventions is necessary to develop an understanding of their influence in the healthcare context.

Healthcare organizations tend to be highly task oriented, labor intensive, and dependent on interdisciplinary teamwork, so the influence of organizational factors within them may differ considerably from the business settings in which they have traditionally been studied (Chau, 2001). Health organizations are also increasingly underresourced due to the global downturn in government social spending, to health-sector privatization, and to aging populations. It is these characteristics that necessitate the rapid uptake of informatics applications, capable of automating aspects of healthcare provision and reducing labor intensity (Coiera, 2004).

From a technical perspective, rapid and fundamental transformation of the healthcare sector through informatics is achievable. However, without a clear understanding of, and ability to, manage organizational factors, it is unlikely that informatics applications will realize their potential in the health sector. This short review provides an overview of the key organizational factors influencing the success of informatics interventions. It begins by positioning informatics interventions in the broader context of organizational change before discussing the current understanding of selected factors.

INFORMATICS IMPLEMENTATION AS ORGANIZATIONAL CHANGE

Implementing informatics applications is essentially “a politically textured process of organizational change” (Berg, 1999, p. 87), aimed at achieving user acceptance and the utilization of informatics applications. Organizational change requires people to be aware of a need for change, to identify a particular course through which the change can occur, and to take actions to make it happen (Lorenzi, 2004). Resistance to change occurs if users are not aware of the need for change, are not convinced of the course of action set out, or are unable to carry out the necessary action. It is the users, not the technology, that should be the centre of the change process as the decision to utilize the system is ultimately theirs (Berg).

Even the best designed and well-intentioned informatics interventions are likely to lead to productivity losses in the early stages and create major changes (Lorenzi, 2004). The timely and effective training of users can reduce the disruption; however, it is not enough to ensure success as even a correctly used system can have far-reaching effects. Informaticians taking a sociotechnical approach view the application as one component of a complex system (the health organization) whose introduction will disrupt other components of the system (e.g., patients and clinicians). They advocate design approaches that aim to create technology that fits within the complex system (Kaplan, 2001).

The multidisciplinary nature of health-sector organizations makes finding the correct fit challenging (Kaplan, 2001). A range of professionals with different needs, expectations, and work norms are likely to use an application, and each will expect it to fit with his or her work practice. When an application does not fit, resistance will increase. This is often due to valid concerns about increased workload or the ability to care for patients (Timmons, 2003). When systems do not fit, the best way to overcome resistance is to change them. However, when they are essentially effective, resistance can be overcome by changing people’s opinions or work norms. Organizational culture and social networks, from which many of these norms and opinions arise, need to be understood and managed.

ORGANIZATIONAL CULTURE

Organizational culture is the set of shared norms, values, and tacit rules within which members of an organization function (Lorenzi & Riley, 2000). “Every culture supports a political and social values system” (Lorenzi et al., 1997, p. 85) that will influence the reaction to an informatics application. Healthcare settings often involve a professional hierarchy between doctors and nurses, are characterized by high levels of informal and disruptive communication, and place value on clinician-patient relationships and patient care.

It is necessary to identify and target the aspects of organizational culture presenting opportunities for and barriers to success when changing the organization through an informatics intervention. Managing change requires mediating the influence of culture on events rather than necessarily aiming to change it (Demeester, 1999). Where organizational culture and informatics applications appear incompatible, adaptation of the application should be considered.

If it is not possible to modify the system, success is dependent on changing the organizational culture to make it compatible. Cultural change directly targeted at the strongly held values of users may only increase resistance. If the organizational culture supports a belief that informatics applications undermine good clinician-patient relationships, attempting to convince clinicians that good relationships with their patients are not important is unlikely to be a successful strategy for winning acceptance of the application. However, it may be possible, through an educational process, to convince clinicians that informatics applications do not necessarily undermine good relationships and in the right conditions can even enhance them. Users may already be convinced of the need to change some aspects of organizational culture that do not threaten the values they are most passionate about. The structure of an organization, and work patterns and roles of individuals within it are influential and may be appropriate areas to encourage change.

Organizational Structure

The structure of an organization will affect the way in which decisions are made, the type of leadership that emerges, and the way resistance is dealt with in the implementation process. Flatter organizational structures tend to encourage the sharing of ideas, the emergence of innovation, and broader involvement in decision making (Leonard, Graham, & Bonacum, 2004). In these types of organizations, management tends to adopt a collaborative approach, working alongside, listening to, and involving those working on the ground rather than making decisions on their behalf and communicating orders. Management is supportive, approachable, and accountable, and shows dedication to continuous learning (Zimmerman et al., 1993). These types of organizations are more likely to include practicing clinicians in formal decision-making bodies such as management committees. They are also more likely to recognize, encourage, and legitimize the role of grassroots leaders with clinical credibility and presence, and have a commitment to involving informatics users in the implementation process.

User involvement throughout the process "leads to increased user acceptance and use by encouraging realistic expectations, facilitating the user's system ownership, decreasing resistance to change, and committing users to the system" (Lorenzi et al., 1997, p. 86). It also allows a better definition of problems and solutions from the user's perspective, and develops a better understanding amongst users of the application (Lorenzi et al.). Involving users in the design and implementation of a system is more likely to result in applications suited to the current work patterns of the intended users.

Work Patterns and Roles of Clinicians

Any informatics application must be compatible with the current work practices and values of the organization (Greenhalgh, Robert, MacFarlane, Bate, & Kyriakidou, 2004; Kaplan, 2001; Lorenzi & Riley, 2000). Compatibility will differ between organizations and cultures, so applications require the capacity to be tailored to the needs of individual organizations. Take an electronic prescribing system, for example. In one hospital, it may be used to enter prescription orders during ward rounds via a laptop computer; however, a different hospital (or even another ward within that hospital) may find it more appropriate to install the system on a computer terminal at the nurses' stations so that orders can be entered retrospectively. Bearing in mind the necessity to consider the unique context of individual organizations, it is possible to make some generalizations regarding the work patterns and roles of clinicians to broadly inform the design of informatics interventions.

Doctors have traditionally worked with a high degree of professional autonomy and status (Gagnon et al., 2003). Applications perceived to undermine their autonomy and status as professionals or "subvert the art of medical practice" (Kaplan, 2001, p. 4) are more likely to meet with resistance. In a qualitative study examining factors influencing the adoption of a CDSS (clinical decision-support system) involving automatic clinical reminders, Rousseau, McColl, Newton, Grimshaw, and Eccles (2004) found clinicians favored on-demand evidence systems to automatically generated reminders. The latter were perceived to be intrusive, often inapplicable, and not particularly useful for making patient-management decisions.

Clinicians tend to be patient focused, require an ability to maintain control of patient care, and make decisions specific to individuals, which computers are not capable of. They must be convinced that applications will not jeopardize their ability to care for patients (Timmons, 2003). Overly prescriptive systems, or those that attempt to take on the uniquely human quality of thinking, are unlikely to be successful. Rouseau et al. (2004) noted inapplicable reminders were a barrier to effective use and found that practitioners formed a habit of ignoring all reminders. In a study of adherence to electronic HIV treatment reminders, Patterson, Nguyen, Halloran, and Asch (2004) found that the inapplicability of reminders to many patients' specific situations and the time taken to document why the reminders were not adhered to were significant barriers to effective system use.

All applications will create some change to normal work patterns and roles. That is essentially what their implementation is intended to do (Berg & Toussaint, 2003). Users need to be realistically informed of and prepared for changes to normal work practice. It is inevitable that users will expect some future benefit from adapting their behaviors, and realistic communication of the likely benefits, particularly if they are indirect or not clearly visible, should form an integral part of the communication strategy.

Communication

Communication binds individuals together and is integral to the implementation process. Without effective communication, it is impossible to lead, learn, make decisions, prepare individuals for an intervention, or use the intervention effectively (Zimmerman et al., 1993). The lack of communication, or ineffective communication that lacks trust, can negatively influence the uptake of a technology (Ash, 1997). There is no magic formula for effective communication; however, there are some key principles that should be applied to enhance the effectiveness of communication.

Communication must be timely. People require time to digest information and prepare for changes (Tilley & Chambers, 2004). However, they also forget. Users who are trained to use a system months before its implementation may not remember how to use it when it finally arrives. Information must also be communicated at an appropriate time in the day. Disrupting a lunch break or patient care may not be the most appropriate way to inform users about a new application.

Communication must be sincere and truthful. Users must be offered an honest and realistic assessment of the potential negative consequences and expected benefits of an application. For example, if an electronic medical-record application is introduced, it may be realistic to expect its use to be more time consuming in the initial stages while users master the system. However, in time, its ability to provide comprehensive patient information at the click of a button may save time and provide better quality information. It is also important to acknowledge unexpected benefits and problems if they occur.

Sources of communication have an immense impact on the perceived message, its credibility, and its influence (Kaplan, 2001). A manufacturer's leaflet declaring the new system as easy to use is unlikely to carry much weight with users; however, a respected clinical peer conveying this information may be quite influential.

It is essential that the communication be recognized as multidimensional rather than a one-way channel from management to users. Mechanisms for receiving feedback must be created, and where it is not forthcoming, it should be actively elicited from users. Magrabi, Westbrook, Coiera, and Gosling (2004) incorporated two mechanisms for feedback into an online DSS. Users could volunteer feedback at any time; however, it was also actively elicited by randomly prompting users. Once received, feedback should be acted upon to adapt the application to meet user needs (Greenhalgh et al., 2004). For example, in response to clinician feedback, the patterns of automatically generated reminders in a CDSS were altered to limit the number of reminders perceived by clinicians to be inapplicable. Communication through informal structures (e.g., gossiping in the tea room) is inevitable, and the ability to manage it directly is limited. However, the less effective formal communication mechanisms are, the more likely it is that communication will take place through informal channels in the social network.

SOCIAL NETWORKS

A social network consists of the individuals, groups, and organizations with whom, and patterns of communication by which, individuals in an organization interact. It is through social networks that organizational culture and behaviors are reinforced and adapted (Lorenzi et al., 1997). The culture represented within a social network can be influential. For example, Gagnon et al. (2003) found that physicians who perceived social and professional responsibility from others in their social network to adopt telemedicine applications had a stronger intention to do so. The culture represented in each individual's social network will differ, and each individual within will interact and be influenced differently.

To properly understand a social network, it is necessary to examine the interactions, not the individuals. The frequency of interaction is important but should not be viewed in isolation from the style of communication (e.g., formal or informal), the type of communication (e.g., synchronous or asynchronous), the strength of ties between the participants in an interaction, and the power relations involved (Katz, Lazer, Arrow, Contractor, 2004). When implementing an informatics application, interactions that occur in clinical teams and with respected opinion leaders are particularly influential in relation to individuals' decisions to utilize applications.

Effective Teamwork

The clinical team has been identified as the organizational unit most influential in the diffusion of innovation (Gosling, Westbrook, & Braithwaite, 2003). Well-functioning teams facilitate effective communication, encourage continuous learning, and offer a trusting environment in which ideas and issues can be raised. Through these interactions, teams develop shared visions and common goals that support the introduction of new innovations to fulfill these goals. As work in health organizations is highly dependent on teamwork, and it is teams, not individuals, that must adopt informatics applications, well-functioning teams are a prerequisite for successful informatics implementation (Goldstein et al., 2004).

In organizations where well-functioning teams do not exist, an informatics intervention may be an opportunity to develop teams by uniting individuals around a shared vision and common goals that the application can fulfill. Doctors and nurses take on different roles and responsibilities in the process of caring for patients, so a shared vision may not be immediately apparent. However, both groups ultimately work toward the goal of providing optimal patient care, so incorporating an application into a vision of improved patient care may be a way to unite users.

Consideration should also be given to the size and composition of teams. Teams of more than 15 tend to fragment into subteams, while very small teams have a tendency to become cliquish, so ideally they should consist of 10 to 15 members (Gosling et al., 2003). The work environment may largely dictate a team's composition. In healthcare settings, teams are usually multidisciplinary. Individuals, however, are more likely to create ties with those they perceive to be similar, so identifying similarities amongst multidisciplinary teams is pertinent (Katz et al., 2004). Identifying and utilizing the influence of respected clinicians within the team can also be useful.

Clinical Champions and Opinion Leaders

Respected clinicians with influence amongst their peers and who support the intervention play an important role in convincing others of an application's worth, as do those who oppose the intervention. They are commonly referred to as clinical champions or opinion leaders. As the name champion tends to imply a positive influence and does not necessarily imply influence amongst peers (Lolock, Dopson, Chambers, & Gabbay, 2001), opinion leader will be used in this article.

Opinion leaders are individuals with the ability to influence others in the social network and who make a major personal commitment to diffusing information about an informatics application (Lorenzi & Riley, 2000). Such diffusion may have a negative or positive influence and can discourage or encourage the adoption of the application. For example, Ash (1997) identified the presence of champions as a significant factor in the diffusion of e-mail in academic health-science centres. Conversely, Timmons (2003) discusses a "strong and articulate" ward sister whose resistance to using an electronic system in the ward was successful in preventing its implementation. It is not unusual to have both negative and positive opinion leaders within a network.

Whatever their persuasion, they tend to be charismatic individuals with good interpersonal relationships based on trust and understanding. They act through clinical conviction, generally outside of the formal structures, and give applications credibility at a local level. Their role is essentially informal, and, on the proviso their colleagues respect them, they tend to be self-appointed (Lolock et al., 2001). The role is largely dependent on personal motivation and conviction, and therefore it is difficult to formalize. The potential alienating effect of being an opinion leader means that those without sufficient commitment are likely to be reluctant to take on such a role (Lolock et al.)

Despite a general consensus that successful interventions are more likely when champions are present, there is a lack of understanding about exactly what it is they do, the circumstances in which they will be influential, and how best to describe them. As with each of the factors mentioned here, opinion leaders are one small component of a large and complex system (the health organization), and it is difficult to isolate their effect. There is also considerable evidence to suggest that other factors, in particular the suitability of the application, influence the emergence of opinion leaders. Further research is needed to identify how opinion leaders influence, and how the champions amongst them can be encouraged.

CONCLUSION

There is a general recognition that numerous organizational factors will influence the success of an informatics intervention. This is supported by a body of evidence from multidisciplinary and health-specific research. In particular, research has noted the influence of organizational culture and social networks. Organizational culture, the shared norms and values within which members of an organization function, influences the organization's structure and patterns of communication. Social networks, the individuals and groups with whom one interacts and the interactions that occur between them, are the social space in which teams are formed and work, and in which individuals are influenced, particularly by those individuals known as opinion leaders or clinical champions. Organizational factors are highly interrelated, and the exact nature and contribution of each to the success of an intervention is not clear. A health-specific understanding and recognition of these factors is necessary if informatics applications are to reach their potential in healthcare settings.

REFERENCES

Berg, M. (1999). Patient care information systems and healthcare work: A sociotechnical approach. *International Journal of Medical Informatics*, *55*, 87-101.

Berg, M., & Toussaint, P. (2003). The mantra of modeling and the forgotten powers of paper: A sociotechnical view on the development of process-oriented ICT in health care. *International Journal of Medical Informatics*, *69*, 223-234.

Chau, P. Y. K., & Hu, P. J.-H. (2002). Investigating healthcare professionals' decisions to accept telemedicine technology: An empirical test of competing theories. *Information and Management, 39*, 297-311.

Coiera, E. (2004). Four rules for the reinvention of healthcare. *British Medical Journal, 328*, 1197-1199.

Demeester, M. (1999). Cultural aspects of information technology implementation. *International Journal of Medical Informatics, 56*, 25-41.

Gagnon, M.-P., Godin, G., Gagne, C., Fortin, J.-P., Lamothe, L., Reinharz, D., et al. (2003). An adaption of the theory of interpersonal behaviour to the study of telemedicine adoption by physicians. *International Journal of Medical Informatics*, *71*, 103-115.

Goldstein, M. K., Coleman, R. W., Tu, S. W., Shankar, R. D., O'Connor, M. J., Musen, M. A., et al. (2004). Translating research into practice: Organisational issues in implementing automated decision support for hypertension in three medical centers. *Journal of the American Medical Informatics Association, 11*(5), 368-376.

Gosling, A. S., Westbrook, J. I., & Braithwaite, J. (2003). Clinical team functioning and IT innovation: A study of the diffusion of a point-of care online evidence system. *Journal of the American Medical Informatics Association, 10*(3), 244-251.

Greenhalgh, T., Robert, G., MacFarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organisations: Systematic review and recommendations. *The Millbank Quarterly, 82*(4), 581-629.

Kaplan, B. (2001). Evaluating informatics applications. Some alternative approaches: Theory, social interactionism, and call for methodological pluralism. *International Journal of Medical Informatics, 64*, 39-56.

Katz, N., Lazer, D., Arrow, H., & Contractor, N. (2004). Network theory and small groups. *Small Group Research, 35*(3), 307-332.

Leonard, M., Graham, S., & Bonacum, D. (2004). The human factor: The critical importance of effective teamwork and communication in providing safe care, quality and safety in health care, *13*(Suppl. 1), i85-i90.

Lolock, L., Dopson, S., Chambers, D., & Gabbay, J. (2001). Understanding the role of opinion leaders in improving clinical effectiveness. *Social Science and Medicine, 53*, 745-757.

Lorenzi, N. M. (2004). Beyond the gadgets: Non-technological barriers to information systems need to be overcome. *British Medical Journal, 328*, 1146-1147.

Lorenzi, N. M., & Riley, R. T. (2000). Managing change: An overview. *Journal of the American Medical Informatics Association, 7*(2), 116-124.

Lorenzi, N. M., Riley, R. T., Blythe, A. J. C., Southon, G., & Dixon, B. J. (1997). Antecedents of the people and organizational aspects of medical informatics: Review of the literature. *Journal of the American Medical Informatics Association, 4*(2), 79-93.

Magrabi, F., Westbrook, J. I., Coiera, E. W., & Gosling, A. S. (2004). Clinicians' assessments of the usefulness of online evidence to answer clinical questions. *Proceedings of the 11th World Congress on Medical Informatics* (pp. 297-300).

Patterson, E. S., Nguyen, A. D., Halloran, J. P., & Asch, S. M. (2004). Human factors barriers to the effective use of ten HIV clinical reminders. *Journal of the American Informatics Association, 11*(1), 50-59.

Rousseau, N., McColl, E., Newton, J., Grimshaw, J., & Eccles, M. (2003). Practice-based, longitudinal, qualitative interview study of computerised evidence based guidelines in primary care. *British Medical Journal, 326*, 314-321.

Tilley, S., & Chambers, M. (2004). The process of implementing evidence-based practice: The curates egg. *Journal of Psychiatric and Mental Health Nursing, 11*, 117-119.

Timmons, S. (2003). Nurses resisting information technology. *Nursing Inquiry, 10*(4), 257-269.

Zimmerman, J. E., Shortell, S. M., Rousseau, D. M., Duffy, J., Gillies, R. R., Knaus, W. A., et al. (1993). Improving intensive care. Observations based on organisational case studies in nine intensive care units: A prospective multicenter study. *Critical Care Medicine, 21*(10), 1143-1451.

KEY TERMS

Clinical Champion: Clinical champions are opinion leaders who champion or encourage the uptake of an application.

Opinion Leader: An opinion leader is an individual, respected amongst his or her peers, who acts out of clinical conviction to influence the opinions of others vis-à-vis an informatics application.

Organizational Culture: Organizational culture is the set of shared norms, values, and tacit rules within which members of an organization function.

Organizational Factors: In an informatics context, organizational factors are factors relating to the culture and functioning of an organization that, negatively or positively, influence its ability to adapt to an informatics intervention.

Social Network: A social network consists of the individuals, groups, and organizations with whom an individual interacts, and the interactions that take place between the individual and other components of his or her social network.

Sociotechnical Approach: A sociotechnical approach is one that views informatics applications as part of the broader social and political context within which they are implemented.

Teamwork: The cooperative effort of a small group to achieve a specified outcome.

This work was previously published in Handbook of Research on Informatics in Healthcare and Biomedicine, edited by A. A. Lazakidou, pp. 410-418, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 5.2
Information Imbalance in Medical Decision Making:
Upsetting the Balance

Jimmie L. Joseph
The University of Texas at El Paso, USA

David P. Cook
Old Dominion University, USA

ABSTRACT

This chapter explores the ethical implications of a reduction in information asymmetry between health care providers and their patients. In many human interactions, asymmetry of information and experience typically raises ethical dilemmas for the party with the greater degree of information. This chapter illustrates that it is the reduction in information asymmetry that is raising ethical dilemmas in dealing with medical issues. Understanding this phenomenon may assist in identifying and managing future ethical quandaries that may occur as Internet resources provide broad access to information previously distributed only to a subset of the population.

INTRODUCTION

Information asymmetry is an imbalance in the information available to the parties in an interaction. This situation is not atypical and can be considered the basis for many of our most meaningful conversations. Human interactions would be far less interesting if everyone knew and "shared" the same thoughts, feelings, and discoveries at all times. While information asymmetry makes questions such as "How is the weather?" and "How was your day?" at least superficially interesting, in numerous human interactions, asymmetry of information and experience can raise ethical dilemmas for the party with the greater degree of information.

Significant academic research exists in the area of information asymmetry (e.g., finance and economics). The presumption of much of this research is that market efficiency can be increased by reducing the degree of asymmetry (Akerlof, 1970; Milgrom & Roberts, 2001; Payne, 2003). Furthermore, it is commonly held that information parity (or information access that moves the parties towards this point) can be the linchpin for the elimination of the ethical quandaries introduced by information asymmetry (Akerlof, 1970; Diamond, 1984; Hellwig, 2001).

In the field of medicine, for example, it is clear that the physician has historically possessed more information than the traditional patient. This disparity has created a peculiar set of ethical issues for medical providers, based on their fiduciary responsibility to their patients. As with many relationships, the relationship between caregiver and patient changes, however, as information asymmetry diminishes. Relaxing the assumption that there is a significant difference in the information available to the physician and patient causes a new set of ethical issues to arise. In the extreme case, information inversion exists: the patient knows more about his or her condition than the attending medical personnel. The anecdotes about doctors making poor patients approach the status of urban myth (Porter, 1992). This chapter focuses on these very issues: how do the physician-patient relationship and the resultant ethical issues change as the patient's information disadvantage decreases.

This chapter examines the effects of changes in technology and societal norms on the ethical posture of practitioners in the medical profession. Ethical, social, technical, and regulatory issues associated with consumers' increased access to medical information are explored. It is posited that situations exist where a reduction in information asymmetry does not lead to fewer, or less severe, ethical dilemmas — it leads to more severe ethical quandaries. This chapter illustrates a counterintuitive example to the presumption that increased information parity enhances market efficiency and diminishes the number of ethical quandaries in the provision of medical care. While the chapter seeks to illustrate a counterintuitive example and the ethical dilemmas it raises, it does not purport to provide a solution to these still evolving ethical issues.

INFORMATION ASYMMETRY AND ITS COSTS

Problems resulting from information asymmetry are generally classified as either moral hazard or adverse selection (Akerlof, 1970; Diamond, 1984; Payne, 2003). Moral hazard problems are related to the buyer's inability to observe the actions taken by the seller (Holmstrom, 1985). Adverse selection problems are related to the buyer's inability to observe either pertinent seller characteristics or the contingencies under which the seller operates (Nayyar, 1990). Adverse selection can occur in the medical environment due to the patient's inability to determine on an *ex ante* basis the competence of the physician, the value of the service, or its quality.

Given the existence of such problems, and that they arise as a result of information asymmetries, numerous strategies have arisen to attempt to remedy or mitigate the effect of these asymmetries. Several potential remedies (e.g., contingent contracts, certification, monitoring, and reputation) exist for information asymmetries (Holmstrom, 1985; Nayyar, 1990). Each of the mentioned remedies has its shortcomings. In contingent contracts, for example, the capability must exist to identify all relevant contingencies prior to the enforcement of the contract.

Certifications only indicate an entity has met some minimum standard of performance or knowledge (as established by the certifying authority). For certification to be of benefit, it is necessary for the party with the information paucity in a transaction to understand the nature or quality of the certification or certifying agency (Diamond, 1984; Hellwig, 2001). Further, some certifying authorities may not make their criteria for certification public knowledge (Diamond, 1984; Hellwig, 2001).

Monitoring introduces additional information asymmetries between the monitor and the entity being monitored. An agency problem can arise if the party depending on the monitoring organization has no way of ascertaining the diligence of the monitoring (Diamond, 1984; Hellwig, 2001). This is especially problematic in self-monitoring organizations such as bar association monitoring boards composed of lawyers and state medical boards composed of doctors (Davies & Beach, 2000; Fournier & McInnes, 1997; Marks & Cathcart, 1974). While each of these mechanisms may have been fomented with good intentions, each is fraught with difficulty.

Another mechanism to mitigate or eliminate information asymmetry is greater informational transparency and availability. The Internet is a good example of a vehicle that can facilitate these goals. The Internet presents a medium through which unprecedented volumes of data can be made available to anyone with access and interest. However, just as other forms of asymmetry mediation have unique shortcomings, the Internet may pose its own issues.

Examples of Information Asymmetry

Many of our interactions are predicated on an asymmetry in the information possessed by the parties involved. Though numerous examples of information asymmetry in interactions could be presented, along with their associated effects and resultant moral dilemmas, two situations involving information asymmetry (the sale of automobiles with known defects and insider trading) are briefly examined here. Information asymmetry involving parties to medical interactions will be examined more closely in the ensuing sections of this chapter, and so will not be discussed here.

Automobile Sales

In used car purchases, the prospective purchaser knows exactly how much she can spend on the vehicle, but the seller has no such information. Conversely, the seller may be privy to details of a vehicle's defects, accident history, and undisclosed repairs to which a prospective buyer is not (Akerlof, 1970). For the seller, knowledge of defects creates an ethical dilemma in that he must decide whether to disclose the existence of the defects and possibly nullify the interest of a prospective buyer (Akerlof, 1970).

To help address the issue of seller information advantage, governments (both Federal and state) have developed and imposed "Lemon" laws and other statutes for consumer protection (Akerlof, 1970). The intention of such laws is to facilitate commerce and to protect consumers from the problems created by information asymmetry. These laws also reduce the moral dilemma faced by the seller: it now makes moral and legal sense to disclose all known problems with the automobile (Akerlof, 1970).

Insider Trading

When a person with access to information not available to the general public wishes to trade stock, he or she is faced with a different ethical problem. Insider trading is the result of trading stock on the basis of privileged information (Anonymous, 1998, 2003b). Such trading continues because of the tremendous potential for profit that can accrue from buying or selling ahead of the disclosure of the information to the public (Anonymous, 1998, 2003b). Few would argue that trading stock on the basis of privileged information is a fair practice; as a consequence, laws exist that disallow such activities (Anonymous, 2003b). The purpose of such laws is to level the playing field so that parties without access to privileged information are not disadvantaged on that basis.

Why Medicine is Different

There is an exception to the generally accepted proportional relationship between information asymmetry and ethical concerns: the relationship between medical providers and patients. This chapter posits that as patients achieve a reduction in information asymmetry (RIA) vis-à-vis their medical care providers, the ethical dilemmas of the formerly better informed medical professionals increase in number and/or patient impact, thus inverting the generally accepted relationship between information parity and ethical dilemmas. This trend is of concern as increasing portions of the population have Internet access at work, and high-speed Internet access at home. Omnipresent Internet access allows the general public to view information available only to medical professionals or medical researchers in prior decades.

Medical Information Access

Two characteristics of medical information that have traditionally affected its dissemination are availability and readability (Majno, 1975; Marti-Ibanez, 1962). Availability is defined here as the ability of a person to discover that information about a topic exists, and then to acquire that information. Readability entails the possession of sufficient domain knowledge (assumed to be present in the target audience when writing the information) to understand and utilize available information (Aguolu, 1997).

Prior to the invention of the Guttenberg printing press in 1436, books were hand-crafted and, consequently, too expensive for (unavailable to) the average citizen (Borgman, 2000). Magazines (journals) had not even been conceptualized at that time. Most persons did not know how to read (since it was of little practical value to their daily lives), and books were unlikely to come into their possession. Even most medical practitioners could not read, and medical knowledge was typically passed from generation to generation via an apprentice-like system of medical education (Borgman, 2000; Majno, 1975; Marti-Ibanez, 1958, 1962).

A HISTORY OF MEDICAL INTERACTIONS

The practice of medicine has largely been a dyadic one between the patient and the care provider (doctor and/or nurse, midwife, witch doctor, tribal shaman, acupuncturist, etc.) (Digby, 1997; Majno, 1975; Marti-Ibanez, 1958; Summers, 1997). In the traditional doctor-patient relationship, doctors diagnosed and prescribed a course of treatment, and patients underwent treatment as directed (Majno, 1975). News of treatments available on other continents, in other nations, or in distant regions of their own land may not have reached residents of a locale until long after the discovery (Magner, 1992; Majno, 1975; Marti-Ibanez, 1962; Robinson, 1931).

Given the lack of a formal dissemination mechanism such as "continuing education" (through which medical providers could be fully trained in new techniques and procedures), even if new medical treatment ideas arrived, there was no guarantee that the local medical caregiver would understand how to implement the new regimen. While Marco Polo may have returned with information concerning the existence of acupuncture as a treatment for headaches and other ailments, this did not provide the training necessary for European doctors to perform the procedures (Majno, 1975; Marti-Ibanez, 1958).

Since there were few books prior to the widespread distribution of the printing press, availability of information was the primary limiting factor in medical innovation dissemination (Marti-Ibanez, 1962). The availability of relatively inexpensive books brought about a renaissance in reading, and public education brought reading to the masses. This change, however, did not alter the readability issue of medical texts.

In the 19th century, and through the late 20th century, the primary means of disseminating information regarding the effectiveness of various forms of medical treatment was medical journals. Current and archived volumes of these journals were available at medical libraries and hospitals. Medical professionals could also subscribe to pertinent journals, often at substantial discounts from the retail price, through their medical associations (Majno, 1975).

However, for a citizen who was not a professional in the medical field, the journals presented two familiar hurdles: availability and readability. Unless an individual lived near a medical library, the availability of medical journals was usually limited. Medical journals were (and are) prohibitively expensive for the average citizen. Along with the cost of a current subscription, there is the cost of purchasing past issues to track streams of research. Given the large number of medical journals that are/have been published and the high cost of each journal, local libraries carry a very small selection of what may be seen by the funding community to be a highly specialized item.

If the journals are available, their readability is still problematic for most nonmedical professionals. As with most publications, medical journals are targeted to a particular audience. The target audience does not typically enjoy having to wade through the same introductory and historical information for each successive article in a stream of research. It is also a waste of publication space to use excessive verbiage when there are specific terms, phrases, and acronyms widely used in the field of specialty that convey the intended meaning quickly, succinctly, and with far greater alacrity. In medical writing, authors can assume their target audience has a college education, years of medical school, and for many medical journals, specialized training in a medical sub-field. Readability for the target audience is usually not considered a problem; readability for nonmedical professionals is not a priority.

The issue of providing information concerning new treatments or medical techniques was presumed to be the purview of the medical professional. Thus, until the latter decades of the 20th century, the dissemination of medical treatments and advances had not changed significantly from the dissemination methodologies of the middle ages (Majno, 1975; Osler, 1921; Robinson, 1931). The general forms of the traditional ethical issues facing the medical profession included:

- **Traditional Dilemma 1:** How much time to spend on learning new medical techniques (since this was time that could otherwise be spent with patients); and
- **Traditional Dilemma 2:** Whether or not to discuss a medical innovation with a patient if the medical professional was not proficient with that innovation, or if that technology or innovation was not available locally.

Traditional Medical Information Access

The primary conduit of medical information to the public is the medical practitioner (Majno, 1975). From the earliest medical discoveries of tree bark for soothing aches and pains to using leeches to remove "bad humors," medical practitioners have been the repositories of knowledge on the most effective means of treating injury, regaining health, and reducing pain (Majno, 1975; Marti-Ibanez, 1961, 1962). Once medical books and journals became practical, doctors became the local repository of medical knowledge. The size of a medical provider's private medical library could dictate the breadth and depth of treatment available in a locale (Majno, 1975). New discoveries were made known to the local medical practitioner, who dispensed this information to the general populace, if needed. Otherwise, the medical practitioner simply performed the procedures as recommended.

In the 19th and 20th centuries, continuing education, conferences, seminars, and medical journals provided doctors and nurses with information on approved, state-of-the-art medical treatments (Felch & Scanlon, 1997). Information was also made available by medical device and pharmaceutical manufacturers concerning the specific implementation and formulations of their new medical treatment technologies and techniques. Patients with questions concerning the best treatments for their conditions simply queried their medical practitioners; the answers were rarely questioned (Majno, 1975).

MODERN MEDICAL INFORMATION DISSEMINATION

Television and the Late 20th Century

Mass media discussions of medical issues, techniques, and technologies added potential ethical concerns for the medical professional. These media made it feasible for patients to at least hear of new techniques and technologies through channels that were not under the control of the medical professional. However, in-depth discussion of the conditions under which a particular technique or innovation was optimal, side effects, interaction effects with other treatments or medications, and conditions that would eliminate a patient from being a candidate for the treatment were still left to the physician (Friedman, Furberg, & DeMets, 1998; Spilker & Cramer, 1992). Additionally, government control or licensing of mass media outlets meant that techniques, technologies, medications, and innovations that had not been approved by a particular nation's medical establishment or government regulatory body were unlikely to gain significant airtime for discussion in that country's media.

The proliferation of cable television channels dedicated to news, science, and medicine in the late 1980s began to change the dynamic of medical information dissemination. Shows could dedicate more time to discussing medical problems, symptoms, and treatments. Consequently, patients could ask more probing questions concerning treatment options in consultations with their physicians. The practical effect of such channels of information dissemination was to reduce the frequency of the occurrence of Traditional Dilemma 2. Interested patients of this era were rapidly reducing the information gap between patient and medical provider. Traditional Dilemma 2 became less problematic as physicians were increasingly bombarded with questions from patients referring to research studies and treatment regimens seen on television shows.

The question for physicians was not whether to broach the subject of certain treatment options, but rather, how to respond effectively to patient queries. A companion effect was that physicians were also forced to keep abreast of the types of studies and medical treatment options about which patients were asking questions, if for no other reason than to be able to adequately respond to patient inquiries (Felch & Scanlon, 1997; Pennachio, 2003; Shaneyfelt, 2001).

The Internet and the World Wide Web

The 1990s witnessed a significant change to the patient/provider dyad: widespread access to the Internet. The Internet solved the availability problem for patients by providing global access to medical information. Patients could access information on innovations without having the information limited by finances (personal or regional) or filtered by their medical provider, a media censor, or their national government.

Patients with Internet access were potentially one degree of freedom closer to the source of new medical innovations. Individuals could decide which innovations interested them and seek additional information. Individuals still faced the task of comprehending the information that they uncovered. The Internet, however, made it possible for interested parties to seek out multiple presentations of the same material with relative ease. This increased the potential for the dedicated information seeker to find a presentation of the material that was comprehendible, in accordance with the individual's level of subject and domain knowledge.

Medical Information Without Borders

Information available online is not limited by geopolitical boundaries. Patients can find information on treatments, pharmaceuticals, technologies, and innovations that have not been vetted or approved by their medical establishment or national government. E-mail, chat and PC-based video conferencing potentially allow direct access to medical professionals in other countries. These practitioners may provide an in-depth analysis of a particular medical treatment that is not available to the patient from traditional local information sources (e.g., television or radio shows). Medical providers face patients well versed in the efficacy, side effects, and potential benefits of a wide range of treatment options that may not be legally available in a particular locale, or about which the medical provider may have little knowledge or expertise.

Internet access has provided patients participating in medical experiments with an economical means to find and communicate with other experiment participants. While such actions may seem farfetched, consider the motivation many patients with serious ailments have to seek out information and develop a support network. These individuals are often in pain, scared, and desperate to find a cure for their illness. The ease with which USENET news groups (or threads within an existing group) can be created, combined with the ease of setting up IRC chats or Web-based chat rooms, makes contacting others in real time (or near real time) relatively simple.

Individuals participating in an experiment for a new drug, technique, or technology can now discuss their treatment and its results directly with other subjects in the experiment. This allows individuals to potentially discern if they are receiving the same benefit as others in an experiment. Participating patients who find that a different treatment is providing better results can go to their physicians or the experimenter and demand that other treatment. Information on experimental treatments and clinical trials is readily available. Patients interested in an experimental treatment can access CenterWatch at www.centerwatch.com, which provides access to information related to the clinical testing of drugs and recent FDA-approved drug therapies.

DOCTOR'S DILEMMA

Direct, and near global, access to medical information allows patients to petition their medical practitioners for pain-reducing or lifesaving treatments approved in other countries but not approved in their country of residence. Patients may now share and review information concerning treatment availability and costs from around the world. A doctor's local patients can compare their progress with that of patients in other nations taking different treatments or who are involved in medical experiments. Patients can compare symptoms and try to determine if their situation is the same as (or relatively close to) that of another sufferer.

Widespread information access on the part of patients has created ethical dilemmas for medical practitioners based on the fact that:

1. there may be a better (more effective or less painful) treatment available than the medical professional is currently offering;
2. regardless of the efficacy of the treatment identified by a patient, the medical professional may be prohibited from prescribing

it to, and possibly even from discussing it with, her patient;
3. there may be no clinical trials on which a medical professional can base an opinion as to the efficacy of the treatment the patient is seeking;
4. there may be no means for patients to avail themselves of the treatment they are seeking, even if the medical professional confirms that it is a better treatment than that which is currently available locally; and
5. patients involved in experiments may discover that they are receiving the placebo treatment, while ferreting out data indicating that the experimental treatment is providing relief to other participants.

The issues arising from a reduction in information asymmetry (RIA) can be generally categorized into the following ethical dilemmas for the health care provider:

- **RIA Dilemma 1:** Whether to address the patient's question;
- **RIA Dilemma 2:** Whether to provide the treatment, in possible violation of the law;
- **RIA Dilemma 3:** Whether to place the patient in contact with a medical practitioner or source for the treatment being sought;
- **RIA Dilemma 4:** Whether to petition the experimenters to provide the patient with access to the experimental treatment once the patient has made an explicit request.

Locally Unapproved Medical Treatments

National laws and approval bodies guide the approval of medical treatments, and medical practitioners are generally not permitted to prescribe treatment regimens or drug therapies that have not been approved by their governing bodies. This restriction, in part, helps to ensure that medical professionals are properly trained in the use, effects, and side effects of a particular treatment. It also helps to assure that the treatment regimens and drugs are efficacious and reasonably free of undesirable or unintended side effects.

In *RIA Dilemma 1*, the medical professional must decide if he or she will address a patient's concern. Should the medical professional address the patient's questions concerning unapproved medical treatments, the dyad may be venturing into territory proscribed by law. Refusing to answer the question leaves the patient without the knowledge to decide for him/herself whether to investigate the matter further. Indeed, not addressing the question may discourage the patient from pursuing a course of treatment that may provide him or her with relief or a cure. If the patient has the resources, he or she may travel to where the treatment is provided. If consultation with a medical professional after arrival contraindicates the sought after treatment, the patient would have wasted time and money that could have been better spent on the current treatment.

Providing an unapproved treatment to the patient may result in the medical professional being subject to censure, jail, or loss of medical license. In some rural areas, the loss of a single medical professional would deprive one or more communities of access to critical medical services (Hassinger, 1982). Thus, the medical provider is faced with the dilemma of helping one patient and possibly depriving a community of needed medical care.

Providing a patient with a treatment that is less efficacious than the best available may violate the Hippocratic philosophy of "First Do No Harm" (Majno, 1975; Marti-Ibanez, 1962; Osler, 1921). The physician's deliberations on whether to provide patients access to locally unapproved treatment options necessarily revolve around this oath. Knowingly denying a patient possible access to a treatment believed to be efficacious may constitute a violation of the Hippocratic Oath. On the other hand, governing medical bodies and governments have created protocols and procedures with the intention of protecting the populace from "snake oil" remedies (Anonymous, 2003a), with attendant penalties for willful violation. If the physician chooses to prescribe an unapproved drug therapy, for example, he or she may be unwittingly placing a patient at risk for serious medical complications while concurrently placing his or her medical license in jeopardy.

Locally Unavailable Medical Treatments

Medical treatments may be approved for distribution in a nation or region but may not necessarily be made widely available to a specific locale, leading to *RIA Dilemma 2*. This may be for commercial reasons, or it may be the result of government rationing (Morgan, 2003; Schwartz, Morrison, & Sullivan, 1999). In countries with nationalized healthcare, it may not be economically feasible or politically desirable to provide all treatments in all locations (Morgan, 2003). Thus, a treatment that is nationally approved may not be accessible in a given region (Schwartz et al., 1999).

Direct access to medical information from other regions and countries allows patients to petition for approved pain-reducing or lifesaving treatments that are not available in their area. Patients may now share and review information concerning treatments in other regions, and become aware of treatments being denied them on the basis of geography (Morgan, 2003; Schwartz et al., 1999).

Should the medical professional address the patient's questions, he or she may be venturing into territory proscribed by law, especially in the case of nationalized healthcare (Schwartz et al., 1999). To not answer the question is to leave the patient without the knowledge to decide for herself whether to investigate the matter further. Indeed, not addressing the question may discourage the patient from pursuing a course of treatment that may provide her with relief or a cure.

A patient may seek a recommendation from his or her local practitioner for a contact in another locale for a consultation, resulting in the practitioner facing *RIA Dilemma 3.* This can also have legal ramifications. Medical personnel may be prohibited from recommending persons not approved in their country to perform medical treatments.

Locally More Expensive Treatments

If a less expensive but still efficacious treatment is available, the medical practitioner is likely to inform the patient, as long as the treatment is governmentally authorized. If the less expensive alternative is not authorized for local patients, the medical practitioner is not likely to volunteer information on that treatment. However, should the patient raise the issue of the treatment, the medical provider is forced into a position of discussing an unauthorized treatment or leaving the patient in ignorance.

Prior to patient access to transnational information on treatment pricing, treatment cost differentials were not a major ethical issue for medical practitioners, since the patient was unlikely to be knowledgeable enough to broach the topic. If a less expensive alternative to the current treatment was not available in his or her country, the medical professional was under no obligation to mention the treatment, so *RIA Dilemma 1* was not an issue. With direct access to information concerning pricing worldwide, patients can now request information concerning specific alternatives to their current treatment. Patients may be able to find pricing information, but they may still have questions on efficacy, side effects, or interactions with current treatments.

An example illustrates this point. The presence of pharmaceuticals in Canada that are less expensive than their US counterparts is not an unheard of occurrence. In some cases, pharmaceuticals formulations used in Canada are identical to their US equivalents and are manufactured in the US.

Doctors and pharmacists are placed in the position of having to decide whether to respond to their patients' or customers' inquiries about the ability to purchase the same items for less via the Internet or by traveling to Canada. For patients with fixed retirement incomes facing rapidly increasing prescription drug prices on multiple prescriptions, this may be a life-or-death decision. The practitioner is faced with the same problems found in the situation of locally unavailable treatments: he or she could be subject to disciplinary action or potential loss of medical license, if he or she is found to be recommending locally unapproved pharmaceuticals.

Unaffordable Medical Treatments

Cost is a concern for many medical providers and hospitals and is often a determinant of the medical treatment made available to patients. Cost containment is an issue common to both nationalized and privatized healthcare systems (Barry & Raworth, 2002). Some countries and medical establishments simply cannot afford to provide very expensive treatments (Barry & Raworth, 2002).

The dilemma for medical practitioners is whether to inform the patient of treatments that cannot be afforded or leave the patient in ignorance (Barry & Raworth, 2002; Gert, 2002). The former option provides the patient with the information on an alternative treatment, but also the pyrrhic knowledge that his or her suffering may be in vain because it could be alleviated — if the nation or healthcare facility could afford the treatment or technology (Barry & Raworth, 2002). The latter option denies the patient the opportunity to travel to a locale where the treatment is provided, if personal resources permit (Gert, 2002).

Medical Experiments: Request for a Specific Treatment

The final situation that we will discuss concerns medical experiments and *RIA Dilemma 4*, and has been alluded to already. Traditionally, scientifically valid experiments require a control group that received a placebo (e.g., the classic "sugar pill") treatment, and an experimental group that received the treatment to be tested (Matthews, 2000; Spilker & Cramer, 1992). There may also have been a third group that received the current treatment, although in situations where removing a patient from, or denying him access to, the traditional treatment may be deemed too detrimental to the patient, the placebo is often the currently approved treatment (Barber, 1980; Cowles, 1976; Matthews, 2000; Spilker & Cramer, 1992).

Medical practitioners treating patients with a particular condition were recruited to invite their affected patients to participate in the experiment. To ensure that medical practitioners did not inadvertently influence the patient, participating medical practitioners were usually not informed whether their patient was a member of the control group or the experimental group (Barber, 1980; Cowles, 1976; Matthews, 2000; Spilker & Cramer, 1992).

Access to the Internet and chat rooms has altered this equation. Patients now have the potential to discover that an experiment is underway or planned and request that their practitioner volunteer them for the experiment. The medical practitioner must then seek out the experimenter(s) to request permission for his or her patient to be included in the experiment. This could have the effect of excluding a patient who would be a better candidate for the experiment, but who did not know to request inclusion (Barber, 1980; Spilker & Cramer, 1992).

Internet access can also provide patients with a forum to discuss their participation in experiments. Thus, it is now possible for patients to talk among themselves, compare treatment results, and determine the success of treatments using resources such as www.spinalcord.org. By comparing results, it is possible for patients to deduce whether they are in the control group and, possibly, the efficacy of the experimental treatment.

If the experimental treatment appears successful, patients in the control group can ask their physicians for the new treatment. The medical provider is placed in the position of asking the experimenter to provide the new treatment. As a consequence, the experimenter is thrust into the position of denying the experimental treatment to a subject who knows he or she is receiving the placebo. The subject who deduces that he or she is in the placebo group is no longer valid for data analysis, because the placebo effect is destroyed; continuing with the ruse serves no useful scientific purpose.

If the experimental treatment is not helping the experimental group, patients in that group may request that their medical practitioners remove them from the experiment, and return them to the existing treatments (Friedman et al., 1998; Spilker & Cramer, 1992). Since informed consent is the cornerstone of all enlightened medical research, and since that consent can be rescinded at any time, once the physician reports the patient's feelings to the experimenter, the experimenter must release him or her from the experiment (Barber, 1980; Cowles, 1976). Thus, new medical discoveries could be imperiled.

FUTURE TRENDS

An ethics discussion of the factors affecting medical professionals frees us from the constraint common to economic discussions of the topic: the need for equilibrium. As noted by Milgrom and Roberts (2001), "Almost all of economic theory is equilibrium analysis…predictions arise only once equilibrium behavior is assumed." The ethical dilemmas facing medical professionals increase as patients gain access to information, not simply once they reach information parity with their medical practitioners. Indeed, the ethical dilemmas for the medical provider continue, and may in fact increase, if a patient should surpass his or her knowledge of a specific topic, disease or treatment.

For most dyadic interactions, a reduction in the imbalance of information between parties will reduce the ethical dilemmas related to the interaction. The increasing availability and readability of information concerning new medical advances has produced the opposite effect. For medical professionals, the decrease in the traditional information imbalance between patients and medical providers has increased the number and nature of ethical dilemmas faced.

Serious illness or chronic pain provides intense motivation to pursue information on a particular affliction or condition with single-minded dedication. While a medical professional may need to track the symptoms and diseases of dozens or hundreds of patients, a particular patient can concentrate on a specific problem. Thus, a patient may be more knowledgeable about the current discoveries and findings regarding a specific problem than the medical practitioner. With the Internet providing an efficient conduit for researching a problem, a patient may be motivated to explore all available information concerning his or her affliction. Since the Internet is borderless, this research can uncover the most current medical findings worldwide.

This chapter focuses on moral dilemmas brought about through the continued spread of information access. While other chapters in this book focus on threats to privacy and intellectual property, it is important to note that there can be ethical threats engendered in the very availability of the information itself. As ever-greater volumes of information migrate online, it is possible, even likely, that other arenas of human interaction may face similar quandaries. An examination of the current issues may help prepare for a discussion of these future imbroglios.

Concomitant with such developments, the world's citizens should be prepared to deal with their governments' attempts to mitigate the effects of the ethical quandaries for which this expanded information availability is the genesis. Existing laws governing medical devices, treatments, and technologies (e.g., regulation of locally unapproved drugs) suggest that governments will likely take steps to deal with perceived problems arising from greater information flow. There is every reason to expect that regulations will be developed to counter the problems governments perceive in other arenas as well. A lively ethical debate can facilitate the development of well-reasoned laws and regulations.

CHAPTER SUMMARY

National laws are based on geography: the physical limitations of jurisdiction. As with other laws in the Internet age, laws regarding the actions of medical professionals within national borders are facing the reality of borderless information access. The Internet purchase of less expensive prescription drugs in Canada by citizens in the United States is but one example of physical laws not keeping pace with the telecommunications revolution.

This chapter is designed to illuminate the issues surrounding an interesting problem with information parity vs. asymmetry in Medicine. The changing dynamics between patient and caregiver are intricate and complex, especially with the omnipresent impact of laws governing medical treatment. The impact of RIA, combined with the presence of information sources beyond the control of a single national government, complicates the ethical decisions for medical practitioners geometrically. No easy solutions are possible to this problem, and this chapter does not purport to resolve such complex issues. As desirable as a solution to these problems is, a workable solution is not practical at this juncture due to the interplay among technology, law, and ethics.

A technological solution is not tenable since it is the technology that leads to many of the ethical dilemmas facing medical practitioners: new technology may in fact introduce additional complexity in medical relationships. A legal solution is equally fraught with difficulty. National laws currently exist in many countries that are intended to address medical issues. However, the borderless society created by modern information technologies may require an international solution that addresses the realities of physician responsibility and modern data interchange. It may be impossible, however, to un-ring the bell of borderless medical information flow. The reality is that the caregiver is ultimately responsible for the ethical quandaries posed, but that researchers, politicians, and ethicists can assist by fostering an open discussion of the moral calculus.

Finally, this inverse relationship between information parity and ethical dilemmas is of concern because it may not be a singular exception to the generally held belief that information parity decreases ethical dilemmas. As high-speed Internet access becomes more pervasive in society, other situations may develop in which information that becomes available to the general public generates an increase in ethical quandaries. Future research should explore the social impact of unfiltered information becoming available in other knowledge domains. Without awareness of this possibility, ethical training and legal debate may be unprepared to confront a potent, counterintuitive, social phenomenon. It is the nature of such issues that they do not lend themselves to formulaic solutions, but require deliberations and debate to find workable options. It is our goal to begin to foster such debate with this chapter.

REFERENCES

Aguolu, I. E. (1997). Accessibility of information: A myth for developing countries? *New Library World, 98*(1132), 25.

Akerlof, G. A. (1970). The market for “lemons”: Quality uncertainty and the market mechanism. *Quarterly Journal of Economics, 84*(3), 788-500.

Anonymous. (1998). What is “insider trading”? U.S. Securities and Exchange Commission [Web page]. Modified: 09/01/1998. Retrieved October 13, 2003 from: *http://www.sec.gov/divisions/enforce/insider.htm*

Anonymous. (2003a). Approvals of FDA-regulated products. US Food and Drug Administration. [Web page]. Retrieved March 30, 2004 from: *http://www.fda.gov/opacom/7approvl.html*

Anonymous. (2003b). The laws that govern the securities industry. [Web Page]. U. S. Securities and Exchange Commission. [Web Page]. Modified: 06/27/2003. Retrieved from: *http://www.sec.gov/about/laws.shtml*

Barber, B. (1980). *Informed consent in medical therapy and research*. New Brunswick, NJ: Rutgers University Press.

Barry, C., & Raworth, K. (2002). Access to medicines and the rhetoric of responsibility. *Ethics & International Affairs, 16*(2), 57(15).

Borgman, C. L. (2000). *From Guttenberg to the global information infrastructure: Access to information in the networked world*. Cambridge, MA: MIT Press.

Cowles, J. (1976). *Informed consent*. New York: Coward, McCann & Geoghegan, Inc.

Davies, C., & Beach, A. (2000). *Interpreting professional self-regulation*. London: Routledge Taylor & Francis Group.

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *The Review of Economic Studies*, *51*(3), 393-414.

Digby, A. (1997). The patient’s view. In I. Loudon (Ed.), *Western medicine* (pp. 291-305). Oxford, UK: Oxford University Press.

Felch, W. C., & Scanlon, D. M. (1997). Bridging the gap between research and practice: The role of continuing medical education. *JAMA, 277*(2), 155(152).

Fournier, G. M., & McInnes, M. M. (1997). Medical board regulation of physician licensure: Is excessive malpractice sanctioned? *Journal of Regulatory Economics, 12*(2), 113-126.

Friedman, L. M., Furberg, C. D., & DeMets, D. L. (1998). *Fundamentals of clinical trials* (3rd ed.). New York: Springer-Verlag.

Gert, H. J. (2002). Avoiding surprises: A model for informing patients. *The Hastings Center Report, 32*(5), 23(11).

Hassinger, E. W. (1982). *Rural health organization*. Ames, IA: Iowa State University Press.

Hellwig, M. F. (2001). Risk aversion and incentive compatibility with *ex post* information asymmetry. *Economic Theory, 18*, 415-438.

Holmstrom, B. R. (1985). The provision of services in a market economy. In R. P. Inman (Ed.), *Managing the service economy: Prospects and problems*. Cambridge, UK: Cambridge University Press.

Magner, L. N. (1992). *A history of medicine*. New York: Marcel Dekker.

Majno, G. (1975). *The healing hand: Man and wound in the ancient world*. Cambridge, MA: Harvard University Press.

Marks, F. R., & Cathcart, D. (1974). Discipline within the legal profession: Is it self-regulation? *University of Illinois Law Review, 193*(Spring), 236.

Marti-Ibanez, F. (1958). *Centaur: Essays on the history of medical ideas*. New York: MD Publications.

Marti-Ibanez, F. (1961). *A prelude to medical history*. New York: MD Publications.

Marti-Ibanez, F. (1962). *The epic of medicine*. New York: Clarkson N. Potter.

Matthews, J. N. S. (2000). *An introduction to randomized controlled clinical trials*. London: Arnold, A member of the Hodder Headline Group.

Milgrom, P., & Roberts, J. (2001). Informational asymmetries, strategic behavior, and industrial organization. *Game Theory and Industrial Organization*, *77*(2), 184-193.

Morgan, J. M. (2003). New device indications: Practice and cost implications in Europe. *Cardiac Electrophysiology Review*, *7*(1), 49-53.

Nayyar, P. R. (1990, November/December). Information asymmetries: A source of competitive advantage for diversified service firms. *Strategic Management Journal*, *11*, 513-819.

Osler, S. W. (1921). *The evolution of modern medicine*. New Haven, CT: Yale University Press.

Payne, R. (2003). Informed trade in spot foreign exchange markets: An empirical investigation. *Journal of International Economics*, *61*, 307-329.

Pennachio, D. L. (2003). CME: How to get yours? Requirements are multiplying, but so are the opportunities to meet them. *Medical Economics*, *80*(16), 21(23).

Porter, R. (1992). The patient in England, c. 1660-c. 1800. In A.Wear (Ed.), *Medicine in society*. Cambridge, UK: Cambridge University Press.

Robinson, V. (1931). *The story of medicine*. New York: Tudor Publishing Co.

Schwartz, L., Morrison, J., & Sullivan, F. (1999). Rationing decisions: From diversity to consensus. *Healthcare Analysis*, *7*(2), 195-205.

Shaneyfelt, T. M. (2001). Building bridges to quality. *JAMA*, *286*(20), 2600(2602).

Spilker, B., & Cramer, H. A. (1992). *Patient recruitment in clinical trials*. New York: Raven Press.

Summers, A. (1997). Nurses and ancillaries in the Christian era. In I. Loudon (Ed.), *Western medicine* (pp. 192-205). Oxford, UK: Oxford University Press.

This work was previously published in Information Ethics: Privacy and Intellectual Property, edited by L. Freeman & A.G. Peace, pp. 197-218, copyright 2005 by Information Science Publishing (an imprint of IGI Global).

Chapter 5.3
How to Start or Improve a KM System in a Hospital or Healthcare Organization

A.H. Rubenstein
IASTA Inc., USA

E. Geisler
Illinois Institute of Technology, USA

ABSTRACT

One of the key factors that distinguishes enterprises of the 21st Century is the emphasis on knowledge and information. Knowledge management is an important means by which organizations can better manage information, and more importantly, knowledge. Unlike other techniques, knowledge management is not always easy to define, because it encompasses a range of concepts, management tasks, technologies, and practices, all of which come under the umbrella of the knowledge management. This chapter deals with two aspects of knowledge management systems: (a) why KM systems are needed, and (b) how to get started on designing and rolling out a new or improved KM system. The inferences are drawn from the direct experiences the authors have had during their academic and consulting activities in many health sector organizations.

INTRODUCTION

Hospitals and other healthcare organization need bricks and mortar, human resources, medical technology, financial systems, and other infrastructure items. In addition, they need effective ways of capturing, preserving, transforming, retrieving, and applying *knowledge* about past experiences and subsequent "lessons learned" to *current* and *future* needs. These are commonly called *knowledge management (KM)* systems.

Here are some brief definitions:

1. *KM* is the management of intellectual capital (IC) in the interests of the enterprise.
2. *IC* includes knowledge, information, lessons learned, and other data held for use over time.

3. Intellectual nuggets (*nuggets*) are the units that carry knowledge, lessons learned, etc.
4. *Super-nuggets* are items tied specifically to express or implied needs of the organization.
5. The most valuable structure of a nugget (N) is a compound statement which has an "If x, ...then y" format: If you do this in a certain way, ... then you can achieve or avoid the following consequence(s).
6. Some nuggets are, in fact, *causal statements*, but many others only suggest strong *correlations* or *influence* of actions on outcomes (consequences).

QUESTIONS SURROUNDING KM SYSTEMS

Why does a hospital or healthcare organization need a new or improved KM system?

- To bridge knowledge "*silos*" in specialties and support functions.
- To learn from its own experiences and experiences of other organizations.
- To avoid repeating or duplicating major or even disastrous mistakes in, for example:
  - disease management;
  - infection control;
  - misuse or abuse of instruments;
  - duplication of expensive equipment.
- To support training at all levels.
- To support weaker or resource-poor units with experience from stronger or resource rich units.
- To share "tricks of the trade."
- To help avoid pitfalls in:
  - organizational design;
  - staffing;
  - work flow.
- To exchange methods of:
  - productivity improvement;
  - cost savings;
  - patient services.

Does everyone need a KM system?

- In general, *yes*, to help: organize, preserve, retrieve, and use lessons learned and other nuggets about his/her job, and the internal and external environment.
- However, there are vast differences among individuals and individual information needs and style — some people still pine for the all-but-obsolete card catalog in the public library. Others have gone "all electronic" and will never look back.
- There are also differences in the relevance and value of past experience or the experience of others.
- Although it is hard to imagine or accept, people in responsible positions may be completely ignoring lessons from the past.
- It is important that the organizational KM system is able to accommodate a wide range of individual differences and bring lessons learned across organizational lines.

Many areas of professional and staff practice have *common elements* where experience can be fruitfully codified and exchanged. Such commonalities can transcend narrow medical, healthcare, and administrative specialties. KM systems can also help transcend temporal and spatial distances in the organization and between organizations. Two such areas are: managing medical technology and medical and staff errors.

Managing Medical Technology

Application areas include computerized patient records, telemedicine, and improved utilization and management of equipment. These all provide opportunities for sharing knowledge and experience, but the pace of such sharing lags far behind the leading edge in industry and the evolving state of the art. Progress is slow toward achieving ease of use, standardization, and connectivity between databases and different record systems. A key challenge is to incorporate existing and new knowledge systems into integrated and comprehensive systems that can be used effectively and cooperatively by caregivers, administrators, payors, and regulators.

Medical Errors

Dealing with medical errors systematically requires improved codification, increased sharing, e.g., the "mortality and morbidity" sessions conducted in hospitals, but generally not recorded and disseminated.

How to Get Started

- Who should initiate a new or redesigned KM system — the role of "champions."
- The design and rollout teams — size, skills, experience, knowledge of the organization.
- How to keep it simple.
- How to get started in a modest way — not across the whole organization at once.
- Possible modest field experiments to test ideas and components of the system.

The champion should be a sponsor at a high enough level to provide credibility as well as material support — e.g., a CEO, other top manager, an operating division manager. The team should include three to six people with: (1) technical understanding of information and communication systems, (2) credibility and experience within the organization, and (3) good people skills.

How to Keep it Simple

- Don't start with a "greenfields" approach.
- Utilize existing components where feasible. All healthcare organizations have many existing information systems that might be adapted.
- Get something useful going quickly.
- Do not ignore possibly useful legacy systems.
- Design the size and "shape" to fit the existing organization — not an ideal one.
- Don't start on too broad a front.
- Select a functional or operating area with high success potential for KM.
- Design a fast prototype that can be readily rolled out to demonstrate the KM principle.

How to Get Started

- Identify information/knowledge needs and styles of key individuals and groups of potential users.
- Examine the current, if any, versions of KM in use or being developed.
- Develop metrics for assessing:
  - The operation of the KM system:
  - The impacts — immediate or direct, intermediate, and ultimate.
- Design and test a systematic procedure for the extraction and refinement of knowledge "nuggets."
- Design and test a monitoring and assessment procedure for system usage and effectiveness.

Some methodologies we have used for designing systems from the viewpoint of the users include:

1. Personal communication charting and communication flow matrix.
2. Individual information-seeking propensity profile.
3. Critical incident method — Effective and ineffective use of the system.

4. Black hole analysis — Where do key pieces of information disappear to?
5. Gatekeeper analysis — Who are the key intermediaries in searching for information?
6. Knowledge transfer and application mapping — What happened to a nugget or super-nugget in terms of application?
7. Individual and group risk-taking propensities.
8. Blame-placing behavior.
9. Bureaucratic analysis and the role of organizational politics.
10. Indicators for statistically out-of-control behaviors.
11. As to KM technology, adequate technology is in hand or on the way. The key problems in designing, installing, using, and management of KM systems are primarily human (user behavior) and organizational (politics, culture, and communication).

SOURCES

Davenport, T. & Prusak, L. (1997). *Working Knowledge: How Organizations Manage What They Know*. Boston, MA: Harvard Business School Press.

Dixon, N. (2000). Common *Knowledge: How Companies Thrive by Sharing What They Know.* Boston, MA: Harvard Business School Press.

Geisler, E. (1999). Mapping the Knowledge-Base of Management of Medical Technology. *International Journal of Healthcare Technology and Management*, 1(1), 3-10.

Geisler, E., Krabbendam, K. & Schuring, R. (2003). *Technology, Healthcare and Management in the Hospital of the Future*. Westport, CT: Greenwood Publishing Group.

Martin, W. (2000). Approaches to the Measurement of the Impact of Knowledge Management Programs. *Journal of Information Science*, 26(1), 21-27.

Nonaka, I. & Nishiguchi, T. (eds.) (2001). *Knowledge Emergence: Social, Technical and Evolutionary Dimensions of Knowledge*. New York: Oxford University Press.

Rapport, M. (2001). Unfolding Knowledge. *Knowledge Management*, 4(7), 44-49.

Rubenstein, A. H. (1989). *Managing Technology in the Decentralized Firm*. New York: John Wiley & Sons.

Rubenstein, A. H. (2001). Knowledge-Management Systems: Effective Rollout and Operation to Meet User Needs. *Annual Review of Communications, International Engineering Consortium*, 54, 1-9.

Rubenstein, A. H. & Geisler, E. (1992). Intelligent Support Systems: An Analysis of User Needs. *Siemens Review*, 6, 32-36.

Rubenstein, A. H. & Geisler, E. (2003). *Installing and Managing Workable Management Systems.* Westport, CT: Praeger.

Rubenstein, A. H. & Schwaertzel, H. (eds.) (1992). *Intelligent Workstations for Professionals.* New York: Springer Verlag.

This work was previously published in Creating Knowledge-Based Healthcare Organizations, edited by N. Wickramasinghe, J.N.D. Gupta & S. Sharma, pp. 44-49, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 5.4
Patient Centric Healthcare Information Systems in the U.S.

Nilmini Wickramasinghe
Illinois Institute of Technology, USA

INTRODUCTION

Healthcare expenditure is increasing exponentially, and reducing this expenditure (i.e., offering effective and efficient quality healthcare treatment) is becoming a priority not only in the United States, but also globally (Bush, 2004; Oslo Declaration, 2003;Global Medical Forum, 2005). In the final report compiled by the Committee on the Quality of Healthcare in America (Institute of Medicine, 2001), it was noted that improving patient care is integrally linked to providing high quality healthcare. Furthermore, in order to achieve high quality healthcare, the committee has identified six key aims, that is, healthcare should be:

1. **Safe:** avoiding injuries to patients from the care that is intended to help them
2. **Effective:** providing services based on scientific knowledge to all who could benefit, and refraining from providing services to those who will not benefit (i.e., avoiding under use and overuse)
3. **Patient centered:** providing care that is respectful of and responsive to individual patient preferences, needs, and values, and ensuring that patient values guide all clinical decisions
4. **Timely:** reducing waiting and sometimes harmful delays for both those receiving care and those who give care
5. **Efficient:** avoiding waste
6. **Equitable:** providing care that does not vary in quality based on personal characteristics

Most of the poor quality connected with healthcare—such as loss of information or incomplete information pertaining to patient medical records, allergic reactions that can be life threatening, or the ordering of wrong tests—is related to a highly fragmented delivery system that lacks even rudimentary clinical information capabilities resulting in inadequate information flows and poorly designed care processes characterized by unnecessary duplication of services, long waiting times, and delays (Institute of Medicine, 2001; Chandra, Knickrehm, & Miller, 1995). In addition, poor information quality is also a major contributor to the numerous medical errors that permeate throughout the system (Mandke, Bariff, & Nayar, 2003). The introduction of the Health Insurance Portability and Accountability Act (HIPAA, 2001) in the United States into this context only makes matters more complex, since it imposes a further level of convolution to the design and management of information and its flows throughout the healthcare system. The aims of HIPAA are indeed laudable, since they focus on establishing better governance structures and compliance so that healthcare information can be protected and secured; however, in practice, given the current platform-centric nature of healthcare organizations, this only serves to create further informational challenges.

Healthcare is noted for using leading-edge technologies and embracing new scientific discoveries to enable better cures for diseases and better means to enable early detection of most life-threatening diseases (Stegwee & Spil, 2001; McGee, 1997; Johns, 1997; Wallace, 1997). However, the healthcare industry has been extremely slow to adopt and then maximize the full potential of technologies that focus on better practice management and administrative needs (Stegwee & Spil, 2001). In the current complex healthcare environment, the development and application of sophisticated patient-centric healthcare systems and e-health initiatives are becoming strategic necessities, yet healthcare delivery has been relatively untouched by the revolution of information technology (Institute of Medicine, 2001; Wickramasinghe, 2000; Wickramasinghe & Mills, 2001; Stegwee & Spil, 2001; Wickramasinghe & Silvers, 2002). To address this dilemma, healthcare organizations globally require a systematic methodology to guide the design and management of their respective IC^2T adoptions, not only to be compliant with regulations like HIPAA but also to be able to capture, generate, and disseminate information that is of high integrity and quality, and thereby be both technically sound and meet the highest ethical and security standards. An integrative compliance framework is an appropriate solution strategy.

REGULATORY REQUIREMENTS

In the United States, HIPAA (2001) is the minimum governing regulatory compliance standard to which healthcare organizations must adhere. Essentially similar standards exist in other countries, for example, the EU Directive 46 of 1995 is currently being implemented throughout all EU countries, as well as revisions to this, including privacy law (675/96) (Inchingolo, 2003). These are developed by countries or respective governments within the EU to ensure security and privacy of sensitive patient healthcare information. Irrespective of which policy we look at (HIPAA or the EU Directive), the fundamental areas pertaining to compliance and security of health information are similar. A closer examination of HIPAA reveals three key elements: security, privacy, and standards for electronic submissions and exchange of healthcare information (HIPPA, 2001; Moore & Wesson, 2002).

Security

According to HIPAA, a number of security criteria must be met, not only by the housing of information but also by all electronic healthcare transactions that contain healthcare information. Some of these criteria directly affect how healthcare systems can be accessed as well as how the key healthcare players (governments, providers, payers, and patients) may interact with these systems. The HIPAA security requirements1 focus on:

- establishing trust partnership agreements with all business partners
- instituting formal mechanisms for accessing electronic health records
- establishing procedures and policies to control access of information
- maintaining records of authorizing access to the system
- assuring that system users receive security awareness training, and the training procedures are periodically reviewed and updated
- maintaining security configuration including complete documentation of security plans and procedures, security incident

reporting procedures, and incident recovery procedures;
- ensuring communication and network control, including maintaining message integrity, authenticity, and privacy; encryption of messages is also advocated for the open network transmission portion of the message; and
- authenticating data to ensure it is not altered or destroyed in an unauthorized manner.

The principal security tenets of HIPAA fall into three categories—administrative, physical, and technical—each subdivided into several subcategories consisting of multiple levers (Wickramasinghe & Fadlalla, 2004). Table 1 summarizes the major issues and levels under each of these categories.

Transaction Standards

The standards for electronic health information transactions cover all major transactions, including claims, enrollment, eligibility, payment, and coordination of benefits. HIPAA discusses two major categories with relation to transaction standards: practice standards and technical standards. The key practice standards are:

1. Health Care Common Procedure Coding System (HCPCS)
2. ICD-9–Diagnosis Codes
3. ICD-9–Procedure Codes

The technical standards focus on the adoption of electronic data interchange (EDI) using health care industry implementation guidelines and other standards such as XML and X12. Plans and providers can comply with these standards directly or via a healthcare clearinghouse (Wickramasinghe & Fadlalla, 2004).

Privacy

The final element of HIPAA focuses on ensuring the privacy of healthcare information. Patient healthcare data is sensitive in nature and must be protected from the potential and possibility of misuse or abuse. Specifically, the *Federal Register* (vol. 67, no. 157) details all the rules that must be adhered to with respect to privacy. The purpose of these rules is to maintain strong protection for the privacy of individually identifiable health information. Thus, these privacy requirements cover the use and disclosure of treatment and payment information, while creating a national standard to protect individuals' medical records and other personal health information. Specifically, they aim to:

- give patients more *control* over their health information;
- set *boundaries* on the use and release of health records;
- establish appropriate *safeguards* that healthcare providers and others must achieve to protect the privacy of health information;
- hold violators *accountable,* with civil and criminal penalties that can be imposed if they violate patients' privacy rights; and
- strike a balance when *public responsibility* requires disclosure of some forms of data, for example, to protect public health.

Healthcare providers and organizations must ensure that all their IC^2T initiatives are compliant with all regulatory requirements. For example, in the case of HIPAA in the U.S., non-compliance has major negative implications such as legal action and severe financial penalties. A framework that enables healthcare organizations to design and build HIPAA compliance (or other relevant regulatory compliance) into the IC^2T architecture then becomes an indispensable tool and a critical success factor for any patient-centric healthcare system.

Table 1. HIPAA security issues and key focus areas[2]

Security Issues	Key Areas of Focus
Security Management Process	• Risk Analysis • Risk Management • Sanction Policy • Information System Activity Review
Workforce Security	• Authorization and/or Supervision • Workforce Clearance Procedure • Termination Procedure
Information Access Management	• Isolating Healthcare Clearinghouse Function • Access Authorization • Access Establishment and Modification
Security Awareness and Training	• Security Reminders • Protection from Malicious Software • Log-in Monitoring • Password Management
Security Incident Procedures	• Response and Reporting
Contingency Plan	• Data Backup Plan • Disaster Recovery Plan • Emergency Mode Operation Plan • Testing and Revision Procedure • Applications and Data Criticality Analysis
Business Associate Contract and Other Arrangement	• Written Contract or Other Arrangement
Facility Access Controls	• Contingency Operations • Facility Security Plan • Access Control and Validation Procedure • Maintenance Records
Device and Media Controls	• Disposal • Media Re-Use • Accountability • Data Backup and Storage
Access Control	• Unique User Identification • Emergency Access Procedure • Automatic Logoff • Encryption and Decryption
Audit Controls	
Integrity	• Mechanism to Authenticate Electronic Protected Health Information
Transmission Security	• Integrity Controls • Encryption

Information Integrity and Quality

At the center of the information flows in current healthcare systems is the healthcare information system (HCIS). Not only does the HCIS connect the key healthcare players within the healthcare system in an efficient and effective manner, but it also forms the central repository for critical information such as patient medical records, billing, and treatment details. Hence the HCIS provides the underlying network structure for supporting the information flows which in turn support decision-making activities throughout the healthcare system. Healthcare procedures such as medical diagnostics, treatment decisions, consequences of these decisions, prevention, communication, and equipment usage can be thought of as iatric in nature since they are connected to treating patients (Perper, 1994). Integral to these iatric procedures is the generation and processing of multi-spectral data and information such as patient history, research findings on treatments, and protocols for treatment (Mandke et al., 2003). The patient provides key information at the time of a clinical visit or other interaction with his/her provider. Such a visit also generates other information including insurance information, medical history, and treatment protocols (if applicable) which must satisfy regulatory requirements, payer directives, and the healthcare organization's informational needs. Thus, from a single intervention, many forms and types of information are captured, generated, and then disseminated throughout the healthcare system. All this information and the consequent flows must, in addition to satisfying regulatory requirements such as HIPAA, satisfy some common integrity characteristics such as accuracy, consistency, reliability, completeness, usefulness, usability, and manipulability. In this way the information product will be truly useful and useable while generating a high level of trust and confidence in its content by the user or decision maker. Since the generated healthcare information flows across various departments and/or organizational boundaries, the challenge of ensuring information integrity is further compounded because any integrity problems will propagate with ripple effects following the same trajectory as the information itself. The consequences of poor quality information are multiplied as they propagate throughout the system, thus are not only significant but have far-reaching consequences.

Given the critical role information plays both within and between the many healthcare players, it is imperative that the information flowing both within the HCIS and between these key players must exhibit both the attributes and the dimensions of the information integrity construct, as well as satisfy the six healthcare quality aims (stated earlier).

Information Integrity

Information integrity is an emerging area that is "not just about engineering the right properties of information but it also includes sensitivity to the context in which information is used and the purpose for its usage" (Geisler, Lewis, Nayar, & Prabhaker, 2003, p. 5). More specifically it encompasses the accuracy, consistency, and reliability of the information content, process, and system. By focusing on the privacy, security, and standards aspects of healthcare information, it would appear that HIPAA implicitly assumes certain characteristics of this information product such as its accuracy and reliability. However, in practice this may not always be the case, and from the perspective of the healthcare organization, it is not sufficient to be HIPAA compliant, rather it must also ensure that the information product satisfies the principles of information integrity standards. Hence, the information should display the attributes of accuracy, consistency, and reliability of content and processes, as well as the dimensions of usefulness, completeness, manipulability, and usability (Mandke et al., 2003; Geisler et al., 2003).

Implicit in taking an information integrity perspective is the shift from viewing information as a byproduct to viewing it as an essential product (Huang et al., 1999). This requires following four key principles; namely that the information must (Huang et al., 1999; Mandke et al., 2003; Geisler et al., 2003):

1. meet the consumers information needs
2. be the product of a well-defined information production process
3. be managed by taking a lifecycle approach
4. be managed and continually assessed vis-à-vis the integrity of the processes and the resultant information

Healthcare Quality Aims

In the final report of the Committee on the Quality of Health Care in America (Institute of Medicine, 2001), six key quality aims were identified. They are listed again here for completeness. Healthcare should be:

1. **Safe:** Avoiding injuries to patients from the care that is intended to help them
2. **Effective:** Providing services based on scientific knowledge to all who could benefit, and refraining from providing services to those who will not benefit (i.e., avoiding under use and overuse)
3. **Patient centered:** Providing care that is respectful of and responsive to individual patient preferences, needs, and values, and ensuring that patient values guide all clinical decisions
4. **Timely:** Reducing waiting and sometimes harmful delays for both those receiving care and those who give care
5. **Efficient:** Avoiding waste
6. **Equitable:** Providing care that does not vary in quality based on personal characteristics

These aims can only be negatively impacted by poor information quality, flow, and integrity. Conversely, a higher quality, flow, and integrity of information will positively impact these aims by helping to reduce the large number of medical errors that currently permeate the healthcare system (Mandke et al., 2003; Geisler et al., 2003). This is because more accurate data and information will be entered into the systems, and if erroneous data and/or information is entered, it will be identified quickly and corrected.

What becomes critical then is to incorporate these quality aims into the manufacturing of the information product so that the output is quality information. This requires the establishment of an information quality program which serves to (Huang et al., 1999):

1. articulate an information quality vision in healthcare business terms
2. establish central responsibilities for information quality within the information product manufacturing processes
3. educate the producers and consumers of information on information quality issues
4. institutionalize and continuously evaluate and develop new information quality skills

PROPOSED FRAMEWORK

Healthcare in the United States and globally is facing multidimensional challenges, including structuring efficient and effective healthcare systems that first and foremost deliver value to patients, enable-cost conscious healthcare treatments, and simultaneously support the statutory as well as operational information requirements. IC^2T is becoming a key enabler and strategy necessity for organizations, irrespective of their business sector (Haag, Cummings, & McCubbrey, 2004; Scott Morton, 1991; von Lubitz &

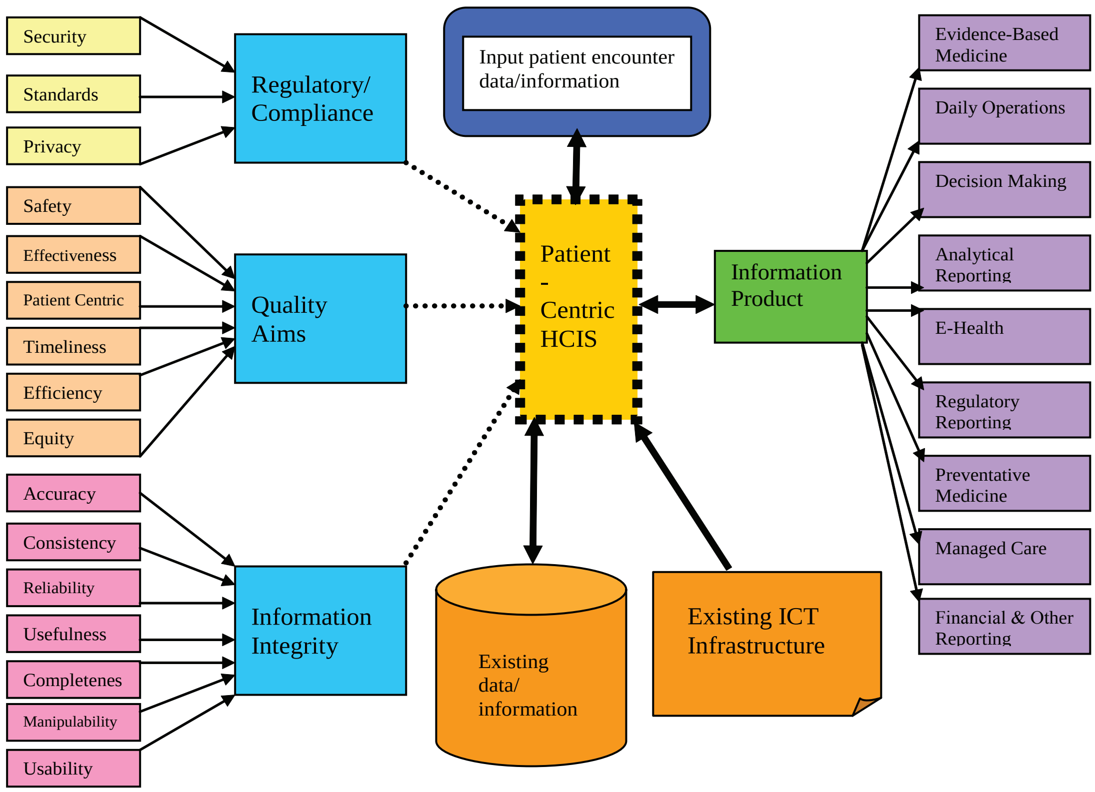
Wickramasinghe, 2005). Hence, it is reasonable to expect that HCIS in general, and e-health in particular, should be able to play a similar role for healthcare organizations; specifically these technologies should be viewed as key enablers for healthcare to meet the current challenges of escalating costs, an aging population that presents more complex treatment issues, and the need to have high-quality information and provide superior healthcare delivery. In order to systematically maximize and facilitate the full potential of such initiatives to enable healthcare organizations to cope with current challenges, it is important to have a guiding integrative framework to facilitate the design and management of robust HCIS and e-health initiatives. The following integrative compliance framework serves such a role.

The proposed framework integrates the key challenges of:

1. the three elements of regulatory compliance taken from HIPAA, including transaction standards, privacy, and security;
2. the six healthcare quality aims; and
3. the core principles of information integrity.

Further, the framework also recognizes the multifaceted nature of the key players within the healthcare system—in particular, the dynamics of their information requirements with respect to capturing, generating, and disseminating the necessary information. Moreover, the framework acknowledges the existence of other challenges such as existing IT infrastructure limitations, gathering of data/information from the patient encounter, and integrating it with (when necessary) other stored data and information. These challenges, though, are beyond the scope of this

Figure 1. An integrative compliance framework for patient-centric healthcare systems



article but are reflected in the framework for completeness. The key component of this framework is the protocols for ensuring compliance, observing the principles of information integrity, and satisfying the healthcare quality aims (Q). Figure 1 depicts the proposed framework. Finally, the framework highlights the major deliverable from the HCIS—namely, the information product and its key applications to various healthcare practices and processes.

The patient-centric healthcare information system received input of patient data at the time of the patient encounter. This information is cross-referenced with any existing data/information stored in databases and also validated using ICT infrastructure, for example, intranet capabilities. In addition the data and information are structured to satisfy regulatory and compliance standards (such as HIPAA), the six key quality aims, and the tenets of information integrity. Once this information is deemed to be correct, the resultant information product is then applied to the healthcare treatment or used to develop the evidence base. The information is input usually by a human, but is continuously corrected if required to ensure that accurate information permeates the system.

DISCUSSION

The proposed framework described above highlights the major information flows within a healthcare system. As can be seen information and data are critical to all healthcare activities, and thus the benefit of a robustly designed patient-centric HCIS that supports healthcare organizations in creating, generating, and disseminating germane data, information, and knowledge is crucial. Such a well-designed HCIS will also assist healthcare organizations as they try to address the major healthcare challenges. Ultimately, it is through having pertinent and relevant information that superior decision making can ensue, and such superior decision making is paramount if cost-effective quality healthcare delivery is to prevail (von Lubitz & Wickramasinghe, 2006). Specifically, the framework provides a useful diagnostic and monitoring tool to help healthcare organizations identify the key challenges they face regarding their healthcare data, and consequently design and/or revise the IC^2T architecture accordingly so that it takes these challenges into consideration, complies with regulatory requirements, and produces information of the highest quality and integrity. This occurs at:

1. the micro level, identifying when and how errors enter the system or if the data is in fact non-compliant with governance criteria
2. the meso level, where comparisons can occur between departments pertaining to utilization of resources and/or treatment protocols
3. the macro level, where comparisons across the healthcare organization can be made with national averages to see if, in fact, cost-effective quality healthcare delivery is occurring and where, if required, changes should be made

This will then ensure that the resultant information product will be of the highest caliber and hence support superior decision making and excellence in healthcare delivery. Given the iatric nature of healthcare procedures, the significance and need for superior information or conversely the ramifications of inferior information cannot be stressed strongly enough. It is worth noting that the framework is as relevant to existing HCIS in meeting these key challenges and refining the information product as it is to the design and subsequent management of new HCISs; hence the framework serves both as a diagnostic and prescriptive tool.

CONCLUSION

The healthcare industry is currently facing constant and relentless pressures to satisfy conflicting goals such as lower cost while maintaining and increasing the quality of service. In addition, healthcare organizations must also contend with regulatory pressures such as HIPAA compliance. In such a challenging environment, the need for well-suited and high-quality information that flows throughout the Web of the healthcare system becomes paramount. Patient-centric HCIS and e-health initiatives can be used effectively and efficiently to facilitate the information flows and consequent decision-making activities throughout this healthcare system Web. This is inevitable, given the general success of IC^2T in many business settings and the fact that healthcare is a data- and information-rich industry. What is critical, given the omnipresent and critical nature of data and information in healthcare organizations, is to take a holistic and comprehensive approach to designing a robust HCIS and subsequently effectively managing this asset.

The proposed integrative compliance framework provides a solution to the pressing problem faced by healthcare organizations of addressing the myriad of ethical, security, and informational considerations regarding their HCISs. Specifically, the framework considers the three key challenges of HIPAA compliance, the need for embracing the principles of information integrity, and the need to embrace the six healthcare quality aims. Hence, the framework provides a solid foundational step in trying to design an appropriate architecture for a patient-centric HCIS that collectively structures the key challenges facing healthcare organizations, analyzes their constituent elements, and develops protocols that attempt to ensure that not only are the data inputs into the HCIS of high quality, but also and of equal importance that the resultant information product is of a high quality and integrity. This in turn will ensure superior decision making and support excellence in healthcare delivery.

REFERENCES

Applegate, L., Mason, R., & Thorpe, D. (1986). Design of a management support system for hospital strategic planning. *Journal of Medical Systems, 10*(1), 79-94.

Bush, G.W. (2004). *State of the Union address.* Retrieved from http://www.whitehouse.gov/news/release/2004/01/20040120-7.html

Chandra, R., Knickrehm, M., & Miller, A. (1995). Healthcare's IT mistake. *The McKinsey Quarterly*, (5).

Croasdell, D.C. (2001). IT's role in organizational memory and learning. *Information Systems Management*, *18*(1), 8-11.

Geisler, E., Lewis, D., Nayar, M., & Prabhaker, P. (2003). *Information integrity and organizational performance: A model and research directions.* Working Paper.

Global Medical Forum Foundation. (2005, April 4-6). *GMF IV: New frontiers in healthcare*. Retrieved from http://www.globalmedicalforum.org/gmf_iv/session3.asp

Haag, S., Cummings, M., & McCubbrey, D. (2004). *Management information systems for the information age* (4th ed.). Boston: McGraw-Hill Irwin.

Hammond, C. (2001). The intelligent enterprise. *InfoWorld*, *23*(6), 45-46.

HIPAA (Health Insurance Portability and Accountability Act). (2001, May). *Privacy compliance executive summary.* Protegrity Inc.

HIPAA. (2002). *HIPAA security requirement matrix*. Retrieved from http://www.hipaa.org/

Holt, G.D., Love, P.E.D., & Li, H. (2000). The learning organization: Toward a paradigm for mutually beneficial strategic construction alliances. *International Journal of Project Management*, *18*(6), 415-421.

Huang, K., Lee, Y., & Wang, R. (1999). *Quality information and knowledge.* Englewood Cliffs, NJ: Prentice Hall.

Inchingolo, P. (2003). Picture archiving and communication systems in today's healthcare. *Business Briefing: Next Generation Healthcare,* 93-97.

Institute of Medicine. (2001). *Crossing the quality chasm—a new health system for the 21st century.* Washington, DC: National Academy Press.

Johns, P.M. (1997). Integrating information systems and health care. *Logistics Information Management, 10*(4), 140-145.

Mandke, V., Bariff, M., & Nayar, M. (2003). Demand for information integrity in healthcare management. *Proceedings of the Hospital of the Future Conference.*

McGee, M. (1997). High-tech healing. *Information Week*, (September 22).

Moore, T., & Wesson, R. (2002). Issues regarding wireless patient monitoring within and outside the hospital. *Proceedings of the 2nd Hospital of the Future Conference*, Chicago.

Oslo Declaration. (2003, June 12-13). Oslo declaration on health, dignity and human rights. *Proceedings of the 7th Conference of European Health Ministers,* Oslo, Norway. Retrieved from http://odin.dep.no/hod/norsk/aktuelt/taler/monister/042071-990209/dpk-bn.html

Perper, J. (1994). Life threatening and fatal therapeutic misadventures. In M. Bogner (Ed.), *Human error in medicine* (pp. 27-52). Hillsdale, NJ: Lawrence Erlbaum.

Scott Morton, M. (1991). *The corporation of the 1990s.* New York: Oxford University Press.

Stegwee, R., & Spil, T. (2001). *Strategies for healthcare information systems.* Hershey, PA: Idea Group.

Thorne, K., & Smith, M. (2000). Competitive advantage in world class organizations. *Management Accounting, 78*(3), 22-26.

Troisfontaine, N. (2000). *Preparatory report: Entrepreneurship in healthcare.* University of Twente, The Netherlands.

von Lubitz, D., & Wickramasinghe, N. (2006). Healthcare and technology: The doctrine of network-centric healthcare. *International Journal of Electronic Healthcare.*

Wallace, S. (1997). Health information in the new millennium and beyond: The role of computers and the Internet. *Health Education,* (3), 88-95.

Wickramasinghe, N. (2000). IS/IT as a tool to achieve goal alignment: A theoretical framework. *International Journal of Healthcare Technology Management, 2*(1/2/3/4), 163-180.

Wickramasinghe, N., & Fadlalla, A. (2004). An integrative framework for HIPAA-compliant I*IQ healthcare information systems. *International Journal of Health Care Quality Assurance, 17*(2), 65-74.

Wickramasinghe, N., Fadlalla, A., Geisler, W., & Schaffer, J. (2003). Knowledge management and data mining: Strategic imperatives for healthcare. *Proceedings of the 3rd Hospital of the Future Conference,* Warwick, UK.

Wickramasinghe, N., & Mills, G. (2001). MARS: The electronic medical record system, the core of the Kaiser galaxy. *International Journal of Healthcare Technology Management*, *3*(5/6), 406-423.

Wickramasinghe, N., & Schaffer, J. (2006). Creating knowledge driven healthcare processes with the intelligence continuum. *International Journal of Electronic Healthcare.*

Wickramasinghe, N., & Silvers, J.B. (2002). IS/IT: The prescription to enable medical group practices attain their goals. *Healthcare Management Science.*

KEY TERMS

E-Health: The use of Web-based technologies, especially Internet capabilities to support healthcare delivery.

Evidence-Based Medicine: The practice of medicine where treatments and/or procedures must be justified from sound past case history or "evidence."

Health Information Portability and Accountability Act (HIPAA): Enacted by the U.S. Congress in 1996.

Healthcare Actors: All participants in the healthcare system including patients, providers, regulators, insurance companies, and healthcare organizations.

Healthcare Delivery: The process of offering treatment to patients.

ICD-9: Ninth revision of International Classification of Diseases, the classification used to code and classify mortality data from death certificates.

Information Integrity: Key tenets for ensuring that information entered or derived is accurate, correct, pertinent, and useful.

Patient-Centric HCIS: Healthcare information system centered around the patient, hence all data and information pertaining to a patient is recorded and stored together, and can be accessed and used throughout the network in a seamless fashion.

Quality Aims: There are six: healthcare should be safe, effective, patient centered, timely, efficient, and equitable.

ENDNOTES

1 Readers interested in the complete HIPPA security requirements are referred to HIPAA (2002).

2 Table elements developed from http://www.hhs.gov/ocr/hipaa/

This work was previously published in Encyclopedia of Information Ethics and Security, edited by M. Quigley, pp. 498-506, copyright 2007 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 5.5
E-Learning in Healthcare and Social Care

Maria Kalogeropoulou
National and Kapodistrian University of Athens, Greece

Maria Bastaki
National and Kapodistrian University of Athens, Greece

Polyxeni Magoulia
National and Kapodistrian University of Athens, Greece

ABSTRACT

E-learning has the potential to transform learning for healthcare and social care, supporting the aims of the NHS Plan and raising standards of care for patients and service users across health and social care. This chapter sets out a vision of healthcare and social care services in the 21st century, and a strategy for making it a reality. The authors present and discuss here the basic principles and benefits of e-learning for healthcare professionals, medical students, and patient education.

INTRODUCTION

E-learning is the use of interactive technologies to support and improve learning. It is not just about online courses. E-learning can include a range of technologies from CD-ROMs (compact disc read-only memory) to electronic whiteboards or online simulations. It should usually include some form of support, whether face to face or electronic, and can often be blended with classroom methods. It can offer learners and tutors many services, including access to resources, information, and advice. It can reduce the time spent on administration, and help with the planning, recording, and tracking of learning and development. An e-learning strategy is therefore really an aspect of a strategy for effective learning.

E-learning has the potential to transform learning for health and social care, supporting the aims of the NHS Plan and raising the standards of care for patients and service users across health and social care. This document sets out a vision of health- and social-care services in the 21st century and a strategy for making it a reality. The vision is of a health and social sector in which:

- Patients and service users have the information they need to be involved in their own care, and know that staff have the skills and expertise to give them the highest standards of care,
- All health- and social-care staff can access the learning opportunities and support they need to develop personally and professionally,
- Flexible learning is a central part of everyday work for everyone,
- The highest standards of professionalism are found throughout all occupations and communities,
- People share knowledge, resources, expertise, and good practice within and across their communities, and
- Resources are used effectively to provide lifelong learning and continuous development opportunities for all staff now and in the long term.

In the past, involvement in learning has been largely a matter of personal preference and opportunity, governed by the individual's own motivation, their seniority, the availability of suitable learning, and the support of colleagues and supervisors. Increasingly, factors such as the ones listed here mean that learning is becoming a central part of everyone's working life:

- The rapidly changing workplace
- A more competitive job market
- Increased emphasis on teamwork
- Informal learning
- Technology
- The need for professionalism

E-learning is increasingly widely used by learners in schools, colleges, and universities. It is also widely used in work-based learning and corporate education, and in industry and the public sector. E-learning is therefore a significant factor in the personal and professional development of the 1.2-million-plus people who work in the health sector, and the 1.4 million who work in social care. Clearly, e-learning is important when it comes to acquiring job-related knowledge and skills.

CAPABILITIES AND BENEFITS OF E-LEARNING

E-learning strategy sets out generic e-learning capabilities for the education sector. It is useful to look at how these could apply to health and social care.

- **Individualized Learning:** Meeting the needs of all staff, including those working in remote locations, in the home, or in small organizations, or whose work requires them to be mobile
- **Personalized Learning Support:** Exploring learning pathways and resources, finding the right courses and materials, and tracking work-based learning
- **Collaborative Learning:** Including collaboration between learners on work-based projects or action research (on, for example, national service frameworks), and supporting health informatics communities or health- and social-care interprofessional groups
- **Tools for Educators and Employees:** Support for innovation by customizing or creating learning resources or simulations
- **Virtual Learning Worlds:** Online master's classes and simulations, and access to virtual campuses or wider learning environments
- **Flexible Study:** On-demand learning, which people can access when and where they need it

- **Online Communities of Practice:** Bringing together specialist communities, practitioners, learners, community or voluntary workers, and service users and care givers
- **Quality at Scale:** Providing access to e-learning resources and services right across the sector, without variations in standards, that are linked to information, and HR (human-resource) and management systems

E-learning can offer huge practical benefits. The value of e-learning for all parties is clear:

- **For Individuals:** Freedom to develop, both personally and professionally, through accessible learning opportunities
- **For Employers:** Engaging staff and promoting a sense of ownership and involvement
- **For Managers:** Achieving business and performance targets
- **For Health Professionals:** Better collaboration and communication, and creating more development opportunities
- **For Providers:** Widening participation in learning, at work, in the community, and at home
- **For organizations:** Becoming partners in workforce-development functions and promoting knowledge management
- **For Patients and Service Users:** Getting individuals and communities involved in improving care outcomes
- **For all Staff:** Interlinking the technologies used for learning and for work

E-LEARNING FOR HEALTHCARE PROFESSIONALS

In recent years, as the demand for lectures without the limitations of time and place by those who have jobs and require lifelong education grows, there are more and more expectations on the implementation of distance learning systems (Sakai, Mashita, Yoshimitsu, Shingeno, Okada, & Matsushita, n.d.). There is a double source of demand for services: institutional and personal. Initially, the demand will basically arise from the public education system of health professionals and technicians. In this initial phase, recipients of information and courses will be of two kinds: technical staff of health divisions or institutions in every state, and managing, medical, and nursing staff of hospitals and ambulatories in individual states (Pulido & Requena, 2003).

Until now, professionals were required to take time away from their practices and personal lives to attend in-person training sessions. The introduction of an Internet-based training curriculum provides physicians with the flexibility and convenience of learning at their own pace and on their own time.

In-depth, efficient training for physicians enables them to deliver an even higher standard of care to patients. E-Learning provides physicians with the means to address their training needs immediately (Raz, 2002). The e-learning program will help participants assimilate the basic principles and practices of therapies, including patient selection, where the treatment should be positioned in the care continuum for diseases, and the significant symptoms of these diseases. The e-learning curriculum consists of modules, and some of them are free for physicians. The modules are interactive and include animation and streaming video, as well as exercises and review questions. In addition, participants receive a CD-ROM containing educational surgical or other kinds of videos, and audio programs on patient-physician communication. Finally, physicians have the option of communicating with e-learning faculty members.

E-LEARNING FOR MEDICAL STUDENTS

It is important to note that a lot of people believe that traditional training combined with a strong implementation process and coaching can create measurable results for students. E-learning can increase the retention of learned information by an additional 35% to 45% when blended with other training approaches and technologies.

Undergraduate medical students are a group of individuals who often have a list for knowledge along with the skills and opportunities to use information technology. Furthermore, as the doctors of the future, they will be increasingly involved with technology in their professional practice.

Past experience in developing e-learning packages for students learning tutorials has shown them to be powerful tools for learning as there are exciting opportunities for the use of images, video, and links to external sites or relevance. However, they are extremely labour intensive to set up and, unless created judiciously, can lead to information overload. Furthermore, whilst they are valuable for developing learning in the cognitive domain, they are less useful for improving interpersonal skills and changing attitudes (Marshall & Kirwan, 2004).

The best balance seems to be reached by providing small-group sessions with a facilitator to consolidate the subject, and to place it within the current clinical context. Furthermore, the involvement of the students in the production of e-learning packages helps to impart a sense of ownership to the students and overcome any resistance to the use of alternative teaching methods. General Medical Council (GMC) guidance on reducing the amount of information imparted to students (GMC & Tomorrow's Doctors, 2003) must be squared with ever increasing amounts of information freely available on the Internet—human contact is vital for this.

The University of Hull's School of Nursing, Social Work, and Applied Health Studies has developed an interdisciplinary, part-time module called Facilitating E-Learning in Health and Social Care Education and Practice to provide health- and social-care professionals involved in education with the opportunity of considering how they can integrate e-learning into their practice and teaching. A major step in implementing the university's e-learning strategy within the school has been to implement a virtual learning environment: an integrated set of electronic teaching tools that are available to students and teachers online, allowing information transfer, learning-material delivery, and communication through e-mail, discussion, and chat to support student learning (Santy, 2003).

The module has the potential to attract students not only nationally, but also internationally because the module can be delivered at any time to any place with an Internet connection. E-learning can make these students more independent and lifelong learners. New technology has been transforming the way we think about education in health and social care, and the increasing use of technology in practice settings.

E-LEARNING FOR PATIENT EDUCATION

GeneEd develops new e-learning systems for patient education for major pharmaceutical clients. As the leading provider of advanced e-learning solutions for the life-science industry, it announced that it has extended its e-learning curriculum to span all subject areas, from drug discovery and molecular medicine to patient education.

The GeneEd patient-education curriculum that was announced (in addition to the over-100 therapeutic courses released over the last two quarters) represents a significant expansion of GeneEd's product line. GeneEd products are extensively used in the research, development, and sales divisions of major global pharmaceutical organizations. In combination with the soon-to-be-released regulatory and compliance curriculum, the patient-education program greatly enhances the GeneEd product portfolio to address all of the mission-critical training needs of modern life-science organizations. As the role of molecular medicine broadens its reach within pharmaceutical organizations, employees, physicians, and patients alike are required to absorb the complex scientific and medical issues this new science brings. GeneEd is the only organization providing validated e-learning across the spectrum. These techniques range from genomics and proteomics, through disease states, all the way to physician and patient education. GeneEd is utilizing its award-winning, proprietary Repurposing™ and Viewing™ engine technology to offer compelling, intuitive, and effective e-learning to some of the most demanding users of information: everyday patients trying to better understand their diseases and the therapeutics that may help them (Patel, 2003).

Patient-education e-learning courses empower physicians, allied health professionals, and patients alike with the knowledge critical to the understanding of diseases and the therapeutic regimens used to treat them. E-learning courses are actively responding to the ever-changing training needs of the life-science industry, and the introduction of this new patient-education curriculum reiterates the strength of our science-focused curriculum design.

DISCUSSION

While there are clearly many benefits and advantages to e-learning, we also need to take into account the barriers, challenges, and disadvantages associated with it. These include high development costs, barriers to access for disadvantaged learners or those with disabilities, and the misconception that online learning is a solitary and unsupported activity. It is particularly important to address any barriers relating to potential users so that e-learning really does benefit all target groups. To gain full benefit, we will need to take steps to guard against a potential digital divide by addressing both access and skills. We also need to achieve the right balance between e-learning and traditional methods. While e-learning can make a powerful contribution to large-scale engagement in learning, as well as the tailoring of learning to individual needs, it should not and cannot replace all other approaches to learning. An e-learning strategy should be one aspect of a wider learning strategy.

CONCLUSION

E-learning is developing rapidly worldwide. Any strategy will therefore need to have an international dimension so that e-learning for the NHS can be genuinely world class. We may need to harness international knowledge and best practice, and use research to benchmark NHS e-learning against the very best globally. Furthermore, with the increasing pace of globalization, learning resources and opportunities that originate abroad will increasingly be available, and health and social care will require systems that are compatible and interoperable with those in use in other fields at home and abroad.

E-learning is not an end in itself, nor is it a marginal activity related only to online courses or distance learning. It will increasingly embrace all aspects of learning and will therefore form a fundamental part of how people will learn in 10 or 20 years. Although we cannot predict exactly which technologies or which models of learner support will be most widely used, existing examples of leading practice, whether in the United Kingdom or elsewhere in the world, provide some indications. These examples, as well as alternative scenarios, should inform the emerging e-learning strategy for health and social care.

REFERENCES

Department of Education and Skills (DfES). (2003). *Towards a unified learning strategy.*

Department of Health (DOH), Cheshire and Merseyside, Cumbria and Lancashire, and Greater Manchester WDCs. (2003). *Delivering e-learning in the NHS. Getting the blend right: A strategic approach for the north west.*

General Medical Council (GMC) & Tomorrow's Doctors. (2003). *Recommendations on undergraduate medical education.*

Marshall, R. W., & Kirwan, J. R. (2004). *Development of e-learning tutorials in rheumatology.*

Pulido, P., & Requena, J. (2003). *Alternatives in "e-learning" for health professionals in Latin American and the Caribbean.* Panamerican Federation of Associations of Medical Schools (PAFAMS).

Sakai, S., Mashita, N., Yoshimitsu, Y., Shingeno, H., Okada, K., & Matsushita, Y. (n.d.). *An efficient method of supporting interactions for an integrated distance learning system.*

URL

http://www.cmwdc.nhs.uk/elearning/elearning-strategy.pdf

http://www.dfes.gov.uk/consultations2/16/

http://www.dfes.gov.uk/elearningstrategy/

http://www.edu.gov.mb.ca/ks4/tech/wbc/wbc-gloss.html

http://www.geneed.com

http://www.google.com

http://www.interstim.com

http://www.medtronic.com

http://www.nhsu.nhs.uk./webportal/learning/e-learning/StrategyDocument.pdf

http://www.training.interstim.com

http://www.UniversityofHull.com

KEY TERMS

CD-ROM: High-capacity optical storage medium.

Distance Learning: Learning in which students and instructors are separated by distance or time.

E-Learning: A wide set of applications and processes including Web-based training, virtual classrooms, digital collaboration, and computer-based training.

Face to Face (F2F): Students and teachers are in the same location at the same time.

Internet-Based Learning or Training: Courses delivered via Internet technologies (these were text based before the Web).

Learning Networks: Communities of learners connected via computer networks.

Learning Objects: Course materials developed according to a standard (e.g., IMS) that allows the easy sharing of materials. Materials are modular and can be used for a variety of purposes and outputs.

Online: Connected to a computer network.

Online Community: Learners who, although separated by distance and time, share a common experience.

Online Conferencing: Communications happening over a network. They can be synchronous via chat or asynchronous via discussion groups.

Web-Based Course (WBC): Distance education course materials supported by computer-mediated communications and delivered asynchronously via the World Wide Web. Ancillary materials such as print, videos, and CD-ROMs may be required.

This work was previously published in Handbook of Research on Informatics in Healthcare and Biomedicine , edited by A. A. Lazakidou, pp. 112-118, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 5.6
Fostering Meaningful Interaction in Health Education Online Courses:
Matching Pedagogy to Course Types

Richard G. Fuller
Robert Morris University, USA

Gary Kuhne
Penn State University, USA

ABSTRACT

This research study examined the best interactive practices of effective health care education faculty from six major universities that offer online health care programs. Program directors from six major universities identified effective faculty, from which twelve faculty members were interviewed to uncover effective practices and an additional thirty faculty participated in a Delphi study to identify and prioritize effective practices. The findings for this study indicate that different types of facilitation approaches are needed to generate adequate interaction in four distinct types of health care courses, i.e., foundational classes, skills classes, analysis/synthesis classes, and hybrid type courses.

INTRODUCTION

Wlodkowski (1999, 1985) suggests that effective instructors must have expertise, empathy, enthusiasm, and clarity, a conclusion that is relevant whether the teaching takes place in traditional face-to-face settings or in online formats. Interaction with the student is a central factor in demonstrating each of these elements. Effective instructors have discovered interactive practices that work well in face-to-face classrooms. The challenge for online instructors is to discover how to replicate effective interaction practices within the online, asynchronous learning environment. As with traditional face-to-face teaching, there are methods and techniques that work in some venues and not in others. Strategies that foster effective interaction in an engineering classroom may not offer the same efficacy in a healthcare program. Effective teachers are willing to explore why certain interaction techniques work and don't work in order to discover the most effective techniques for their particular educational programs.

Knowles (1999, 1980) and Rogers (1969) argue that adult education teachers serve as facilitators, providing the resources to enhance and facilitate the self-directed learning opportunities of their students. Such an understanding of the role of instructors is particularly pertinent to the online asynchronous arena of higher education systems, where educators must design a variety of courses for a variety of learners. Effective instruction requires that teachers understand the changing needs of their learners based upon the nature of the educational program. In other words, an effective instructor does not approach each learning situation with the same pedagogy and style. Different styles of course require different techniques to facilitate success learning. Knowles (1980) suggests the specific learning needs of the particular participants of a given learning activity must be diagnosed. Understanding the unique needs of different university and college programs at graduate and undergraduate levels in different disciplines will go far toward enhancing interactive teaching practices online. Regardless of the mediated nature of the communication, "It is the teacher's responsibility to precipitate and facilitate learning that has purpose and is focused on essential concepts and worthwhile goals" (Garrison & Archer, 2000, p 48.). Adults and distance-education students relate in an interactive collaborative construction of knowledge, a system that typifies many of the concepts of adult education theory (Anderson, et al., 2002). The dilemma facing online instructors is how best to accomplish the designing, facilitating, and guiding of a predominantly text-based learning arena to best foster the different levels of interaction required for learning success.

BACKGROUND

In conducting online teaching, interaction needs to be planned to facilitate learning. Vrasidas and McIsaac (1999) found that structure can affect interaction, and concluded that educators need to design courses to foster learner-to-learner interaction and dialogue. Kozma (1991) agrees with the need for less structure and more dialogue and suggests that learners should actively collaborate in order to construct knowledge rather than relying solely on knowledge gained from direct instruction. For such collaboration to occur, learners must feel a sense of connectedness with the group (Gibb, 1995).

Howland and Moore (2002) found that when students initiated interaction with instructors and other students, knowledge was often built spontaneously, such as through students guiding the direction of discussion-board threads. Such student leadership then led to positive results for others. One student said, "Several times, I have seen questions asked by others that had not even occurred to me to ask and the answers benefited me" (p. 188). Swan (2001) found that students with higher levels of interaction with their classmates through online discussion also reported higher levels of learning and satisfaction from courses. Rovai and Barnum (2003) also found evidence that student perception of learning from online courses was positively related to course interaction, lending further support to the need to provide opportunities for online students to learn by active interaction with each other and with instructors. Effective online instructors develop highly interactive material and facilitate participation in online discussions. Rovai and Barnum also suggested that passive interaction, analogous to listening to,but not participating in, discussions, was not a significant predictor of perceived learning in the present study. Consequently, using strategies that promote active interaction leads to a greater perception of learning and higher levels of learner satisfaction.

Beaubien's research (2002) described instructor characteristics that contribute to effective online courses. Students need to feel that the instructor is online regularly. The instructor does not need to be intrusive to the online dialogue but his/her presence needs to be known. Short postings are good for the most part but the teachable moment should be capitalized upon to provide sufficient information and clarification as is necessary. Sometimes the instructor can pose questions that will stimulate or lead the discussion in a direction. Instructor modeling of a high level of presence sets a positive norm for the class and encourages students to do the same. Moore (2001) suggests that instructor interaction should have the goal of establishing a culture of independent learning and peer participation. Positive instructor feedback tends to bring out the best in people and motivate them to invest discretionary effort (Braksick, 2000; Daniels, 2000). Positive instructor feedback can energize the learning system and increase interaction frequency.

Vrasidas (1999) examined the conceptual framework of interaction in online courses and found that the factors influencing interaction were learner control, social presence, structure, feedback, and dialogue. In a follow-up study, Vrasidas and McIsaac (1999) found that each of these factors has specific implications for teaching practice. For example, activities can be structured to increase interaction with the instructor, other students, and the course content. Discussing a paper outline with an instructor; collaborating activities with peers, and participating in required online discussions were found to increase interaction among participants. This study, not unexpectedly, found that higher-than-expected workloads contributed to decreased interaction. Thus, the appropriateness and on-task time of course requirements must be suitable to facilitate online interaction.

PURPOSE OF THE STUDY, HYPOTHESES, AND QUESTIONS

The purpose of this study was to investigate the interactive practices of effective online health care graduate and undergraduate instructors as gleaned from the experience of successful faculty. This study focused on understanding how the online teaching technologies in the courseware of an Internet portal system were being employed to promote interaction. The goals of this research study were: (a) to better understand the phenomenon of successful online computer-based education in graduate and undergraduate healthcare education through the identification and description of online educational constructs that exemplify effective interactive practice, (b) to better understand how effective distance educators in health-care education utilize the innate capabilities of online courseware to support interactive constructs, and (c) to better understand what techniques and strategies faculty employ to foster and facilitate the sense of interaction. It was hypothesized that effective faculty employ certain methodologies, practices, and mindsets in planning and active teaching phases to promote interaction when utilizing online courseware. The research asked what successful online distance education faculties do to make their teaching more interactive, and do they use different approaches to generate interactivity in different situations.

The Study or Methodology

The research design for this study involved a triangulated three-tiered process. The first phase was to identify graduate and undergraduate health-care faculty from major university schools of nursing and health professions who provide effective interactive education that fosters learning. Potential faculty were identified by program directors from six university health education programs (The University of Pittsburgh, George Washington University, West Virginia State

University, West Chester University, University of Maryland, Baltimore County (UMBC), and Drexel University), who based their identification on the criteria that the instructor, (a) promotes a high level of student-to-student interaction through threaded e-mail and discussion-board activities as well as other activities that allow learners to construct or formulate an idea in a deeper sense, and raises the interest and motivation of the students, (b) the instructor promotes a high level of student-to-instructor interaction through both quantity and quality of assignments that maximize the impact of interactions, and (c) the instructor promotes a high level of student-to-content interactions through offering a variety of activities and resources that offer students a variety of alternatives for learning.

The second phase, utilizing the program director's listings, involved interviews with twelve (12) selected faculty from the list developed in phase one. These instructors were requested to participate in phenomenological interviews, either face-to-face or by telephone to establish trends and common themes in effective online instruction.

The third phase of the research involved taking the 12 phenomenological interviews and synthesizing the results to create a Delphi questionnaire for use with an experienced group of 30 faculty members drawn from the list developed in phase one. Trends and common threads were identified to assist in categorizing the data. To assure that all common themes and trends were identified, a second evaluator was utilized to assure inter-rater reliability. The questionnaire was operationalized following the Delphi Technique with the 30 additional faculty. This tool allowed a group of defined experts to come to a consensus of opinion when the decisive factors were subjective, and not knowledge-based. Through a series of questionnaire exchanges, the experience group identified additional ideas through individual brainstorming and communicating ideas with the investigator to clarify and validate the findings from the previous questionnaire. Questionnaires were exchanged through e-mail to maximize efficiency and minimize time associated with conventional mail. A series of three exchanges with progressive fleshing out of ideas as well as generation of new ideas beyond those attained through the phenomenological interviews was facilitated. It was determination after the third questionnaires that no new ideas had emerged.

Limitations

The attempt of this research was to capture interactive fundamentals of practice among health care online educators. It is acknowledged that in the context of the interview, that some may have issued their espoused theory and not their actual theories in use (Argyris, 1999). Argyris suggests that people consistently act inconsistently, unaware of a contradiction existing between their espoused theory and their theory-in-use (what they actually do in the practice setting). Despite this possibility, participants offered what they envisioned as the most effective pedagogies to maximize interactivity in online health care programs.

It is further recognized that one of the limitations of this study is that the definition of effective practice is based upon the single lens, or perspective, of the identified instructors. This research did not take into consideration student perspectives of effective practice. This research also relied upon the identification of effective faculty from the perspective of program directors who based their recommendations upon a provided set of criteria.

Interview Findings

The 12 faculty members interviewed suggested that different types of courses require different types of facilitation to generate effective interaction. As one instructor for this study stated:

There are different types of courses, such as the hard sciences classes versus the social science classes. There may be classes such as skills classes, such as a clinical class, or a research class where students need to leave with a skill. There may be social science or discussion/opinion classes that take information and apply it in different situations. This is both online and in the traditional courses. There is a big difference. At times the conceptual pieces in putting things together as opposed to the nitty-gritty facts and research and data collection are trickier to do. The relating this in the everyday life can be more difficult and more challenging.

Four types of courses were distinguished, including foundational classes, skills classes, analysis/synthesis classes, and hybrid type courses.

- **Foundational or rote memorization courses:** Such courses provide the foundation or knowledge that will be used to build upon in other courses. The prioritized interactions in such courses focus on helping the student make more "student-to-content" connections. As one instructor from this study stated, "In the foundational course, you either get it or you don't." Foundational courses include courses like anatomy and physiology, pathophysiology, medical economics, and the business of health care. Many of the assignments in such courses are e-mailed weekly based upon module content to assure interaction with the content and understanding of key concepts. Some "student-to-student" is fostered, primarily through having assignments shared and discussed on discussion boards, something that does not routinely happen in a face-to-face classroom. As one instructor stated:

Health Care economics is a foundational course (not all economics courses are this way). In this course it is more of the memorization and regurgitation that is important. I can't ask them to compare and contrast John Maynard Keyes to Karl Marx because they are not there yet. I facilitate this differently than an upper level class. I use more discussions based upon the facts. They come into their first class and think that they understand this stuff. They think that they have all the answers. But they give what I call, "man on the street" answers. I have to remember that what I am doing here is building a "foundation." This type of course and a rote memorization course is foundational for other courses.

Such interactions are not generally open-ended and are designed primarily to reinforce the memorizing of basic concepts that will be applied later in their professional studies and practice. For example, an online nursing student may not understand all the reasons for memorizing anatomy and physiology, but will come to understand better the value of such memorization as the anatomy and physiology facts are applied in future courses.

The faculty interviewed believed that while foundational courses do not innately lend themselves to great online discussions, providing interactive e-mailed assignments, more instructor presence for Q & A, and instructor interaction in discussion boards helps to foster meaningful learning interactions with such courses.

- **Skills based courses:** Such courses require students to gain a particular skill(s) set that is applicable to a specific environment. The prioritized interactions in such courses focus on content, however, and expand beyond "student-to-content" connection, requiring instructor facilitation and presence to promote dialogue specific to the skill and the application of the skill. Instructors facilitate critical thinking and understanding of the concepts through the skillful use of questioning, such as "if we did this what would happen" or "how could we do this if…?" or "Great idea. Does anyone have any other directions?" Examples of such skill-based courses would include Nursing Research, Medical Informatics, and Patient Assessment. Skills classes such as Nursing Research or Health Education Research should be project-based, as students need to flesh out ideas and application of concepts. Interaction focuses on how and where to perform the skills as well as facilitating students toward a finite answer and how they get that answer (example: Identifying research questions). As one instructor stated:

Part of my goal is that I want them to understand the complexities of research and how you work with other people, so I do include a group project into the course. I believe that you learn from your peers. There is a skill set in how do you develop a proposal, how do you interpret statistics, how do you design research. The goal isn't to make them an expert when they are finished but to give them a set of skills that they can apply.

Another example drawn from an instructor teaching a Physical Assessment course:

I found teaching Physical Assessment online a difficult course to teach online because there is, or should be, so much hands on, and if you teach it in the classroom there is. So it is very difficult to deliver the physical assessment content Web-based because there has to be a video portion. The student has to be able to see how you percuss (thumping of the different body areas to determine if air or fluid is present), what the assessment of the abdomen looks like, and one of the big issues that we have is not bandwidth on our end at the university but among the receiving students. So I can decide to send them this fabulous thirty-minute clip, the best I've ever seen, and they are not going to be able to download it if they don't have the technical capabilities. I have to make a decision about how much information I can chunk into a block.

Assignments that facilitate student-to-content interaction in a skills-based class are weekly assignments that are e-mailed back to the instructor to assure that they are interacting with content and understanding it. Interaction with students and instructor focuses on the processes that they are going through in learning the content.

- **Synthesis/Analysis Courses:** Such discussion-based courses are used to teach students to analyze a situation and engage in problem solving. Instructors find that authoritative postings tend to shut down dialogue or that students simply parrot instructor ideas or postings (most students do not want to challenge instructors and risk receiving a lower grade). Dialogue among participants provides regular opportunities for reflection and inquiry and requires the least intervention in the discussion boards of all the course types. Simulated interaction in this manner through subject matter presentation can subsume part of the interaction by causing students to consider different views, approaches, and solutions and generally to interact with a course.

These are courses where core information is presented but there is not necessarily a right or wrong view. An example may be an Issues in Health Care, Health Care Policy, or Nursing Practice course where a module or lesson would focus on "Compare the value of the Canadian versus the USA health systems." This is the type of course or topic that will prompt many opinions and views. The key is to flesh out all angles of the subject and have the students explore and support the differing viewpoints with the facts. For the instructor in this type of class, the key was not to intervene too much. As one instructor stated:

I've found that if I post my particular opinions about a topic, then I change the discussion in that students stop posting or they just restate *my opinions. I think you have to be very careful. What I do is post more personal experience than personal thoughts and beliefs. You have to control that and it can be very difficult. You have to post more with "what I've found is." Any time you give them a clue on which direction you lean the majority of the class is going to lean that way too because they want a good grade too. So you need to be very careful how you do that. I think that if you come down with a very dogmatic statement then you shut them down as people don't want to be wrong. They are still learning and fleshing out their own thought processes and if you post too much you shut down their thought processes or they may not agree but are not willing to take this instructor on.*

There is a need to establish the culture of independent collaborative learning. Instructors do not respond to the majority of postings in this type of course but read them all and respond to key ideas and elements and, through additional questions, guide the learning process. If the discussion is getting one sided or negative, the instructor can, through a posted thought, direct the dialogue to view all sides. "That is true but what would 'so and so' say regarding this and why?" One technique in this type of course is to create an online debate where students don't get to pick but defend an assigned point of view, which forces them to see all sides. In a Nurse Practice course, one technique reported was that students were facilitated to choose a topic, interview someone, and then write a paper and discuss their issue and what they have found by leading a discussion-board thread.

- **Hybrid Courses:** Such courses have a combination of the above three and require a mixing of techniques to facilitate interaction. An example of this may be a Health Care Management or Leadership course where there is specific theory to understand and employ but also where you want students applying and understanding their own personal leadership style in different situations.

THE DELPHI FINDINGS

The Delphi process suggested that interaction strategies vary according to the type of course. Those participating in the Delphi component agreed with the majority of the interview findings, and in many cases further elaborated on the four types of course offerings.

Foundational or Rote Memorization Courses

It was agreed that while these types of courses don't generally lend themselves to great online discussions of the material, the interaction generated was centered on assisting students with learning the content material. It was also believed that a greater instructor presence was necessary so that some students don't have the feeling of learning alone or in a vacuum.

Skills-Based Courses

It was agreed that the interaction in a skills-based course is centered on the content, and that interaction was again used to make connections between content and skill application. This type of course was again found to require more instructor presence for students to gain the skill, and that the instructor's role was to generate thinking and a better understanding of the concepts. A majority of the Delphi participants utilize the weekly assignments format for a skills class. The concept of forcing a student-to-student interaction just for the sake of having one was not described to be effective. One instructor stated: "Certain topics don't lend themselves to meaningful discussion. Having a discussion assignment because there should be student-student interaction is not effective, does not facilitate mastery of the material, and frustrates the students." Another articulated that: "Instructor interaction in this type of course takes a lot of instructor facilitation and takes on a greater role with telephone follow-up when they are having difficulty in understanding concepts."

Synthesis/Analysis or Discussion-Based Courses

It was agreed that in these types of courses, the free flow of ideas monitored and facilitated by the instructor is the best technique. This requires that the instructor monitor carefully, and he or she must be diligent in fostering good online dialogue about the topics at hand. In this type of course, a majority of instructors reported using a "search and report" technique, where students go out and research a topic and then report back as a catalyst for generating good discussion boards. Additionally, instructors reported using weekly discussion board topic questions based upon the readings and students' own research to generate discussions. The instructor would monitor and then post as was appropriate to guide, stimulate, and assure good dialogue. As one instructor stated: "In my nurse practice course they will choose a topic and go interview someone and then write a paper and discuss their issue and what they have found by leading a discussion board thread."

Hybrid Courses

It was agreed that there are these types of courses but not all agreed on which type of course was in which category. Some reported that a research course was felt to be a skills class while others felt that it was a hybrid class where multiple techniques are employed. It becomes less of an interest to this research to categorize courses as it is to understand that there are different courses that may need facilitation using different techniques differently to maximize the online learning.

CONCLUSION AND DISCUSSION

One of the more significant findings in this research is the identification and confirmation that different types of courses require different types of facilitation to generate interaction. The four types of courses in the health-care educational systems (foundational classes, skills classes, analysis/synthesis classes, and hybrid type courses) require the employment of specific facilitation techniques on the part of online instructors. This finding has implications for interactive design and online teaching.

Foundational or Rote Memorization Courses

The interaction in this type of course is more focused on providing a greater understanding of material to serve as the basis for future learning. The central focus here is on student-to-content interaction. Health care programs have innately established a framework of knowledge and skill that serves as a foundation for future learning. The information learned in a Medical Terminology course is the building block for an Anatomy and Physiology course. The knowledge gained in these courses serves as the universal knowledge for the rest of the health care clinical education, whether in nursing or any other allied health field. These courses are typically very structured in nature and require the instructor to clearly articulate the material for ease of consumption by the learner. The learner of online foundational material must be independent in learning the material but, as with a face-to-face class, the interaction is designed to assist the learner to that end. Carnwell's (1999) concept of developing internal dialogue is supported here. His research indicated that the students in this situation desire more highly structured materials. Design of text materials is also important since the level of structure within the text may create either independence or dependence in students. Jones and Kemper (1994) suggested that independence can be fostered by requiring students to use self-study packages in an unsupervised manner.

Instructors need to design interaction that is more finite in assisting participants to understand the material. This interaction takes its form in assisting students to learn and memorize the material, to see how it all fits together. The design of the interaction is seen through e-mailing assignments to the instructor for feedback and assistance in learning the key objective elements of the foundational course. While this type of course does not lend itself to great online discussions, but is more focused on learning the concepts, it is also paramount for the instructor to practice a greater presence that provides the student with a sense of contact with the instructor (Townsend, 2002; Kanuka, Collett, & Caswell, 2002; Beaubien, 2002).

Skills-Based Courses

The implications of the research suggest that instructors need to design their interaction in this type of online course to assist students to gain the skills necessary to function in the particular health-care environment. Effective interaction requires designing discussion board activities to present applications and allow students to see other's work and learn from each other toward the final goal of attaining the skill, whether it is learning the physical skills of chest percussion, or lung sounds (where there are significant limitations), or the mental skills of designing and conducting medical research, or the skill of learning database construction and manipulation in a medical informatics class. The ability to demonstrate competence for the physical skills innate to health-care practice is one that has significant limitations in the current online environment within the present available technological structures, and assessment of competency often requires a face-to-face environment with credentialed professionals. The implications for the design of interaction are for the instructor to use a more facilitative role to assure that student-to-student interaction is being assisted here toward applying the information from the class to the field. Kennedy (2002) supported this premise as he suggested that learner-to-learner interaction is a valuable part of the online learning experience and that the distance education format is particularly well suited to engaging students in this type of interaction. These discussions can be deeper and more reflective, covering a broader range of issues that assists students to gain the skills taught in the course. Encouraging greater interaction among learners not only enhances student learning and application of the new skill set, but also places the instructor in a more supportive, facilitative role, which results in more efficient use of instructor time (Udod & Care, 2002). Instructor interaction in this type of course needs to take on greater monitoring of the learning, and the utilization of personal e-mail contact and telephone conversation may assist with students having difficulty in understanding concepts.

The interaction to facilitate these skills in online teaching needs to be planned to include learner-to-learner contact to assist students to learn from each other in how they design and apply the skills discussed above. Vrasidas and McIsaac (1999) supported this finding, as a need for educators to structure for dialogue including learner-to-learner interactions was found important. Kozma (1991) supported this premise seeing the need for more dialogue by visualizing learners actively collaborating with the medium to construct knowledge and skill.

Synthesis/Analysis Courses

The implications for these courses are for the instructor to understand that there may be multiple views of a situation that require exploration by the participants to fully realize the depth and breadth of the concepts. The obvious nature of these types of courses is for students to explore all aspects of the topic and then draw conclusions based upon the information presented and explored. The online instructor in this type of course must take care not to be too authoritative, as stated above, otherwise he or she will shut down the exploration and the insights that can be gleaned from students interacting and dialoguing about opposing viewpoints. Instructors will find that taking an authoritative stance and letting participants see their opinions on an issue will close dialogue, or they will find that students will simply repeat or "parrot" the instructor's view on the issue. This is in direct conflict of what the instructor may wish to accomplish by allowing students to analyze and synthesize all views of the problem or issue. Dialogue among participants provides regular opportunities for reflection and inquiry (Wesley & Buysse, 2001) and requires the least intervention in the discussion boards of all the course types. Simulated interaction through subject matter presentation in pre-produced courses can subsume part of the interaction by causing students to consider different views, approaches, and solutions and generally to interact with a course (Holmberg, 1999, 1989).

Hybrid Courses

Courses that reflect a combination of the above three forms require a mixing or blending of techniques to facilitate interaction. The advantage for an online instructor is to be aware of the first three types of courses and the methodologies that are effective in facilitating those types of courses, and then applying that on the micro-application level for the hybrid course.

If we frame the curriculum in the context of the four different types of courses, then we provide a better opportunity in the planning phases to maximize the learnings, as we are focused on the goals and best structure for interaction given the different types of course limitations and opportunities available in the online courseware. Knowlton (2000) believed that learning and teaching are reconceptualized in the online course to allow maximum independence among students by framing the curriculum and student interactions through the providing of resources and opportunities. Framing is used to facilitate students' desire to develop and implement shared goals in making connections with the curriculum. Students must be able to find space for their own inquires and needs within the assurance of a well-planned, content-rich, and flexible learning environment with adequate navigational tools and support systems (Vandergrift, 2002). This type of in-depth planning is more demanding and time consuming than the traditional classroom planning. A classroom teacher can draw upon his/her innate knowledge and platform skills to provide an impromptu lesson structure that cannot be duplicated in the online text-based arena. Vandergrift demonstrated the need for a deeper understanding of the dynamics that online teachers apply to the deliberate acts of design and interaction that facilitate successful courses.

Visor (2000) supported this as faculty serve to design and conduct a course, which is positive for student learning outcomes and serves to maximize the learning opportunities the same way that faculty foster learning in the traditional setting. They prepare and organize content according to well-established and communicated objectives, consider methodologies, which will assist the student to achieve the objectives in an online format, and be cognizant of time that the student will need to spend on the course. Wright and Thompson (2002) support this, as faculty establish a pedagogical strategy and begin to understand how online activities will occur in their discipline for maximum learning, and that faculty create templates in which they can insert their specific academic content. The four types of courses identified in this study have different goals and objectives for learning outcomes. Instructors need to facilitate interaction specific to the needs of the students in the confines of these course types.

The innate properties of the course shell portal allow options for organizing the pedagogy of an interactive course. This also becomes part of the consistency described to establish a "think-forward" type of lesson. The shell portals have innate properties built in that provide a "think forward" consistency of structure that becomes a comfort for students if interaction is designed consistently by the instructor. If the online course facilitates interactions in the nature described in this study, the constructivist and andragogical models of learning predict that successful learning is likely to result. These models require students to create their own meaning to knowledge in a self-directed manner and take more responsibility for their own learning (Knowles, 1999, 1980). This is not to say that the courseware becomes limiting but, as with the traditional classroom, there are confines of what can and cannot be done successfully. The innate properties of the course shell portal allow options for organizing the pedagogy of an interactive course and provide a level of flexibility to allow instructors and designers freedom to explore a number of pedagogies. Instructors need to be open to exploring new pedagogies that require different thinking.

FUTURE DIRECTIONS

Insight into establishing some of the best methodologies utilized in each of these types of courses would assist online instructors to establish the interaction that is employed and how best to maximize the competencies attained by students. Applying the practices used in the four types of courses would further establish a number of pedagogies that would work in different online health-care educational settings and may open the door for more nontraditional online courses to be offered. Brooks (1999) suggested that curriculum should be examined to determine how technology fits. Using technology, we should identify content that is technology neutral, technology driven, and technology enhanced. Salmon (2000) identified the teacher in the role of e-moderator as the key ingredient for effective teaching and learning online. It is the instructor who is the driver of the technology, given the differing nature of the four types of courses found in this research. Providing online educators with greater insights as to the application of pedagogies in different courses will assist them to create more effective online learning environments and help their students learn the true benefits of online learning communities as established from interactive practice priorities. Until now, there are those who would only see certain types of courses being taught online. With the identification of these four types, and the methodologies that can be employed, additional insights will assist the less traditional online course (clinical-based skills courses) to be opened to this technology.

Applying Holmberg's (2003) conversational theory to the four types of online health-care courses taught will assist in further understanding the dynamics of the interaction required to facilitate learning. This application will aid in offering insights to the pedagogies which promote learning in each of the four course types and the dialogue necessary to promote the interaction. Carnwell's (1999) dialogue is much more than merely transmitting messages to students about requirements of courses. Curriculum planners of distance education programmers and materials designers need to devise a balance of internal and external dialogue which allow students with different learning style preferences and approaches to gain the maximum from their learning experiences. MacDonald (2001) suggests that there may be particular times when extra formative feedback is of particular importance to students. Additional research is necessary to understand the balance of internal and external dialogue and the feedback necessary to maximize the learning opportunities in the four different types of courses.

REFERENCES

Anderson, T., Rourke, L., Garrison, D. R., & Archer, W. (2002). Assessing teaching presence in a computer conferencing context. *Journal of Asynchronous Learning Networks, 5*,(2). Retrieved April 20, 2003 from, http://sloan-c.org/publications/jaln/v5n2/v5n2_anderson.asp

Argyris, C. (1999). *On organizational learning* (2nd Edition). Malden, MA: Blackwell Publishers Inc.

Beaubien, J. (2002). Harnessing the power of complexity in an online learning environment. In K.E. Rudestam, & R. Schoenholtz(Eds.). *Handbook of Online learning: Innovations in higher education and corporate training.* Thousand Oaks: Sage Publications, 221-236.

Braksick, L. W. (2000). *Unlock behavior, unleash profits*. New York: McGraw-Hill.

Brooks, B. (1999). *Professional development for teaching technology across the curriculum*. Minister of Education, Edmonton, Alberta, Canada. Retrieved June, 23, 2005 from, http://ednet.edc.gov.ab.ca/technology

Carnwell, R. (1999). Distance education and the need for dialogue. *Open Learning, 16(2), 58-60.*

Daniels, A. C. (2000). *Bringing out the best in people: How to apply the astonishing* power of_positive reinforcement (2nd ed.). New York: McGraw-Hill.

Garrison, D.R., & Archer, W. (2000). *A transactional perspective on teaching and learning: A framework for adult and higher education.* Oxford, United Kingdom: Pergamon.

Gibb, J. (1995). *Tribes*. Sausalito, CA: Center Source Systems.

Gillis, A., Winston, J., Braid, A., MacDonald, P., & MacQuarrie, M. A. (2000). The learning needs and experiences of women using print-based and CD-ROM technology in nursing distance education. *Journal of Distance Education, 15(1),*1-20.

Holmberg, B. (2003). A theory of distance education based on empathy. In M.G. Moore, & W.G. Anderson (Eds.), *Handbook of distance education*. London: Lawrence Erlbaum Associates, 79-86.

Holmberg, B. (1999). The conversational approach to distance education. *Open Learning, 14(3),* 58-60.

Holmberg, B. (1989). *Theory and practice of distance education*. London: Routledge.

Howland, J.L., & Moore, J. L. (2002). Student perceptions as distance learners in Internet-based courses. *Distance Education, 23*(2), 183-195.

Jones, A., & Kemper, D. (1994). Approaches to learning and student acceptance of self-study packages. *Educational and Training Institute, 1*(2), 93-97.

Kanuka, H., Collett, D., & Caswell, C. (2002). University instructor perceptions of the use of asynchronous text-based discussion in distance courses. *The American Journal of Distance Education, 16*(3), 151-167.

Kennedy, D. M. (2002). Dimensions of distance: A comparison of classroom education and distance education. *Nurse Education Today, 22*(5), 409-416.

Knowles, M. S. (1980*). The modern practice of adult education: From pedagogy to andragogy.* Englewood Cliffs, NJ: Cambridge Adult Education, Prentice Hall Regents.

Knowles, M. S., Holton, E. F., & Swanson, R. A. (1998). *The adult learner* (5th ed.). Woburn, MA: Butterworth-Heinman.

Knowlton, D. S. (2000). A theoretical framework for online classroom: A defense and delineation of a student-centered pedagogy. In R.E. Weiss, D.S. Knowlton, & B.W. Speck(Eds.). *Principles* *of effective teaching in the online classroom, 84*. San Francisco: Jossey-Bass, 5-14.

Kozma, R. B. (1991). Learning with media. *Review of Educational Research, 61*(2), 179-211.

MacDonald, J. (2001). Exploiting online interactivity to enhance assignment development and feedback in distance education. *Open Learning, 16*(2), 181-189.

Rovai, A. P., & Barnum, K. T. (2003). Online course effectiveness: An analysis of student interactions and perceptions of learning. *Journal of Distance Education, 18*(1), 57-73.

Swan, K. (2001). Virtual interaction: Design factors affecting student satisfaction and perceived learning in asynchronous online courses. *Distance Education, 22*(2),306-331.

Salmon, G. (2000). *E-moderating: The key to teaching and learning online*. London: Kogan Page.

Townsend, E., Campbell, C., Curran-Smith, J., McGinn, F., Persaud, D., Peters, P., Bower, I., & Sheffield, S. (2002) Accessibility and interactivity in distance education programs for health professionals. *Journal of Distance Education, 17*(2), 1-24.

Rogers, C. (1969). *Freedom to learn*. Columbus, OH: Merrill.

Udod, S. A., & Care, W. D. (2002). Lessons learned in developing and delivering Web-based graduate courses: A faculty perspective. *Journal of Continuing Education in Nursing, 33*(1), 19-23.

Vandergrift, K. E. (2002). The anatomy of a distance education course: A case study analysis. *Journal of Asynchronous Learning Networks, 6*(1). Retrieved January 28, 2003 from, http://www.sloan-.org/publications/jaln/v6n1_vandergrift.asp

Visser, J. A. (2000). Faculty work in developing and teaching web-based distance courses: A case study of time and effort. *American Journal of Distance Education, 14*(3), 21-32.

Vrasidas, C., & McIsaac, M. S. (1999). Factors influencing interaction in an online course. *The American Journal of Distance Education 13*(3), 22-36.

Vrasidas, C. (1999). *Meanings of online and face-to-face interactions in a graduate course.* Unpublished Ph.D. Dissertation, Arizona State University.

Wlodkowski, R. J. (1999, 1985). *Enhancing adult motivation to learn*. San Francisco: Jossey-Bass Publishers.

Wright, T., & Thompson, L. (2002). Cost, access and quality in online nursing and allied health professions. *Journal of Asynchronous Networks 6*(2). Retrieved January 3, 2003 from, http://www.sloan-c.org/publications/jaln/v6n2_wright.asp

This work was previously published in International Journal of Information and Communication Technology Education, Vol. 4, Issue 1, edited by Lawrence A. Tomei , pp. 44-55, copyright 2008 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global

Chapter 5.7
Potential Benefits and Challenges of Computer-Based Learning in Health

Athina A. Lazakidou
University of Piraeus, Greece

Christina Ilioudi
University of Piraeus, Greece

Andriani Daskalaki
Max Planck Institute of Molecular Genetics, Germany

ABSTRACT

Computer-based learning has been developed for the beginning medical student and the experienced practitioner, for the lay person and the medical expert. There are many advantages to online and computer-based learning when compared to traditional face-to-face courses and lectures. Information technologies are providing new opportunities for linking medical schools around the world for sharing computer-based learning materials. In this chapter, the authors present examples of actual programs that are being used to support medical education for each of these categories of learners.

INTRODUCTION

Information technology is an increasingly important tool for accessing and managing medical information: both patient-specific knowledge and more general scientific knowledge. Medical educators are aware of the need for all medical students to learn to use information technology effectively. Computers can play a direct role in the education process; students may interact with educational computer programs to acquire factual information and to learn and practice medical problem-solving techniques. In addition, practicing physicians may use computers to expand and reinforce their professional skills throughout their careers (Shortliffe & Perreault, 2001).

The application of computer technology to education is often referred to as computer-assisted learning (CAL), computer-based education (CBE), or computer-aided instruction (CAI).

Computer-based learning has been developed for the beginning medical student and the experienced practitioner, for the lay person and the medical expert. In this article, we present examples of actual programs that are being used to support medical education for each of these categories of learners.

TYPES OF COMPUTER-BASED TRAINING

There are four levels of computer-based training (CBT), each based on the application's complexity and its level of interactivity with the user (Dulworth & Carney, 1996):

- **Level I. Customized Linear Presentation:** Training similar to a standard PowerPoint overhead presentation with little interactivity
- **Level II. Instructor-Led, Nonlinear Presentation:** Training by a facilitator accompanied by navigation through the information on a computer without the use of multimedia
- **Level III. Facilitator-Led Training:** A multimedia presentation accompanied by classroom-based training
- **Level IV. Self-Paced Training:** A multimedia presentation that trainees use independently with minimal assistance (also known as stand-alone training). Individuals can train at their own pace, either at an outside lab or on their own desktop computer, and complete the exam provided in the program.

Levels I, II, III, and IV are the types of computer-based training that would be most effective in addressing performance gaps among international health workers. To qualify for these levels, a computer-based training program must meet the following commonly accepted criteria (Dickelman, 1994).

- Be easy to enter and exit
- Provide a simple way to move forward and backward (i.e., from screen to screen)
- Be consistent in its key conventions
- Offer context-sensitive prompts and helps
- Provide tracking feedback (e.g., where have I been? Where am I now? How much more is there to go?)
- Offer bookmarks (i.e., quit now, resume later)
- Always offer a way out

COMPUTER-BASED TRAINING IN HEALTHCARE

In the health setting, CBT can be delivered in a preservice or in-service mode, as follows:

- **Preservice Training:** Computerized training delivered in health-education, nursing, and medical-school curricula through the use of software tutorials with or without professor facilitation, followed by examinations programmed in the computer program or given by an instructor
- **In-Service Training:** Health workers use CD-ROMs independently on their own computers for stand-alone training, meet at a computer lab where facilitator-led courses are coupled with the computer program, or attend the lab according to their own schedules and review the materials at their own pace

Research has shown that computer training is particularly well suited to visually intensive, detail-oriented subjects, such as anatomy and kinesiology. This is because it allows text to be combined with still and moving graphics, with the display of this information controlled by the learner (Toth-Cohen, 1995). For example, computers can be particularly effective in presenting the following (Phillips, 1996):

- Subjects that are difficult to conceptualize, such as microscopic processes
- Material that is three dimensional and difficult to visualize on traditional two-dimensional media such as books or whiteboards
- Simulations of expensive or complex processes, where the mechanical details of performing the process or the impossibility of using the real equipment may hinder understanding

BENEFITS OF COMPUTER-BASED LEARNING

Students may be learning more from using their computers than from attending lectures, according to a study published in July 2004 at studentbmj.com (http://www.studentbmj.com/back_issues/0800/news/265a.html).

Research psychiatrists at the School of Medicine, University of Leeds, compared students' use of a computer-based multimedia package with lecture-based teaching on the subject of anxiety. They found that even though students felt they learned more in the lecture theatre, they gained more from using a computer package.

There are many advantages to online and computer-based learning when compared to traditional face-to-face courses and lectures. There are a few disadvantages as well.

Main Advantages of Online or Computer-Based Learning

- Class work can be scheduled around work and family.
- It reduces the travel time and travel costs for off-campus students.
- Students may have the option to select learning materials that meet their level of knowledge and interest.
- Students can study anywhere they have access to a computer and an Internet connection.
- Self-paced learning modules allow students to work at their own pace.
- There is the flexibility to join discussions on the bulletin-board threaded discussion areas at any hour, or visit with classmates and instructors remotely in chat rooms.
- Instructors and students both report that e-learning fosters more interaction among students and instructors than in large lecture courses.
- E-learning can accommodate different learning styles and facilitate learning through a variety of activities.
- It develops knowledge of the Internet and computers skills that will help learners throughout their lives and careers.
- Successfully completing online or computer-based courses builds self-knowledge and self-confidence, and encourages students to take responsibility for their learning.
- Learners can test out of or skim over materials already mastered and concentrate their efforts on mastering areas containing new information and/or skills.

Main Disadvantages of Online or Computer-Based Learning

- Learners with low motivation or bad study habits may fall behind.
- Without the routine structures of a traditional class, students may get lost or confused about course activities and deadlines.
- Students may feel isolated from the instructor and classmates.
- The instructor may not always be available when students are studying or need help.

- Slow Internet connections or older computers may make accessing course materials frustrating.
- Managing computer files and online learning software can sometimes seem complex for students with beginner-level computer skills.
- Hands-on or lab work is difficult to simulate in a virtual classroom.

THE INTERNATIONAL VIRTUAL MEDICAL SCHOOL

The International Virtual Medical School (IVIMEDS, http://www.ivimeds.org) is a major international collaboration created to meet the challenge facing medical education through innovative approaches that exploit developments in educational thinking and information and communication technologies. Currently, 37 leading medical schools located in 14 countries have committed financial and human resources, and have agreed to share learning resources to make a reality of the IVIMEDS vision.

The International Virtual Medical School provides the following:

- A comprehensive medical-education resource, the Medical Education Service, available to teachers and learners worldwide. It will provide users with educationally and technologically state-of-the-art medical-education resources and services at low cost.
- A cost-effective alternative-track curriculum for undergraduate medical students for the early years of the undergraduate program, and the right to complete the clinical stages of their training in a partner medical school.
- Customized postgraduate and continuing professional development (CPD) programs that can be taken at the time and place of choosing of the learner, thereby facilitating convenient and cost-effective lifelong learning.
- Customised medical and multiprofessional health education through the IVIMEDS Foundation appropriate to the needs and circumstances of developing countries in regard to curriculum, localization, language, and mode of delivery.

Benefits to Partner Institutions

IVIMEDS offers membership to an international network of partner institutions sharing resources to enrich individual member curriculums and to enhance the ability to deliver cost effectively high-quality medical-education programs. Benefits include the following:

- Improved finances and assets by providing the means to develop an enhanced curriculum with additional topics and approaches to learning, open access to medical training for students of different backgrounds, and a global market for homegrown educational resources and strategies.
- Rapid, effective execution of new approaches that may be beyond the budget and scope of any one institution, drawing upon a global body of expertise in subject matter, educational theory, and technology.
- Reduced risk associated with curriculum changes by sharing innovative thinking and benefiting from other schools' experiences in curriculum planning and their use of learning technologies and learning-management systems.
- Quality resources and innovative approaches to medical education, which can contribute to the creation of a curriculum that is coherent, integrated, student centered, and authentic.

Benefits to Students, Trainees, and Practicing Doctors

IVIMEDS will provide an innovative curriculum and/or curriculum elements tailored to the changing educational needs of students, trainees, and medical professionals. Benefits include the following:

- A learner-centered approach with students exposed to just-in-time learning with theory closely linked to practice.
- Adaptive learning or "just-for-you" learning catering to individual learning styles and interests.
- Curriculum frameworks provided by established learning outcomes, a broad curriculum map, and a bank of virtual patients.
- Blended learning including anytime, anywhere electronic study guides, face-to-face and online tutors, and peer-to-peer learning.
- Flexible learning based in a variety of settings (e.g., a university teaching hospital, a local health centre, or home-based study), best suited to the financial, personal, and educational circumstances of individual students.

Benefits to Society

IVIMEDS offers an approach to medical education and training that is both adaptable and cost effective. Benefits to society include the following:

- Flexibility to expand and contract numbers of learners to meet changing circumstances.
- Wider access to medical education for students, including disadvantaged and mature students.
- Training doctors to focus on the needs of particular communities, with the potential that qualified doctors will return to work in these communities.
- Training doctors with an appropriate high level of competence in information handling and an aptitude for self-directed learning and continuing professional development.
- Cost-effective training with schools working together to blend e-learning with face-to-face learning in a variety of educational and clinical settings.

CONCLUSION

We recognize that technology impacts health-care-education, research, and science educators in the areas of research, classroom teaching, and distance education. While the overall effect is not yet fully assessable, the presence of technology in so many different aspects of the profession makes it important to more clearly recognize and appreciate its current and potential role.

Information technologies can be educators' tools in finding creative ways that encourage students to self-test, self-question, and self-regulate learning and in helping them to create solutions to complex problems. Information technologies are providing new opportunities for linking medical schools around the world for sharing computer-based learning materials. They open a wide horizon for acquiring and expending medical knowledge originated in any part of the world without the limitations of time, space, or distance.

The use of computers and information technology in medical education should be regarded as an additional tool and must never be a goal in itself but part of flexible learning. On the contrary, clinical medical education should always be centered on direct patient contact and bedside education. While we urge for direct patient contact, we believe that using stimulations would also benefit the student in training.

REFERENCES

Dickelman, G. J. (1994). Designing and managing computer-based training for human resource development. In C. E. Schneier, C. J. Russell, R. W. Beatty, & L. S. Baird (Eds.), *The training and development sourcebook*. Amherst, MA: Human Resource Development Press.

Dulworth, M. R., & Carney, J. (1996). *Improve training with interactive multimedia: Info-line 9601.* Alexandria, VA: American Society for Training and Development.

Evans, L. A., Brown, J. F., & Heestand, D. E. (1994). Incorporating computer-based learning in a medical school environment. *Journal of Biocommunication, 21*(1), 10-17.

Henry, S. B., & Waltmire, D. (1992). Computerised clinical simulations: A strategy for staff development in critical care. *American Journal of Critical Care Nursing*, *1*(2), 99-107.

Jaffe, C. C., & Lynch, P. J. (1993). Computers for clinical practice and education in radiology. *Radiographics, 13*(4), 931-937.

Johannson, S. L., & Wertenberger, D. H. (1996). Using simulation to test critical thinking skills of nursing students. *Nurse Education Today, 16*, 323-327.

Lauri, S. (1992). Using a computer simulation program to assess the decision making process in child health care. *Computers in Nursing*, *10*(40), 171-177.

Lauriland, D. (1995). Multimedia and the changing experience of the learner. *The British Journal of Educational Technology*, *26*(3), 179-189.

Phillips, R. (1996). *Developers guide to interactive multimedia: A methodology for educational applications.* Perth, WA: Curtin University.

Romiszowski, A. (1994). Individualization of teaching and learning: Where have we been? Where are we going? *Journal of Special Education Technology, 2,* 182-194.

Shortliffe, E. H., & Perreault, L. E. (2001). *Medical informatics: Computer applications in health care and biomedicine* (2nd ed.). New York: Springer.

Toth-Cohen, S. (1995). Computer-assisted instruction as a learning resource for applied anatomy and kinesiology in the occupational therapy curriculum. *American Journal of Occupational Therapy*, *49*(8), 821-827.

Further Reading

http://www.asme.org.uk/

http://www.cdlhn.com

http://www.chime.ucl.ac.uk/

http://www.dso.iastate.edu/dept/asc/elearner/advantage.html

http://www.emedicine.com

http://www.health.state.mn.us/divs/hrm/dl/compbased.html

http://www.ifmsa.org/partners/wfme_he.htm

http://www.interactive-designs.com/cbl1.htm

http://www.ivimeds.org

http://www.lib.uiowa.edu/commons/cbl.html

http://www.med.cam.ac.uk/html/teaching/DepMed/Phase1/computer.html

http://www.personal.dundee.ac.uk/~cdvflore/

http://www.qaproject.org/pubs/PDFs/researchcbtx.pdf

http://www.sph.umn.edu/publichealthplanet

KEY TERMS

Computer-Aided Instruction (CAI): The application of computer technology to education (also called computer-assisted learning and computer-based education).

Computer-Assisted Learning (CAL): The application of computer technology to education (also called computer-aided instruction and computer-based education).

Computer-Based Education (CBE): The application of computer technology to education (also called computer-assisted learning and computer-aided instruction).

Multimedia Content: Information sources that encompass all common computer-based forms of information, including texts, graphics, images, video, and sound.

Simulation: A system that behaves according to a model of a process or another system; for example, a simulation of a patient's response to therapeutic interventions allows a medical or nursing student to learn which techniques are effective without risking human life.

Tutoring System: A computer program designed to provide self-directed education to a student or trainee.

This work was previously published in Handbook of Research on Informatics in Healthcare and Biomedicine, edited by A. A. Lazakidou, pp. 119-126, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 5.8
Teaching Medical Statistics Over the Internet

Rachael Knight
Royal Women's Hospital, Australia

Kate Whittington
University of Brisol, UK

W. Chris L. Ford
University of Bristol, UK

Julian M. Jenkins
University of Bristol, UK

INTRODUCTION

The potential for computers to assist learning has been recognised for many years (Jenkins, 1997), with reproductive medicine benefiting greatly from Internet technology (Jenkins, 1999). Following a detailed survey of information technology facilities and skills for postgraduate education (Draycott, 1999), a pilot Internet training programme in reproductive medicine demonstrated effective methods to deliver online teaching (Jenkins, 2001). Based on this experience, in 2001 the Obstetrics and Gynaecology in the University of Bristol, U.K. launched a postgraduate masters course in human reproduction and development, delivered principally over the Internet (Jenkins, 2002). This course has been under continuous evaluation and development since its launch, refining the application of learning technology to most appropriately meet students' needs (Cahill, 2003). A particularly challenging module of the course considers research methods and statistics. This module was independently evaluated from both a student and tutor perspective, with the objective of identifying learning priorities and optimal educational methodology. This article presents strengths and weaknesses of delivering statistics education online, considering how best to develop this in the future.

BACKGROUND

Teaching statistics to medical and allied healthcare undergraduates and post-graduates has been a challenging area in higher education (Marantz, 2003). Traditionally taught early in the medical course by colleagues from the mathematics department, basic statistical principles are not retained by many medical graduates. A number of reasons have been postulated for this, including the timing of the teaching and lack of practical implications at this stage of training, and lack of motivation among the students (Marantz, 2003; Romero, 2000; Astin, 2002).

As medical students progress through undergraduate education to post-graduate education, the requirement for medical statistics knowledge increases and ranges from day-to-day needs to critique published literature to the ability to design and evaluate research proposals. The knowledge from undergraduate courses is often inadequate to deal with the increased responsibility, and statistics update courses are difficult to attend while working full time as a clinician (Astin, 2002). Many medical statistics textbooks are available; however, they cannot be individually tailored to meet each student's needs. Many candidates report medical statistics as a difficult subject to learn, and the students notoriously lack motivation with regards studying (Romero, 2000). The practical aspects are not always relevant to the clinical aspects of medicine being studied; thus, retention of knowledge is poor (Astin, 2002). Evaluations of the current methods of teaching medical statistics in undergraduate curriculum have focused upon the need for clinical relevance when teaching at earlier levels. Courses based around data analysis have been criticised, while a greater emphasis upon critical appraisal and data presentation has been recommended (Romero, 2000; Astin, 2002; Bruce, 2002). One course reported improved student preparation and participation when a case discussion method was employed to teach epidemiology and bio-statistics (Marantz, 2003).

The Internet provides a comprehensive range of statistics resources that could be used to support an Internet biomedical statistics training programme. Table 2 contains many of the more popular resources, but is by no means an exhaustive list. To highlight several of the Web sites: Statistics at Square One is an online version of a statistics text available via the British Medical Journal Web site. This is an easy-to-read text regarding basic statistics, and has clinical examples at the end of each chapter. Clara is a Web-based computer statistics program that enables the user to perform basic statistical calculations. This assumes some knowledge with regard to the most appropriate test for the data set. Although the scope of this article does not permit discussion of the many other Internet statistics resources, Table 2 provides a guide to many Web sites that may be of interest.

BRISTOL APPROACH

The University of Bristol course aimed to address some of the above challenges of delivering biomedical statistics by providing more student-centered learning delivered over the Internet, with emphasis on clinically relevant practical exercises. This section presents a course evaluation of the Bristol approach to date, identifying the problems that need to be considered with suggestions for future research and possible solutions for Web-based training.

Methods

In 2003, during the second year of the Reproduction & Development MSc in the University of Bristol, an online survey was completed by 18 of the 20 registered students who had completed the statistics and research methods module. Prior to starting the course, students had been sent a copy of the statistical computer programme Stats Direct (wwww.Statsdirect.com) with explanatory

Figure 1. Student responses regarding online teaching in general (n =18)

Student characteristics	Total respondents N=18 n (%)	Part-time subgroup N=14	Clinician subgroup N=12
Full time	4 (22%)	-	2 (17%)
Part time	14 (78%)	-	10 (83%)
Clinicians	12 (67%)	10 (71%)	-
Scientists	6 (33%)	4 (29%)	-
Male	5 (28%)	5 (36%)	5 (42%)
Female	13 (72%)	9 (64%)	7 (58%)
UK based	11 (61%)	9 (64%)	6 (50%)
Overseas based	7 (39%)	5 (36%)	6 (50%)

reference material, and were advised to complete simple tasks to familiarise themselves with statistics prior to a workshop. During the workshop, students were taught by lectures in small groups and completed specific exercises. Following the workshops, further information was delivered online, with interactive questions provided to allow reflection on workshop learning and to develop and test their understanding of statistics. Assessed practical exercises provided an opportunity to apply their knowledge to relevant tasks. Following the period of Internet teaching, students returned for a formal examination by computer-marked assessments, short answers and data analysis exercises.

The online student survey evaluated the distance education component of the course, with particular emphasis on the teaching of the statistics and research methodology module. The students were presented with statements about online learning and asked to respond by grading their views using: strongly agree, agree, neutral, disagree and strongly disagree. Statements specifically targeting the statistics module were

Figure 1. Student responses regarding online teaching in general (n =18)

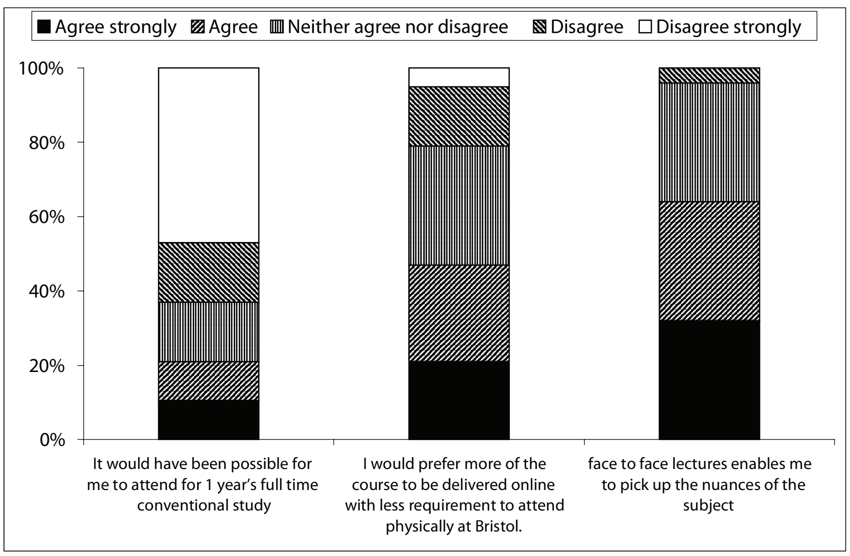
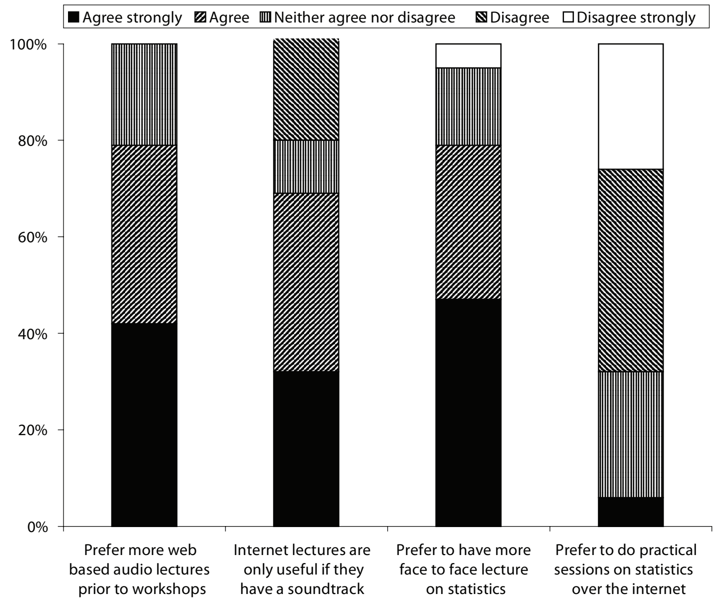


Figure 2. Student responses regarding Internet-based teaching of statistics (n=18)



included, and students were asked to prioritise the statistics lectures and tutorials.

Results from a course appraisal of a distance learning statistics module for medical and paramedical postgraduates

The demographics of the students on the course are indicated in Table 1. When appraising Internet-based vs. traditional face-to-face learning, students reported that it was not possible in the majority of cases to attend a conventional lecture-based university course. Seventy-nine percent of part-time students felt they would not have been able to attend the course if it was arranged as a conventional one-year campus-based study (see Figure 1). Most students reported they could learn efficiently from audio-Web-based lectures and would prefer more audio-Web lectures to be delivered prior to the attending workshops. In contrast, when students were specifically asked about Internet-based learning of statistics, they reported that they would prefer more face-to-face lectures and less Internet-based practical sessions (see Figure 2). They reported that face-to-face lectures allowed them to pick up the nuances of the subject and thus improve their understanding.

Students preferred to have lectures on basic statistical concepts; design of clinical trials; calculating means, medians and confidence intervals; and which statistical test to apply (see Figure 3). Lectures on survival analysis, regression, correlation and advanced statistics were found to be least useful. The overall preference was to have face-to-face lectures for basic statistical principles and common statistical calculations, with less emphasis on more advanced statistics such as regression analysis. Students also appreciated lectures and tutorials on practical aspects of statistics that would be useful for their career and ongoing medical education, such as the design of clinical research, determining which statistical test is most appropriate for a given data set and how to use a computer-based statistics program.

Table 2. Web-based statistics resources

Statistics Resources on the Web
www.stats.gla.ac.uk/cti/links_stats/index.html
Links to online statistics journals, textbooks, university departments and newsgroups. Compiled by Glasgow University.

Bibliography for Computational Probability and Statistics
http://ubmail.ubalt.edu/~harsham/statistics/REFSTAT.HTM
Thorough Web site with many resources, including sites for computational probability and statistics, as well as links to journals, books and relevant societies and organisations. Compiled by Professor Hossein Arsham, University of Baltimore.

BUBL LINK / 5:15 Internet Resources
http://bubl.ac.uk/link/mat.html
Selected Internet resources for academics. Extensive range in all areas, including statistics methods, education and research. Compiled by Andersonian Library, Strathclyde University, 101 St James Road, Glasgow

SOSIG: Statistics
www.sosig.ac.uk/statistics/
Statistics Tools for Internet and Classroom. An extensive list of resources including statistics societies, databases, books, companies, journals, educational material, government publications, software link, papers and reports. Compiled by Southampton University Library.

Darren's WWW links: Stats
www.mas.ncl.ac.uk/~ndjw1/bookmarks/Stats/
List of Internet links for statistics students, including sites on bayesian statistics and bioinformatics, catalogues, conferences, groups and societies, software and statistical teaching sites.
Compiled by Dr. D. Wilkinson Senior Lecturer in Statistics. School of Mathematics and Statistics, University of Newcastle upon Tyne.

Mathematics Department Library: Internet Resources in Mathematics
www.ma.ic.ac.uk/library/ires.html#stat
Provides links to several Internet statistics resources. Compiled by Mathematics Department Library, Imperial College.

Statlib
http://lib.stat.cmu.edu/
This is a Web site for distributing statistical software, datasets and information by electronic mail, FTP and WWW. Created by Mike Meyer, Department of Statistics, Carnegie Mellon University.

University of Ulster Library - Internet Resources for Statistics
www.ulst.ac.uk/library/soc/statnet.htm
This guide provides links to Internet Resources for Statistical Information. Contains comprehensive international links to statistics database sites. Compiled by Ulster University Library Services.

Applied statistics : a brief guide to Internet resources
www.library.rdg.ac.uk/subjects/ir/irappstat.html
A guide to some useful resources available on the Internet in applied statistics, including: information gateways, databases, organisations, museums, e-journals and bureaux of statistics. Compiled by the University of Reading Library.

Statistical Services Internet Statistics Resources
www.utexas.edu/cc/stat/world/
This Web site classifies statistical and statistical software information into several broad categories to help students find an appropriate resource. Includes links to online calculators and data analysis. Developed by the University of Texas Information Technology Service.

Statistics and Statistical Graphics Resources
www.math.yorku.ca/SCS/StatResource.html
This Web site provides an annotated, topic-based collection of available resources for statistics, statistical graphics and computation related to research, data analysis and teaching. It contains more than 580 links. Compiled by Professor M. Friendly, Statistical Consulting Service, York University, Toronto, Canada.

Statistics.com
www.statistics.com/content/about.html
A directory of links to data sources for data analysts, researchers, educators, managers and students. Links to commercial online courses, software and journals.

continued on next page

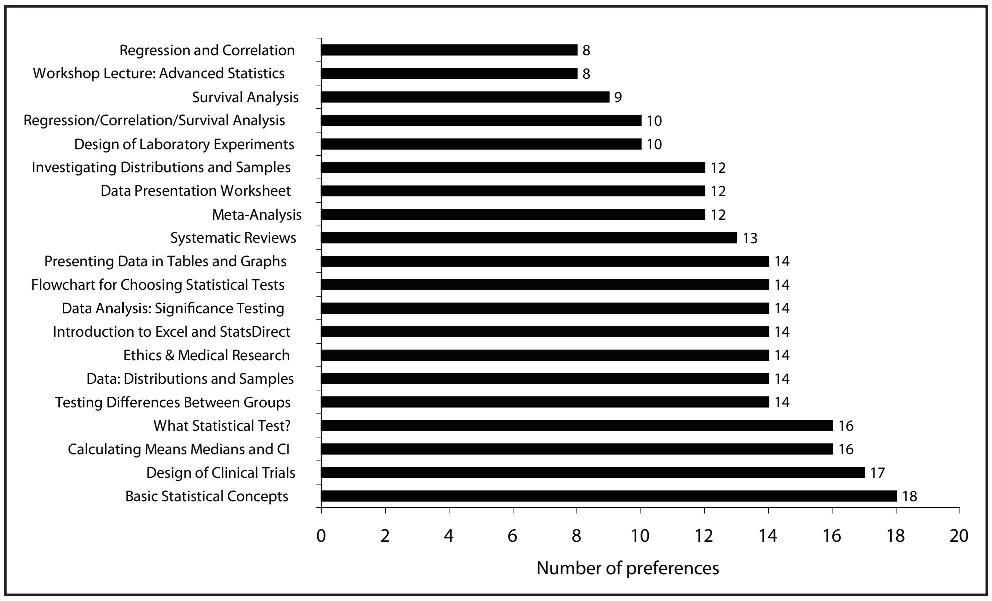
Table 2. continued

UCLA Department of Statistics www.stat.ucla.edu/ This Web site has links to online statistical calculators and probability and statistics journals. The Department of Statistics, UCLA, developed it.
St@tServ www.statserv.com/ This Web site markets itself as an 'Internet statistical community.' It has links to statistics books and journal sites, software and consulting services, statistics organisations and employment services.

FUTURE TRENDS

This survey demonstrates students' desires for more Internet teaching to allow a larger proportion to participate in additional education, but illustrates both the need to support this with workshops and the particular challenges of teaching statistical concepts online. From the recent survey, the Internet approach initially adopted in Bristol did not overcome the problems associated with teaching statistics, as students felt they preferred more face-to-face lectures and more supervision of practical sessions. This highlights that teaching difficult concepts using a Web-based approach can be challenging. This may be explained by students' needs to attend lectures for difficult concepts to pick up the finer points of the subject and to ask questions of the lecturer. Students may also perceive the need for increased assistance in what is traditionally a difficult area, and they may have a lack of self-motivation with respect to this topic, which Internet-based lectures do not overcome.

Figure 3. Summation of lectures included by individual student preferences for top 10 most useful lectures in statistics



In part, this assistance may be addressed through structured online discussion bulletin boards, providing the opportunity to explore the concepts with views from both tutors and fellow students. The use of online discussion bulletin boards requires further research. The preference for lectures that cover basic statistical concepts; design of clinical trials; calculating means, medians and confidence interval; and learning which statistical test to apply indicates that students are looking for a course with more day-to-day, practical emphasis. They want techniques to use when appraising the medical literature and reviewing and designing audit and research. More complicated medical statistics were less popular, such as advanced statistics, correlation and regression, and survival analysis. It may be prudent to design a course emphasising the skills needed on a day-to-day basis, with more advanced concepts defined but not discussed in great detail. A second level of the course could be designed around the teaching of advanced statistical concepts, if this was required.

CONCLUSION

Current postgraduate students are demanding courses adaptable to their lifestyle and more clinically relevant. Geographical location and work commitments involving long hours combined with an increased knowledge and skill of Internet and computer technology, particularly among younger students, places increased challenges for course providers. Thus, courses that place more emphasis on distance learning are becoming increasingly popular, and the Reproduction and Development MSc course at the University of Bristol is now extremely popular, attracting an international student group (Whittington, 2004). Statistics has always been a difficult subject area to teach, and the results of our survey may be explained by students' needs for a practical-based statistics course with emphasis on basic concepts and skills needed for literature appraisal. From the results presented above, we conclude that for postgraduate students distance learning packages are very popular and useful; however, teaching difficult concepts online is more challenging for course organisers and requires careful development.

REFERENCES

Astin, J., Jenkins, T., & Moore, L. (2002). Medical students' perspectives on the teaching of medical statistics in the undergraduate medical curriculum. *Statistics in Medicine*, *15,* 1003-1006.

Bruce, J., Bond, S., & Jones, M., (2002). Teaching epidemiology and statistics by distance learning. *Statistics in Medicine*, *15*, 1009-1020.

Cahill, D.J., Cook, J., & Jenkins, J.M. (2003). How useful are World Wide Web Discussion Boards and email in delivering a Case Study Course in Reproductive Medicine. In C. Ghaoui (Ed.), *Usability Evaluation of Online Learning Programs* (pp. 360-370). Information Science Publishing.

Draycott, T.J., Cook, J., Fox, R., & Jenkins, J. (1999). Information technology for postgraduate education; a survey of facilities and skills in the South West Deanery. *British Journal of Obstetrics & Gynaecology: An International Journal of Obstetrics and Gynaecology*, *106,* 731-735.

Jenkins, J., Cook, J., Edwards, J., Draycott, T., & Cahill, D. (2001). A pilot Internet training programme in Reproductive Medicine. *British Journal of Obstetrics & Gynaecology: An International Journal of Obstetrics and Gynaecology*, *108*,114-116.

Jenkins, J.M. (1997). Computer assisted learning. In E.M. Symonds (Ed.), *Current Obstetrics & Gynaecology Churchill Livingstone* (7, pp. 139-144).

Jenkins, J.M. (1999). The Internet, Intranets and Reproductive Medicine. *Human Reproduction*, *14*, 586-589.

Jenkins, J.M. (2002). ReproMED on the Internet today and tomorrow. *Human Fertility, 5*(1 Suppl), S66-71.

Marantz, P., Burton, W., & Steiner-Grossman, P. (2003). Using the case-discussion method to teach epidemiology and biostatistics. *Academic Medicine, 78,* 365-371.

Romero, V., Berger, D., Healy, M., & Aberson, C. (2000). Using cognitive learning theory to design effective on-line statistics tutorials. *Behaviour Research Methods, Instruments, & Computers, 32*, 245-249.

Whittington, K., Cook, J., Barratt, C., & Jenkins, J. (2004). Can the Internet widen participation in reproductive medicine education for professionals? *Human Reproduction*, *19*, 1800-1805.

KEY TERMS

Biomedical: Relating to biomedicine, the application of natural sciences, especially biology and physiology, to clinical medicine.

Cognitive Learning Theory: This provides a framework to understand learning, suggesting that although learning is not directly observable it occurs through active mental processes where knowledge is progressively assimilated, making a change in behaviour possible.

Distance Education: Students learning remotely from their educational establishment, supported by a variety of methods.

Reproductive Medicine: Speciality of medicine dealing with the reproductive system and related medical issues such as infertility, contraception, menopausal problems and menstrual dysfunction.

Statistics: Branch of mathematics concerned with the collation, analysis and interpretation of quantitative data.

Self-Directed Learning: Students are empowered to learn at their own pace, catering to the different learning speeds and styles of individuals.

Web-Based Training: Use of material delivered via Web browser to support education. This may support distance learning or can be used within an educational establishment.

This work was previously published in Encyclopedia of Distance Learning, Vol. 4, edited byC. Howard, J. Boettcher, L. Justice, K. Schenk, P.L. Rogers, and G.A. Berg, pp. 1770-1776, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 5.9
Tropical Medicine Open Learning Environment[1]

Geraldine Clarebout
University of Leuven, Belgium

Jan Elen
University of Leuven, Belgium

Joost Lowyck
University of Leuven, Belgium

Jef Van den Ende
Institute for Tropical Medicine, Belgium

Erwin Van den Enden
Institute for Tropical Medicine, Belgium

INTRODUCTION

Educational goals have generally shifted from knowing everything in a specific domain to knowing how to deal with complex problems. Reasoning and information-processing skills have become more important than the sheer amount of information memorized. In medical education, the same evolution has occurred. Diagnostic reasoning processes get more strongly emphasized. Whereas previously knowing all symptoms and diseases was stressed, reasoning skills have now become educationally more important. They must enable professionals to distinguish between differential diagnoses and to recognize patterns of illnesses (e.g., Myers & Dorsey, 1994).

BACKGROUND

Authentic or realistic tasks have been advocated to foster the acquisition of complex problem-solving processes (Jacobson & Spiro, 1995; Jonassen, 1997). In medical education, this has led to the use of expert systems. Such systems were initially developed to assist practitioners in their practice (NEOMYCIN, in Cormie, 1988; PATHMASTER in Frohlich, Miller & Morrow, 1990; LIED in Console, Molino, Ripa di Meana & Torasso, 1992) and simulate real situations. These systems were expected to provoke or develop students' diagnostic reasoning processes. However, the implementation of such expert systems in regular educational settings has not been successful.

Instead of developing reasoning processes, these systems assume them to be available. They focus on quickly getting to a solution rather than reflecting on possible alternatives. Consequently, it was concluded that students need more guidance in the development of diagnostic reasoning skills (Console et al., 1992, Cromie, 1988; Friedman, France & Drossman, 1991), and that instructional support was lacking.

KABISA is one of the computer programs purposely designed to help students in the development of their diagnostic reasoning skills (Van den Ende, Blot, Kesten, Van Gompel & Van den Enden, 1997). It is a dedicated computer-based training program for acquiring and optimizing diagnostic reasoning skills in tropical medicine.

DESCRIPTION OF THE PROGRAM[2]

KABISA confronts the user with cases, or 'virtual patients'. The virtual patient is initially presented by three 'characteristics'[3] randomly selected by the program. After the presentation of the patient (three characteristics), students can ask for additional characteristics gathered through anamnesis, physical examination, laboratory and imaging. If students click on a particular characteristic, such as a physical examination test, they receive feedback. Students are informed about the presence of a certain disease characteristic, or whether a test is positive or negative. If students ask a 'non-considered' characteristic; that is, a characteristic that is not relevant or useful in relation to the virtual patient, they are informed of this and asked whether they want to reveal the diagnosis they were thinking about. When they do so, students receive an overview of the characteristics that were explained by their selection and which ones are not, as well as the place of the selected diagnosis on a list that ranks diagnoses according to their probability given the characteristics at hand. If students do not want to show the diagnosis they were thinking about they can just continue asking for characteristics. A session is ended with students giving a final diagnosis. KABISA informs them about the correctness. If the diagnosis is correct, students are congratulated. If the diagnosis is not correct, students may be informed that it is a very plausible diagnosis but that they do not have enough evidence, or they may get a ranking of their diagnosis and an overview of the disease characteristics that can and cannot be explained by their answer. Additionally, different non-embedded support devices – that is, tools are made available to support learners. These tools allow students to look for information about certain symptoms or diseases, to compare different diagnoses or to see how much a certain characteristic contributes to the certainty for a specific diagnosis. Students decide when and how they use these devices (for a more detailed description, see Clarebout, Elen, Lowyck, Van den Ende & Van den Enden, 2004).

FUTURE TRENDS AND CRITICAL ISSUES

KABISA is designed as an open learning environment (Hannafin, Hall, Land & Hill, 1994); that is, students are confronted with a realistic and authentic problem, there is a large amount of learner control and tools are provided to guide students' learning. However, the evaluation study performed revealed some interesting issues. A first revelation was that students do not follow a criterion path when working on KABISA. Prior to the evaluation, two domain experts in collaboration with three instructional designers constructed a criterion path. This path represented the ideal paths students should go through to optimally benefit from KABISA (following the "normative approach" of Elstein & Rabinowitz, 1993), including when to use a specific tool. Only 5 out of 44 students followed this path.

A second issue relates to the use of the tools. KABISA offers different tools to support students. These tools can help students in their problem-solving process. Results suggest that students consult some help functions more than others. However, overall they do not consult them frequently, and if they use them, they do not use them adequately. Students also tend to not use the feedback that can be obtained when asking for a 'non-considered' characteristic.

Although this environment can be described as an open learning environment, it seems that students do not perceive it as a learning environment. Thinking aloud protocols reveal that students think they are cheating or failing when consulting a tool. Giving the limited use of the tools, their impact on the learning process cannot be but limited.

However, in spite of the observation that in only a small number of consultations the criterion path was followed, students do find the right diagnosis in 80% of the consultations. It seems that by trial and error students can also obtain the right diagnosis.

The results of this evaluation suggest that KABISA is currently not used by students to foster their diagnostic reasoning skills. Rather, it enables them to train readily available skills. The results are an example of a well-designed learning environment used by learners as a performance environment. With the use of more open and online learning environments, this raises the issue of how to realize that students see such environments as learning environments with learning opportunities, rather than a performance environment.

In the design and development of KABISA, a lot of time and effort was spent in developing the tools. However, results show that students do not (adequately) use these tools. Other authors have found similar results with other programs (e.g., see Crooks, Klein, Jones & Dwyer, 1996; Land, 2000). This raises questions about the amount of learner control in open learning environments. Should the environment be made less open and provide embedded support devices instead of tools, so that students cannot but use these devices? Or should students receive some additional advice towards the use of these tools? In the first case, support might not be adapted to the learners' needs. This might cause problems, given that either too much or too little support can be detrimental (Clark, 1991). The second option leaves the environment open. But here also, it can be questioned whether this advice should not be adapted to the learners' needs. A possible solution might come out of the animated pedagogical agent-research. These agents are animated figures that aim at helping learners in their learning process and adapt their support based on the paths learners follow (Shaw, Johnson & Ganeshan, 1999). Certainly in online learning environments where there is by definition a large extent of learner control, such agents might provide a solution, and act even as a personal coach for learners.

Another aspect revealed by the analysis is the importance of an evaluation phase in the development of computer-based training programs. For instance, a more thorough analysis of student characteristics could have provided a means to adapt the difficulty level to the level of the students or to identify what guidance students actually need. Apparently, the feedback given to students does not encourage them to adapt their problem-solving process. Being product- rather than process-oriented, feedback may not be adapted to students' actual needs. Likewise, students' instructional conceptions about computer-based learning environments or their perceptions about KABISA (game vs. educational application) may influence the use of the program. Students' instructional conceptions should be taken into account through the design process of the program. One possible way to influence these conceptions might be through an introduction of the program. In the introduction, the aims of the program, the different functionalities and the relationship with the different courses should be clearly defined (see Kennedy, Petrovi & Keppell, 1998).

Given the difficulty of anticipating potential problems and difficulties students might encounter in open learning environments, the introduction of a formative evaluation during the design and development of this program seems warranted. This would enable the redirection of the program while developing it, rather than after implementation. Rather than only evaluating a final product, the development process should be taken into consideration, as well. Rapid prototyping would allow for testing the program at different phases of the development. This leads to a more spiral cycle rather than a linear design process.

CONCLUSION

The evaluation of KABISA addressed some general issues important to consider in the design, development and implementation of open learning environments. Although these environments are advocated to foster the acquisition of complex problem-solving skills, there seems to be a gap between the intention of the designers and the use by the learners. This relates to the issue addressed by Winne and Marx (1982) about calibration. For an instructional intervention to be effective, calibration is needed between the conceptions of the various people involved. The introduction of a pedagogical agent might help to calibrate the conceptions of students towards those of the designers. Moreover, these agents might help in encouraging students to adequately use tools without reducing the openness of the learning environment.

REFERENCES

Clarebout, G., Elen, J., Lowyck, J., Van den Ende, J., & Van den Enden, E. (2004). KABISA: Evaluation of an open learning environment. In A. Armstrong (Ed.), *Instructional design in the real world: A view from the trenches* (pp. 119-135). Hershey: Idea Group.

Clark, R.E. (1991). When teaching kills learning: Research on mathemathantics. In H. Mandl, E. De Corte, N. Bennett, & H.F. Friedrich (Eds.), *European research in an international context: Volume 2. Learning and Instruction* (pp. 1-22). Oxford: Pergamon Press.

Console, L., Molino, G., Ripa di Meana, V., & Torasso, P. (1992). LIED-liver: Information, education and diagnosis. *Methods of Information in Medicine, 31*, 284-297.

Cromie, W.J. (1988). Expert systems and medical education. *Educational Researcher, 17*(3), 10-12.

Elstein, A.S., & Rabinowitz, M. (1993). Medical cognition: Research and evaluation. In M. Rabinowitz (Ed.), *Cognitive Science Foundation of Instruction* (pp. 189-201). Hillsdale: Lawrence Erlbaum Associates.

Friedman, C.P., France, C.L., & Drossmann, D.D. (1991). A randomized comparison of alternative formats for clinical simulations. *Medical Decision Making, 11*(4), 265-271.

Frohlich, M.W., Miller, P.L., & Morrow, J.S. (1990). PATHMASTER: Modelling differential diagnosis as "Dynamic Competition" between systematic analysis and disease-directed deduction. *Computers and Biomedical Research, 23*, 499-513.

Hannafin, M.J., Hall, C., Land, S., & Hill, J. (1994). Learning in open-ended learning environments: Assumptions, methods and implications. *Educational Technology, 34*(10), 48-55.

Jacobson, M.J., & Spiro, R.J. (1995). Hypertext learning environments, cognitive flexibility and the transfer of complex knowledge. *Journal of Educational Computing Research, 12*(4), 301-333.

Jonassen, D.H. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology Research and Development, 45*(1), 65-91.

Kennedy, G., Petrovic, T., & Keppell, M. (1998). The development of multimedia evaluation criteria and a program of evaluation for computer aided learning. In R.M. Cordeory (Ed.), *Proceedings of the Fifteenth Annual Conference of the Australian Society for Computers in Tertiary Education (ASCILITE)* (pp. 407-415). Wollongong: University of Wollongong.

Myers, J.H., & Dorsey, J.K. (1994). Using diagnostic reasoning (DxR) to teach and evaluate clinical reasoning skills. *Academic Medicine, 69*, 429.

Shaw, E., Johnson, W.L., & Ganeshan, R. (1999). Pedagogical agents on the Web. *Proceedings of the Third International Conference on Autonomous Agents*, 283-290.

Van den Ende, J., Blot, K., Kestens, L., Van Gompel, A., & Van den Ende, E. (1997). KABISA: An interactive computer-assisted training program for tropical diseases. *Medical Education*, *31*, 202-209.

Winne, P.H., & Marx, R.W. (1982). Students' and teachers' view of thinking processes for classroom learning. *The Elementary School Journal*, *82*, 493-518.

KEY TERMS

Animated Pedagogical Agents: Animated figures operating in a learning environment and aiming at supporting learners in their learning process and capable of adapting their support to the learners' paths.

Criterion Path: This a representation of an 'ideal path' to go through a specific learning environment. It specifies for each possible step in the program the most ideal subsequent steps.

Embedded Support Devices: These are support devices integrated in the learning environment. Learners cannot but use these devices (e.g., structure in a text).

Instructional Conceptions: These are conceptions about the functionalities (elements) of a learning environment. These conceptions can relate to the effectiveness or efficiency of specific features in a learning environment (e.g., tools) or to the environment as a whole (e.g., KABISA as a learning environment).

Non-Embedded Support Devices (Tools): These are support devices put to the disposal of learners. Learners decide when and how to use these tools.

Open-Ended Learning Environments: Aim at fostering complex problem-solving skills by confronting learning with a realistic or authentic problem in a learning environment with a large amount of learner control and different tools.

Perceptions: Students' perceptions relate to how they perceive a specific environment (KABISA). They are the results of an interaction between students' instructional conceptions and a specific learning environment.

END NOTES

[1] A more extended version of this manuscript with more background information was published in Armstrong, A. (Ed.). (2004). *Instructional design in the real world. A view from the trenches.* Hershey: Information Science Publishing.

[2] The authors are grateful to Stefano Laganà, who invested a lot of effort to adapt KABISA and to develop a log file system.

[3] The term 'characteristic' refers to either a symptom or disease characteristics, either a request for results of a physical examination,

This work was previously published in Encyclopedia of Distance Learning, Vol. 4, edited by C. Howard, J. Boettcher, L. Justice, K. Schenk, P.L. Rogers, and G.A. Berg, pp. 1902-1906, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global)

Chapter 5.10
Gastrointestinal Motility Online Educational Endeavor

Shiu-chung Au
State University of New York Upstate Medical University, USA

Amar Gupta
University of Arizona, USA

ABSTRACT

Medical information has been traditionally maintained in books, journals, and specialty periodicals. A growing subset of patients and caregivers are now turning to diverse sources on the internet to retrieve healthcare related information. The next area of growth will be sites that serve specialty fields of medicine, characterized by high quality of data culled from scholarly publications and operated by eminent domain specialists. One such site being developed for the field of Gastrointestinal Motility provides authoritative and current information to a diverse user base that includes patients and student doctors. Gastrointestinal Motility Online leverages the strengths of online textbooks, which have a high degree of organization, in conjunction with the strengths of online journal collections, which are more comprehensive and focused. Gastrointestinal Motility Online also utilizes existing Web technologies such as Wiki-editing and Amazon-style commenting, to automatically assemble information from heterogeneous data sources.

INTRODUCTION

For the last several decades, Harrison's Principles of Internal Medicine, published by McGraw Hill, has served as a major source of information in the field of Gastrointestinal Motility. This book and its online presentation have been, and continue to be, used by many medical colleges to train the next generation of medical doctors; practitioners in this field also frequently refer to them.

Traditionally, papers and articles in specialty medical journals supplemented the material in textbooks like Harrison. The latter book would itself be updated periodically to reflect the state of the art in medicine and the various specialties, providing a consensus opinion of the standard of care.

The advent of computers and Internet has given rise to online sources of information such as UpToDate (http://www.uptodate.com/) and WebMD (http://www.webmd.com/). While gaining tremendous following and being updated frequently, these sources of online information relate to the medical field as a whole and not to particular specialties. Furthermore, the information on these sites is generally maintained by personnel of the respective organizations, not by specialists in specific disciplines of medical science. These organizations are usually set up as commercial entities, rather than not-for-profit ones.

The progressive transformation of information has seen many journals that were previously in paper format opting to use new electronic technologies; most of them now come out both in paper and electronic formats. Searchable electronic archives, such as PubMed (http://www.pubmedcentral.nih.gov/), now place a plethora of information into the hands of researchers and physicians. However, such searches are very time consuming and often produce irrelevant or poorly supported articles. Sites like Harrison's Online (http://www.accessmedicine.com/) serve as information directories that can be searched, hoping to place the most suitable information on a medical topic in a user's hand.

Students have gradually come to expect information in quick and readily available forms without having to bother about inter-library loans or even hardcopy versions at all.

The goal of the endeavor described in this article was to adapt emerging technologies to improve methods of teaching gastrointestinal material to students and to serve as a more effective source of relevant and accurate information for medical practitioners and specialists.

Evidence-Based Medicine

A study from the School of Information Management and Systems at UC Berkeley estimates that, in 2003, the World Wide Web contained about 170 terabytes of information on its surface alone, equivalent to seventeen times the size of the information in the Library of Congress (Lyman & Varian, 2003). With this increasingly information-rich society, the most precious ability for students and learners is no longer to find the information, but to discern the most relevant pieces of information and to integrate them into practice. The American Library Association describes "information literacy" as the ability of individuals to "recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information" (American Library Association, 1989).

The medical domain version of information literacy is evidence-based medicine.

Evidence-based medicine (EBM) is the integration of best research evidence with clinical expertise and patient values (Guyatt et al., 1992). The Centre For Evidence-Based Medicine in Toronto, Canada, states that the origins of evidence-based medicine date back to post-revolution Paris (CEBM, 2007), but that the current growth is most closely attributed to the work of a group lead by Gordon Guyatt at McMaster University in Canada in 1992. EBM publications, reflecting interest in this field, have grown from a lone publication in 1992 to thousands in 2007.

Studies have become increasingly critical of the value of textbook sources (Antman et al., 1992). Didactic continuing medical information may be ineffective at changing physician performance (Davis et al., 1997), and clinical journals may lack practical application (Haynes, 1993). In addition, physicians are faced with an increasing burden on their time, forced to diagnose patient findings within a matter of minutes (Sackett & Straus, 1998), and can only afford to set aside half an hour or less per week for general medical reading (Sackett, 1997). The staggering mass of information being discovered is also daunting: 500,000 articles are added to the commonly used Medline medical journal database every year, and "if a physician read 2 articles each day, every day for a year, (s)he would still find herself or himself 648 years behind" (Lindberg, 2003). As research increases the quantity of information available, medical practitioners are compelled to find efficient methods to educate themselves.

The Centre for Evidence Based Medicine has cited several examples of strategic, educational, and technical improvements in medicine that have enabled the current explosion in interest in this field. These include the emergence of new strategies for evaluating information; the creation of systematic reviews; the growing emphasis on continuing medical education and lifetime learning; and the advent of online journals, meta analysis of multiple studies, and ready access to such resources through electronic archives. Rapid dissemination of accurate and comprehensive compilations of research results enables medical practitioners to make informed decisions that are supported by the latest research results, and not by outdated trials.

In light of the increasing number of medical journals, especially journals that focus on specialties, the sheer quantity of information threatens to overwhelm medical practitioners. The concept of information mastery has been coined to describe the set of skills that physicians must nurture in approaching, analyzing and incorporating or rejecting new information. The issues mentioned above are not limited to the medical arena alone: Former Vice-President Al Gore described the state of information management as "*resembling the worst aspects of our agricultural policy, which left grain rotting in thousands of storage files while people were starving*" (Gore, 1992).

The Center for Information Mastery at the University of Virginia asserts that the usefulness of medical information is dependent upon its relevance, validity, and the work required to obtain it, as specified in Equation 1 (Slawson et al., 2007).

Further, the increasing quantity of research being performed by commercial enterprises, as well as other organizations with potential conflicts of interest, requires information be filtered for validity before incorporation into medical the canon. Finally, increased effort involved in accessing the relevant pieces of information reduces the accessibility of such information. In addition, healthcare organizations are siloed, gaining the advantage of sub-optimizing local departments, possibly at the cost of the whole (Senge, 1980); this complicates the problem further.

Information Retrieval and Decision-Making

As Stephen Hawking observes in *The Universe in a Nutshell*, the rate of growth of new knowledge is exponential. While 9,000 articles were published annually in 1900 and around 90,000 in 1950, there were 900,000 scientific articles published per annum in 2000 (Hawking, 2001). The explosive growth of information is challenging both the information repositories designed to hold it, and the ability of users to access relevant information.

At the time of this writing, Wikipedia serves as the de-facto standard for online general encyclopedias, and is among the top ten most-visited Web sites (Alexa Internet, 2007a). Its open-source, volunteer-without-accountability approach led to initial concerns about information validity, but these concerns have been largely addressed. Nature magazine studied Wikipedia and Encyclopedia Brittanica and found that the two of them were largely similar in accuracy (Giles, 2005). The growth of Wikipedia's information base further enhances the quality and breadth of coverage, and supports the possible future use of Wikipedia or Wiki-style architecture as an academically respectable source of reference.

Equation 1. Usefulness (Slawson et al., 2007)

$$\text{Usefulness of Medical Information} = \frac{(\text{Relevance})(\text{Validity})}{\text{Work}}$$

In contrast to the indexed and contributed semi-structured format of Wikipedia, Google relies on search-keyword phrases. The usefulness of Internet-crawling indexers, like Google, is based upon the ability to retrieve and capture information from many sites, and to retrieve relevant pages on query. Google's initial strength and rise to stardom was achieved through its superior PageRank algorithm, which still remains a carefully guarded trade secret; this algorithm provides an uncannily relevant list of matches to any user query, ranging from commonplace query phrases to obscure esoteric trivia and even misspellings.

In a small supermarket today, shoppers are bombarded with a selection of 285 varieties of cookies and 95 varieties of chips, leading the consumer to a state of decision overload (Schwartz, 2004). There is a growing need to restructure data to meet the informational and management requirements of an organization or group of people (Carlson, 2003).

Medical Information Repositories

For the medical arena, PubMed is the most widely used information database in the world, accounting for 1.3 million daily queries by 220,000 unique users (Lindberg, 2003). It is a free access search engine, provided by the U.S. National Library of Medicine as the main access point to the Medline database, a cataloged repository of medical literature classified using the descriptors known as Medical Subject Headings (MeSH). A broad range of search features are offered, including combined searches, exclusions, classification by type of article (original research versus review) and related articles. Another feature of the Medline database, known as MedlinePLUS, provides generalized information on health topics, and is aimed at the public or at practitioners outside their specialty domain. At the current time, PubMed serves as the gold standard in comprehensive medical information, despite its dated interface.

HighWire Press, a Stanford-originated endeavor, distributes thousands of journals, and provides its own search engine. In a recent study, the relevance of articles retrieved from HighWire was found to be greater than that of PubMed, but with the disadvantage of a slower retrieval Speed (Vanhecke et al., 2006).

In order to make a comparative evaluation between different approaches, it is appropriate to characterize the information recall ability using three parameters: precision, recall, and fall-out.

Precision (Equation 2) can be defined as the proportion of all retrieved documents that are relevant.

Recall (Equation 3) captures the concept of complete retrieval of all relevant documents.

Fall-out (Equation 4) is a measure of the number of documents that are retrieved but are unrelated to the issue being searched.

Medical research, while generally emphasizing maximal precision and minimal fall-out, occasionally requires increased recall, in the case of obscure diseases, or unusual side effects of medications. Medline serves as a canonical list for such purposes, but at the cost of significantly lower precision.

Medical practitioners using Google for their searches will often find themselves frustrated at the large quantity of articles on obscure and irrelevant topics. A researcher searching for a pharmacologic treatment of a syndrome will turn up with thousands of articles dealing with vari-

Equation 2. Precision (Wikipedia, 2007)

$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

Equation 3. Recall (Wikipedia, 2007)

$$\text{recall} = \frac{\left| \text{relevant documents}\} \cap \text{retrieved documents}\} \right|}{\left| \text{relevant documents}\} \right|}$$

Equation 4. Fall-out (Wikipedia, 2007)

$$\text{fall-out} = \frac{\left| \{\text{non-relevant documents}\} \cap \{\text{retrieved documents}\} \right|}{\left| \{\text{non-relevant documents}\} \right|}$$

ous sub-types, biochemical-signaling processes involved, and even support groups, before finding a therapeutic treatment. Due to the nature of the search engine and the storage methods, there are concerns about Google's or any search engine's ability to maintain a collection of such information. Carlson (2003) showed that due to the relatively small collection of documents indexed by an average search engine, a significant amount of relevant information would not be returned even in the presence of a perfectly formulated search phrase.

In order to accommodate domain-specific areas, Google has introduced the concept of "Refine Your Search" (http://www.google.com/coop). Without altering its main core search methodology, Google allows users to more quickly locate the type of information desired (i.e., treatment or symptoms). These refinement tools, provided by vendors and other private individuals or agencies that are deemed authoritative, subsequently label Web sites with appropriate descriptor tags. The potential conflict of interest created by these corporate associations is a matter of concern, due to omissions or maliciousness of the labeling.

The range of challenges and issues that characterize the medical domain include:

- The presence of an extensive array of synonyms for various drugs and diseases that require semantic knowledge to be encoded into the search engine in order to link concepts that are not lexically related;
- The naming of disease subtypes (often after a major contributor or discoverer), requires that hierarchies be constructed to allow users looking for the subclasses of disease to find information on the main umbrella disease, and vice versa;
- The growth in the understanding of generalized syndromes results in a corresponding need for reclassification based on new etiologies of disease, thereby suggesting a dynamic organizational structure for the online medical information systems.

In view of the growing difficulty in locating desired pieces of information, individuals performing research are in increasing danger of information overload. As such, the next generation of medical information access tools must aim to improve the ability to retrieve the right chunks of information quickly, with zero or minimal extraneous information; this concept is termed as increasing the signal to noise ratio in the field of electrical engineering.

VISION AND GOALS FOR GASTROINTESTINAL MOTILITY ONLINE

Gastrointestinal (GI) Motility Online is an example of a medical information system that seeks to provide access to high quality medical information online related to a particular medical specialty by centralizing the information and presenting relevant information that is customized to the user's information requirements.

The field of Gastrointestinal Motility is complex and interdisciplinary, involving a variety of experts. The possible user base includes laypersons, patients, medical students, biomedical scientists, physiologists, pathologists, pharmacologists, biomedical students, researchers, pharmaceutical staff, house staff, specialty fellows, internists, surgeons, and gastroenterologists. Each role requires a different approach to depth, scholastic relevance, and clinical direction in terms of information presentation. For example, students are interested in innovative research or review papers; researchers would like to know the most recent developments, and practitioners might be more interested in using the information for differential diagnosis purposes. GI Motility aims to serve as a collaboration of medical professionals, approaching diseases and patients from different angles.

In a library, a user interacts with the data in books very differently from the way that she or he interacts with data in an online presentation. The user expects the book to be focused and to address the topic in a linear fashion. Online, the same user navigates quickly, using hyperlinks, to explore secondary topics. In fact, the user *expects* a different presentation and a different style of information; as a result, the nature of the interactions will differ even with the same content. One of the aims of Gastrointestinal Motility Online is to address these different styles and to present the desired subset of information in the manner that the user might expect. For example, a user might be interested in viewing articles from the perspective of case-based, symptom-based, or test result-based diagnosis in order to apply the information to a particular problem at hand. In essence, the vision of Gastrointestinal Motility Online is to present information as framed by the interaction with the particular user at the particular point in time.

The first step in the vision of Gastrointestinal Motility Online was to collect the information in a manner consistent with the goal to acquire the reputation for the highest quality of knowledge. The information base is assembled entirely from material provided by internationally acknowledged experts. All chapters including synopses, articles, and reviews are written by reputed authorities. The pool of information is envisaged to be shared between different types of users and for different purposes. The design of the system emphasizes a one-stop information approach that enables the users to derive information at various depths. This applies to onsite information, as well as to information at offsite locations.

Details of Effort

A two-phase approach is being utilized for the creation of the information system: the first involves full leveraging of commercial technology as it exists today, and the second involves further research on aspects that can be incorporated in future versions of our system. In the absence of a better term, the term gastrointestinal knowledge repository is used for the final system, as well as for the initial concept-demonstration prototype system.

For the first phase, the acquisition of knowledge proceeded with the establishment of titles and themes for chapters, as determined through discussions involving the concerned authors and the editors for this project (Dr. Raj Goyal of Harvard Medical School and Dr. Reza Shaker of the Medical College of Wisconsin). The creation of the gastrointestinal motility knowledge repository began with calls to key gastrointestinal experts inviting them to submit a chapter, in electronic form, for inclusion in this knowledge repository. The inputs from the contributing authors were reviewed by these two editors and by others, on an anonymous peer review basis. Under the aegis of an unrestricted grant from Novartis Corporation, the two editors worked closely with the staff of

Nature Publishing Group on a number of tasks that ultimately led to the creation of the following Web site: http://gimotilityonline.com.

The creation of the site involved an automated conversion process to adapt MS-Word and rich text documents into a Web presentable format, with special emphasis placed on images, tables, and video. The majority of the investment for development lay in formatting and typesetting; the design itself was of less concern as it followed existing Web branding and style guidelines of Nature Publishing Group. Rights for images needed to be obtained; further, images, tables and video needed to be edited to fit a standard look and feel. After receiving author contributions, the project required approximately 18-24 months to complete, with 9-12 months required for the editorial process itself. The current site consists of 1,000 HTML pages, 1,000 images, 500 Powerpoint presentations and 40 videos. Articles are cross-linked by topics, and the current volume is equivalent of about 700 pages of text. This volume is increasing as this endeavor continues to progress.

Site Purpose

The ostensible purpose of the site is to disseminate information on the specialty of gastrointestinal medicine, primarily to gastrointestinal medical practitioners and to other interested parties. The ambition for the site is for it to become the central hub of information on this specialty, aggregating information from many sources, authors, and journals into one central location with the objective of becoming a de facto standard for online gastrointestinal information.

A secondary role is the development of a community of gastrointestinal specialists and other interested parties, who can form an online collaborative, expand upon the information repository, and facilitate peer communications and discussions. In addition, the site aims to explore and to expand the concept of online information repositories, especially the optimal integration of the current generation of electronic journals and online textbooks.

Site Audience

The primary site audience is the set of gastrointestinal specialists, associated medical staff, and staff under training. Members of this primary user base would initially visit the site because of a recommendation from a colleague or because another site (search or advertisement) directed the user to it. Occasional patients are expected, but are unlikely to become the primary users of this site.

Initially, the core attractions for users to this site are the quality and depth of information coupled with the ease of access and the low cost; these same factors will also help to retain the user base. Most physicians and healthcare specialists spend relatively little time online (six hours per week on average); in order to impress the user base, the signal to noise ratio of the site must be high, along with the ease of locating a desired piece of information (through the information architecture) (Friedman, 2000). The primary user of the site is unlikely to demand high interactivity; instead, the interest will be on locating and extracting the information as quickly as possible. As such, the site must be efficiently organized, streamlined, and equipped with powerful search and index functions.

After drawing a user initially, repeat visitors would use the site to browse new topics of interest, as well as to entrust the editorial staff to select articles that represent innovative research in the relevant field. An interacting community base would evolve over time and eventually cause a change in the workflow. Specialists would then visit the site to explore the comments from their colleagues on the topics expressed and to leave their own authoritative inputs on different articles, eventually contributing entire articles as much as a good electronic journal aspires to do today.

While relatively low in terms of being technically savvy, a typical user of Gastrointestinal Motility is likely to be very comfortable with using the Internet for retrieval of scholarly information through PubMed, online textbooks, and ready references such as UpToDate. A sleek, uncluttered design is likely to be the most attractive, even though it may not support applications involving high load times, such as Flash or embedded videos. Users are comfortable reading large articles online, but they also expect printable versions to be available, on an as-needed basis.

The user base, while highly intelligent, is relatively small in numeric terms, and is characterized by a small presence on the Web. The process of attracting a critical mass to build and to maintain a user community is a high priority task; this involves contacting a high percentage of all members of this specialist community.

Site Content

Gastrointestinal Motility Online is a site that provides information on both medical practice and the fundamentals of the gastrointestinal tract. The information must provide both breadth and depth on the topic, and should ultimately serve as an encyclopedia of the domain-specific knowledge.

Information Architecture: As in a library, information must be properly accessible in order to have value. The architecture of the information is partly determined by the methods that the users will use to query the evolving knowledge base. In a library, the name of the author and the title of the book are important. In a journal, the age of the article or the issue in which it was published may be essential. In Gastrointestinal Motility Online, the most likely user scenarios involve searching for information by anatomical section or syndrome. The timing is also important; recent material is favored over older material. Since relevance of the article is also important, the name of the journal and the name of the author are also used prominently in searches.

Furthermore, classification schemes or hierarchies to group-related topics are essential to narrow the branching process used by the search technique. The determination of related information is a complex topic. In order to address the latter issue, medical organizations are creating medical ontologies, such as unified medical language system (UMLS), to quantify and to impose structure on conceptual relationships. With this classification, the selection criteria can be hierarchically built up or finely specified to obtain desired results. A lengthier discussion of this topic appears later in this article.

Look and Feel: The look and feel of the site needs to convey the sense of scholarly authority, but with a sleek technological approach. Medical personnel have high standards concerning the accuracy of articles, and a professional presentation aids in supporting this perception. A site that is gaudy or shows too many bells and whistles, or involves a long load time, reflects poorly on the content, as do garish colors or lack of color.

Images and video must be consistent and should be available to download as needed for reference and for closer examination. The level of interactivity available should be low, as most gastrointestinal specialists have small online presence. Natural language queries, such as those used with AskJeeves (http://www.ask.com/), are unnecessary, as long as the search and browse features are precise and efficient. Pages may be presented as either textbook or journal articles, based on the preference of the user.

Extracting Content, Metadata, and Cross Referencing: One of the most powerful functions of the Internet is the ability for Web users to span several related articles quickly because of cross-referenced links. A user interested in the preparation of a particular chicken recipe can quickly reference how to sauté and with what form of pan, moving quickly to sources to buy the appropriate cast-iron skillet or wok to benefits and comparisons of different brands and retailers. A Wiki of only gastrointestinal specialists, with limited control from an editor, would be appropriate for the collection and dissemination of information: Gastrointestinal Motility Online is striving towards that goal.

A particularly useful feature in Gastrointestinal Motility Online, not available in online journals, is its cross-referencing tool. Articles that are closely related or broach a topic in greater depth can be quickly accessed by a cross reference within the Gastrointestinal Motility Online domain. These links are established by content extraction tools that create metadata and relate that data between different documents. Footnotes are available at the bottom and allow for a broader topical search; however, the inline linking of articles is particularly appealing for tracking particular items or syndromes of interest.

Auxiliary Technologies

Search features are incorporated in Gastrointestinal Motility Online, but are considered secondary to the organization of the information. A dynamic keyword search for anatomical sections of the gastrointestinal tract is less likely to reveal useful information on general function than a manual perusal of the literature through the prepared subject browse option. The efficient organization of the information, based on the anticipated needs and access patterns of users, is an essential feature for building a knowledge repository. By acknowledging that specialists would be more likely to query by anatomy or syndrome, one needs mechanisms for structured order, as compared to mechanisms that order by article size, author name, or recent usage. An additional feature, incorporated within the community-building module, is a user rating system that allows users to rank the importance of articles so that users browsing information can be directed to the most useful and informative articles.

One difficulty with searches is the likelihood that a particular phrase will appear in nearly all documents unless the search is very specific. In cases where the gastrointestinal tract is analyzed as an interactive system, the phrases for anatomical locations may occur multiple times as reference points, but the central theme of the article may not be easy to determine by word frequency. As a result, a number of semantic tools, described later in this article, are used to analyze articles and to classify them appropriately.

Experts of the Nature Publishing Group (NPG), who possess prior experience in online information presentation, based on the online version of Nature magazine, helped to develop the initial vision of the online knowledge repository. The base site is hosted by Nature Publishing Group. While the site was being developed, commercial tools were available to handle both the production of electronic journals and static textbook efforts like AccessMedicine/Harrison's Online (http://www.accessmedicine.com/) and WebMD. However, the gastrointestinal knowledge repository falls somewhere in between these two cases; accordingly, few over-the-shelf tools and algorithms were available for immediate use. As such, a significant fraction of the interface and architecture had to be innovated and refined, through experimentation.

Many of the existing tools for creating electronic journals are geared towards collation of articles, graphics, and layout work. These tools reduce the time needed by the authors and editors for the processes of uploading, formatting, and editing. In the development of Gastrointestinal Motility Online, the use of a software suite facilitated handling of images and consistency of look and feel. One area where tools are lacking is the ability to organize information into a coherent topical fashion, as in a textbook. Searching by keyword is especially difficult on a physician specialty site, where the dialect is limited and the concepts are reused multiple times. As such, editorial staff must impose additional control to prevent the site from becoming a write-only knowledge repository.

Collections of electronic journals, such as Ovid (http://www.ovid.com/site/index.jsp), have been primarily targeted for libraries and research centers. The primary purpose of Ovid is to serve as a warehouse of information, albeit uncategorized. Gastrointestinal Motility Online's current state differs from that of Ovid in terms of the presentation of the material: the former system is specifically formatted to provide an online view as well as a hardcopy output. The long-term goal of Gastrointestinal Motility Online is to collect a comprehensive knowledge on subjects (as Ovid does), as well as to add more intelligent search tools or information utilities. While Ovid does not organize information except into broad categories, Gastrointestinal Motility Online refines classifications, provides responses to search queries that are more accurate, and supports tools that use the knowledge in compelling ways, such as for differential diagnosis.

The advantages of sites built in a textbook style, such as AccessMedicine, is the hierarchical organization and ready access to information. The evolution of online textbooks has generated significant activity on the sites, as teaching tools. Gastrointestinal Motility Online uses over 40 illustrative videos, which are not typical of a journal, but fit the online textbook paradigm. Online textbooks are excellent repositories of information, except that updating the sites to incorporate new information is generally cumbersome because of the level of interactivity involved.

"Gastrointestinal Motility Online" is a hybrid, adopting the best qualities of online textbooks, such as Harrison's Online, and journal collections, such as PubMed. Organized, Well Edited, Frequently Updated and Comprehensive Information are self-explanatory. Search features specifically refer to the ability to search for a keyword or phrase. Differential diagnosis refers to the ability to integrate clinical patient presentations and create a list of possible problem diagnoses.

At a broader technological level, eBooks represent a technology that has been adapted to deliver information electronically. Medical eBooks can be argued to be a natural outgrowth of the eBook movement to electronic media: volume and space requirements are reduced, key phrase searching can be performed, and portability is enhanced. Nevertheless, few medical texts are adapted as eBooks.

Although eBooks have grown in popularity, they have not grown as rapidly as projected by consultants; this could be because of the following reasons:

1. Most readers see no need to replace print books.
2. Due to the limited screen size, limited battery life, and navigation interface issues with eBooks, many people still find paper books easier to handle.
3. Digital rights management causes compatibility and portability problems when attempting to move the eBook from desktop to PDA or laptop.
4. Current pricing of eBooks does not account for the reduced value relative to paper books. When readers finish reading a paper book, they can give it to a friend or sell it to a used bookstore; neither is possible with most eBooks. (Crawford, 2006)

Additionally, delivery of information may not be simply online, but online and to a mobile user using a PDA or other portable device. The constraints involved in transmitting and displaying information on a limited display panel create a new set of challenges. In most markets other than healthcare, the primary applications for PDAs are for scheduling and contact management (as a busy executive might use in lieu of a pen-and-paper daily planner), or as a portable browser or e-mail client (as in the case of technologists and engineers). In such cases, the application of the PDA works within the bounds of the limited display and the modern constraint of minimal bandwidth, often serving as a surrogate cell phone of sorts.

Table 1. The online endeavor

	Online Textbook	Journal Collections	Hybrid – GI Motility Online
Organized	X		X
Well Edited	X		X
Frequently Updated		X	X
Comprehensive Information		X	X
Search Features	X	X	X
Differential Diagnosis			X

However, in medicine, the PDA is often stretched beyond its limits. The current trend is the delivery of detailed information pages into a portable format, downloadable to PDA. Since medical practitioners can no longer maintain a complete mental catalog of all drugs and particularly obscure clinical symptoms or diseases, PDAs assist physicians in their duties without requiring a quick trip to a computer terminal or a large paper binder archive. Harrisons and UpToDate have both moved rapidly in this area, and the list of available drug databases is already large. PDAs fill the role of drug lookup very well, as well as serve as a primer for obscure diseases. The difficulty lies with more graphically intense data that may not display properly, or may need to be downloaded on the fly. In such cases, medical information systems are pushing the technological limits of PDAs.

PDA sales as a whole, however, are in decline, except as a niche application. Analysts at organizations such as IDC and Gartner have predicted downtrend trend in sales of PDA. Dell has withdrawn its PDA line from production (Mechaca, 2007). In the long term, the PDA may carefully constrain its niche to feature more of the portable planner features and less multimedia and display power, rendering it less useful to medical practitioners. Until the advent of revolutionary new display technologies, such as holographic displays or direct-to-eye projection technologies, the ability for PDAs to contribute to medical reference appears to be technologically limited.

As shown in Table 1, Gastrointestinal Motility Online is a hybrid that lies between the two models of electronic journals and online textbooks, exclusively focused on providing authoritative information in an organized fashion within a specific domain. Using this system, a gastrointestinal researcher can find the most recent journal articles because of the frequent updates, and a gastrointestinal clinician can easily locate a detailed diagram of the lower esophageal sphincter. In addition, topical information can be cross-linked between papers—a typical feature of textbooks and very pertinent for teaching and presentations.

CONCEPTS AND INNOVATIONS

The exploration of the technological space between electronic journals and online textbooks is a relatively new idea. All new ideas face challenges in terms of deployment and adoption. Consider the fax number. As the number of fax machines increases, the value of the fax increases—this illustrates the fact that networks attain greater value with larger number of users. The difficulty faced by Gastrointestinal Motility Online and other specialty interest sites is in terms of initial growth and development of specific communities of interest. These sites must be aesthetically attractive, informative, efficient, and up-to-date.

The case of Gastrointestinal Motility Online illustrates one form of evolution of online journals and textbooks into an active online scientific community. Site loyalty is achieved and maintained by the reputation of the authors and contributors. Gastrointestinal Motility Online needs only to achieve a critical user mass before gaining the benefits of Web sites like e-Bay (http://www.ebay.com/) or Amazon (http://www.amazon.com/) in terms of de facto authority and brand recognition. For sites which are less commercially oriented, the loyalty of the user base is perhaps even more heavily emphasized. Two notable sites, which have grown rapidly without such a strong commercial bias, are Wikipedia and Imdb.

Wikipedia has been both maligned and praised for its loyal community and efforts to create a free encyclopedia that is accurate and up to date without any commercial affiliations. Wikipedia began in 2000 as a complementary project for Nupedia, in which articles were written by experts and reviewed by a formal process (Wikipedia, 2007). In 2001, Larry Sanger proposed on the Nupedia mailing list to create a wiki as a “feeder” project for Nupedia (Sanger, 2001); this spawned rapid growth. By 2001, Wikipedia contained approximately 20,000 articles and 18 language editions. As of 2007, English Wikipedia contains over 1.7 million articles, making it the largest en¬cyclopedia ever assembled. The site relies on the goodwill of its members to write, update and contest articles, and has thus far proven that the Internet community as a whole is willing to contribute towards the database, albeit haphazardly (recent events are more likely to be covered in detail, while significant historical figures languish). Given the size and relatively stable growth of the project, the prospect of a peer-reviewed and written information repository might not be so cynically doomed.

The International Movie Database (IMDB), a site in the top 20 (Alexa Internet, 2007b) U.S. Web sites visited, is the largest Internet compilation of movie information (approximately 900,000 titles and 2.3 million names) (IMDB, 2007). IMDB draws a significant portion of its information from the participation of its user base. Beyond subjective reviews, users are also asked to supply cast and crew lists, production details, and actor biographies. IMDB grew from two lists that started as independent projects in early 1989 by participants in the Usenet newsgroup rec.arts.movies. Each list was maintained by a single person, recording items e-mailed by newsgroup readers, and posting updated versions of his list from time to time. The lists were eventually combined, and by late 1990, the lists included almost 10,000 movies and television series. As the contributions continued to grow rapidly, the IMDB formed as an independent company in 1995 and was later purchased by Amazon Inc. (Wikipedia, 2007b).

The approach of the IMDB system closely resembles the envisioned system for GI Motility Online, with the user population submitting proposed changes, followed by an editorial process. Database content is generally provided and updated by a vast collection of volunteer contributors. There are only 17 members of the IMDB who are dedicated to monitoring received data, although 70% of IMDB’s staff serve as editors (IMDB, 2007), reviewing changes, verifying the information before posting the changes, and policing the forums.

Peer review is considered to be one of the most even-handed and least biased methods of scrutinizing articles for publication. The development of a community, as well as the ability for members of that community to voice their opinions about professional papers is pivotal to the dissemination of accurate information. The model of Amazon or eBay is to allow users to comment, and thereby to signify reliability and approval. Gastrointestinal Motility Online allows users to provide feedback, both critical and supportive, in order to enhance the relevance of articles.

According to Harris Interactive poll, only one-fourth of physicians use the Internet to communicate with their patients (Computing in the Physician’s Practice, 2000). In the same poll, although 89% of physicians use the Internet in their practice in search of information, they spend only six hours per week to browse medical developments. Accordingly, Gastrointestinal Motility Online has been configured to serve as an encyclopedic source, as well as a high-value news feed. AccessMedicine uses a Podcast update model, with 10-minute broadcasts generated daily for use in family practice. Gastrointestinal Motility Online caters to a smaller specialty group with correspondingly less frequent pace of developments, and is therefore less time sensitive.

Automation

Automated content generation or extraction from other publications is not feasible, due to the stringent need to maintain relevance and quality. Thus, deployment of a pure peer-reviewed wiki-style community is precluded by the need to maintain quality. If membership of the wiki is restricted to those users whose credentials are accepted, or if changes must be approved at an editorial level, then a wiki would facilitate the rapid exchange of information as well. In the second part of the endeavor, the goal is to enable machine-assisted updating of the material in the gastrointestinal knowledge repository. Currently, all updates must be initiated manually. The pressure to update, but to update accurately applies to many situations: addition of new material, editing of existing material, and deletion of parts of existing material as new study results become available. Initial thoughts and test results are documented in Sharma (2005).

As the site evolves, the need to keep the site relevant to a particular group of specialist physicians may conflict with the preferences of another significant category of users: the researchers. Gastrointestinal Motility Online intends to use user-based customized layout, as presented by Sarnikar et al. (2005). The filtering of details from the main knowledge repository is not intended to block or hide information, but to provide a more relevant source. Classification and weighting are accomplished by a rules-based system that can assess whether an article is clinical- or basic science-related. Login will also provide both the ability to contribute and comment on articles, and to obtain a specialized view depending on the type of user. The application of user-based site layouts is not innovative, but is important to the domain of medical informatics because physicians attach such a high priority on relevance.

Given the large number of articles published weekly and the difficulty in ascertaining relevance and quality, a number of automated tools will be used to optimize the updating process. Sarnikar et al. (2005) present one technique that will assist in filtering journal results and maintaining and updating the site. Their method selects articles ranked by relevance using a combination of both rule-based and content-based methods, using the following principles:

1. Profiles are modeled in the form of rules.
2. The purpose of the rule-based profile is to identify a sub-set of documents of interest to a user.
3. Each role has a set of predefined rules associated with it.
4. Rules specify knowledge sources to access (e.g., nursing journals for Nurses).
5. Rules can specify knowledge depth and knowledge breadth.
6. Rules can specify semantic types of primary importance to roles.

Profiles are used in the gastrointestinal motility context to separate information into categories: for example, new clinical findings versus basic science. Articles may be assigned a category and a relevance weight, given categorization rules based on Unified Medical Language System (UMLS) synonym lists and the categories *sign or symptom, diagnostic procedure therapeutic or preventive procedure and disease or syndrome semantic types.* In addition to a text search in the abstract, Sarnikar and Gupta (2007) also assign weights to the type of journal. These tools can select and filter relevant articles for presentation as an RSS XML news feed to editors or automatically assemble relevant articles for use by the editors or Web site administrators. While these tools will aid the editor, there is no replacement for the role of humans in decisively selecting and classifying information.

Ontologies and semantic networks are prerequisites to the development and classification of information repositories. Ontologies serve many purposes (Kumar, 2005) including:

- Reuse and sharing domain knowledge
- Establishing classification schemes and structure
- Making assumptions explicit

They further enable analysis of information and complement the stricter terminology that is used in straightforward text searches. Examples of ontologies in use today include the National Library of Medicine's Medical Subject Heading (MeSH), disease specific terminologies such as the National Cancer Institute's PDQ vocabulary, drug terminologies such as the national drug data file, and medical sociality vocabularies such as the classification of nursing diagnoses and the current dental terminology. In Gastrointestinal Motility Online, the ontological hierarchy will be used to distinguish between sections of the gastrointestinal tract, from the stomach and esophagus to the large intestine and colon (Kumar, 2005).

A key enabler for development of automatic information processing is the set of ontologies presented in the UMLS semantic network that rely upon the concepts built in the UMLS concept hierarchy. Hierarchical and clustered ontologies allow software to construct knowledge trees, conglomerating relevant knowledge. An overview of the UMLS is available at the unified medical language system: What is it and how to use it? (http://www.nlm.nih.gov/research/umls/presentations/2004-medinfo_tut.pdf).

UMLS is an aggregate of over 134 source vocabularies, including the classifications from such lists as ICD-10 and DSM: IIIR-IV. It represents a hierarchy of medical phrases that can be used to classify most medical articles and textbook entries. For example, using UMLS, the following phrases are grouped similarly: *Deglutition Disorders, Difficulty in swallowing, Difficulty in swallowing (context-dependent category), Dysphagia NOS, Dysphagia NOS (context-dependent category), Can't get food down, Cannot get food down, Difficulty swallowing, Difficulty swallowing (finding), Dysphagia, Dysphagia (Disorder), Swallowing difficult, Swallowing Disorders* (Aronson, 2001).

The system described in Sharma (2005) uses techniques of Natural Language Processing (NLP) to construct a semantic understanding that surpasses text searching. Using the automated integration of text documents in the medical domain (ATIMED) system, the content and order of phrases are related lexically using a concept called Word-Net. Word-Net operates on the verbs, subjects and objects of the sentences, comparing sets and subsets of subject-verb-objects collections in order to determine topic relatedness.

Word-Net further uses a lexical dictionary to determine similarity in all verb pairs and then subject-verb-action pairs. Sharma (2005) uses the following two sentences as examples: **Dysphagia is a disease and defined as a sensation of sticking or obstruction of the passage of food. Dysphagia is related to obstruction of passage of food.** Since both sentences contain similar objects and subjects, and use the verb "is", the sentences are deemed similar. However, within the current mechanism, the phrase, **"Dysphagia relates to obstruction of passage of food"** would result in a poorly scored correlation or low match because the action verb is not similar (Sharma, 2005).

Finally, the same technique allows the creation of new documents by collating sentences and paragraphs from various documents. An initial method of grouping sentences uses the quantity of concepts expressed. This method is further refined by evaluating the sentences based on the following criteria: *similar-subjects, similar-objects, similar-subjects-objects,* and *similar-objects-subjects.* Based on the structure of English grammar, these techniques have been reliably shown to collate relevant data into a readable format.

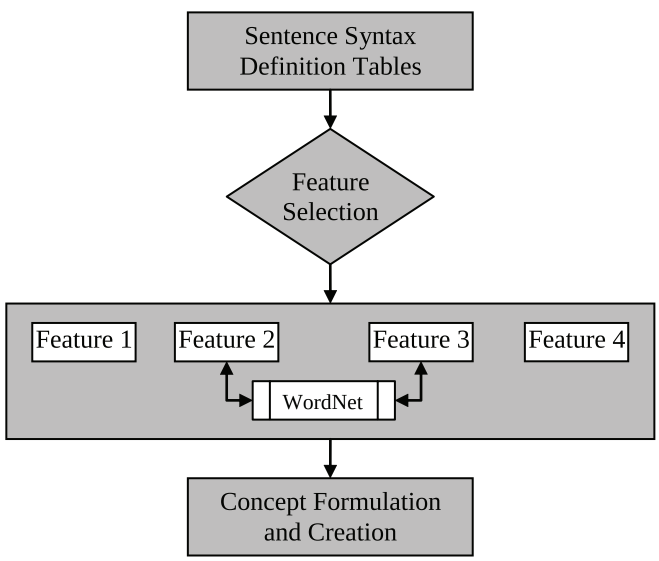
A diagram of the method is shown in Figure 1.

A sample output, based on the use of this technique, is provided below:

REGURGITATION is defined as the spontaneous appearance of gastric or esophageal contents in the back of the throat or in the mouth. In distal esophageal obstruction and stasis, as in achalasia or the presence of a large diverticulum, the regurgitated material consists of tasteless mucoid fluid or undigested food (Sharma, 2005).

The context interchange of heterogeneous sources of information being collated for different classes of users leverages tools from an additional branch of information technology. One level of development is the creation of maps between inputs and outputs, in much the same way that a dictionary might map between languages. The conversion of *n* sources to *m* outputs can grow to be an exponentially difficult problem; this can be addressed with the use of intelligent heuristics and protocols to selectively prune the data. Difficulties become particularly apparent with changes in any client schema that cause a cascade of changes in the mappings (Sarnikar, 2005). Using an independently developed predefined mediating schema would restrict the amount of information that could be exchanged. Apart from relational schema, a client schema could also be specified as a hierarchical schema or as an XML-based message (Sarnikar & Gupta, 2007). In the context of Gastrointestinal Motility Online, the specific source contexts are innovative journal papers, reviews, and textbook articles. The output contexts are specialist clinicians, researchers, students, and other health professionals. Developing a schema to accurately represent journal abstracts and determine the relevance of those abstracts is another method of exchanging contexts. Innovation in this domain will allow Gastrointestinal Motility Online to maintain updated, consistent-quality

Figure 1. Integration of semantic understanding to generate new information (Adapted from Sharma, 2005)



references without requiring an editor to read every journal article published immediately.

LESSONS LEARNED

The GI Motility Online site benefited from a mature development environment for Web-based information retrieval.

Interface: In the case of GI Motility Online, the user base is a readily identifiable group, trained in a similar fashion, with specific needs and expectations of organization and formatting. The following two aspects influenced the design process:

- Pagination is a critical issue for online didactic materials. As people do not like the interface to online books (Crawford, 2006), the development of the site must reflect the reading style and needs of its users. In the case of GI Motility Online, there are currently no page delimiters; this is encouraging users to print the material to be used as a reference. A solution of delivering a formatted print-ready PDF could be applied, but the cost begins to climb with the number of different formats and delivery options supported.
- Tables must be handled intelligently. Tables are used in many medical publications to rapidly and clearly present information. The trend in the mid-1990s for netiquette was to inline the tables with the text, but this can create awkward gaps or poor formatting choices. GI Motility Online chooses the path of Harrison's Online, in linking to the tables outside the main document to preserve readability.

Based on their respective training, many physicians expect a particular language and layout of medical information. The highest-value information for each physician may vary based on specialty, and user-customization of graphical widgets may enhance information value. Since GI Motility Online aims to support a broad base of users with many different roles and information needs, reusable widgets, of the type available in customizable Web-portals, may potentially solve these varying needs.

Workflow: As stated previously, over nine months was required for simply editing the site information, moving documents back and forth between editors, experts and developers. The method involved several inefficient technologies, such as mailing CDs and sending large files through mail servers. The advantage of such hands on interaction and communication is the result: the authors of the site are particularly pleased by the polished appearance of the delivered product.

Electronic collaboration between different sites was difficult due to the nature of document formats, which does not produce a consistent print layout on different computers, and layout formats, which do not allow easy markup or revision to the document.

The use of a standardized input format and efficient conversion of the RTF-formatted documents into Web viewable formats was crucial to the development of the Web site. Though conversion of documents is a minimally complex task, the production of a site that allows conversation on the fly to support concurrent editing and proofing between multiple users is a challenge. Online publishing companies, such as Atypon, offer suites of software to facilitate publishing workflow, allowing multiple authors to upload articles, multiple editors to revise articles, and art editors to manage graphics, all in an organized fashion.

Metainformation: The World Wide Web provides a tremendous asset in connecting related articles seamlessly. In determining related pages, an active agent—a human, machine or some combination thereof—who must isolate the crucial elements of a page and capture that as meta-information, preferably categorized. Organized meta-information allows automated tools to develop connections and links to potentially relevant information.

Extracting crucial elements of a page is best handled by encouraging authors to define keywords, as with scientific articles. Although many automated agents have historically been unsuccessful in information extraction, the medical domain provides some assistance with standardized ontologies, which allow agents to categorize information in a framework, reducing some identification errors.

Maintenance: Many articles in mainstream sites, such as ESPN.com, offer the opportunity for users to comment and leave feedback. This increases user participation and value to the site, but raises additional issues that need to be addressed, such as:

- Profanity or other inappropriate comments must be censored
- Sites are more open to attack and security leaks due to their increased functionality
- Bandwidth usage increases, and may debilitate the site

In addition, a version control framework must be established for sites that allow updates, in order to rollback unwanted changes. Furthermore, hardware resources must be allocated to store site changes.

CURRENT USAGE AND FUTURE DIRECTIONS

Anecdotal evidence provided by the lead creator of GI Motility Online, Dr. Raj Goyal, highlights that the representatives of Nature Publishing Group describe the site as being popular, and that site visits to GI Motility Online are increasing. Due to the delicate nature of conversation with his peers in gastroenterology, Dr. Goyal cannot be certain, but he states that initial impressions are unanimously positive and enthusiastic. Dr. Goyal notes that the majority of the traffic to GI Motility Online currently arrives from Google, and indeed, Google's first returned site in response to "GI Motility" is GI Motility Online (Google, 2007). GI Motility Online does not currently permit any advertising; if this policy is revised, the usage patterns may change. Further, the numbers for site visits are expected to increase if the site were to be more supported by Nature Publishing Group.

The stakeholders are satisfied with the quality of the product, and are ambitiously pursuing an extension of the product to broaden coverage to other parts of the GI tract.

Dr. Goyal has received inquiries to publish the material in a hardcopy format. Custom-built books or full-page colored slides offer additional utility to medical educators, specialists, nurses, and students, and may also offer a revenue stream. These custom-built printable books are a logical extension of the current "print what you want" photo fulfillment services that are popular on photo sharing sites or album hosting services. Given that Harrison's Principles of Internal Medicine contains over 2,000 colored pages in book form, there may be a future for customized books in medicine.

Retrieval: New algorithms for search and ranking are not merely lexically-based, but also combine closely related concepts and relationships between ideas. This facilitates the creation of a richer search language that can account for relationships, such as causation, consequence, association, treatment, or opposite. For example, hypothermia is directly opposite hyperthermia, and may be caused by thalamic alterations. Such basic chains of relationships could be easily captured if the language and storage of meta-information about articles could contain the necessary underlying details.

Presentation: Using Tufte's principles (Tufte, 2001), the real estate of a screen needs to be more efficiently used. Currently, most search-engine results do not often give the sense of the relevance of the article, of the correct sections, or the tone of the article. The information density of the return pages is low, causing users to scroll through potentially hundreds of articles to find relevant information. Graphical interfaces are likely to be a solution, as the ability of the Internet to handle higher bandwidth applications grows. In medical research in particular, an interface, that allows users to quickly see search results of the primary concept and even related concepts, may dramatically affect usability.

Online Sharing: Bulletin board systems, which flourished during pre-Internet days, have re-emerged on the Internet as forums, and are a popular source of information. Corporations with significant Internet presence, and especially gaming companies like Nintendo (http://forums.nintendo.com/nintendo/), have begun adding and using these forums as a method of improving public relations and offering support.

Intranet and Internet file sharing systems have also flourished, as bandwidth rates increase. The rise of Youtube is one phenomenon, but the comfort of using the Internet to disseminate multimedia (such as GI endoscope video) appears to be more solidified. These high bandwidth applications have grown in acceptance, and despite the threat of viruses in downloaded files, file-sharing traffic is increasing daily.

Megaupload and Rapidshare, two prominent file sharing services, are now ranked #18 and #27 in the reliable Alexa ranking (Alexa Internet, 2007b) of most visited Internet sites in the U.S. However, these services are predominantly used by non-commercial entities. Commercial enterprises may be hesitant due to slow adoption, security issues, bandwidth maximums, or unprofessional presentation. Such file sharing sites will certainly cater in the future to commercial entities, perhaps by providing specially developed sites, or providing branded services. The future of the Internet, and especially the ability to deliver high bandwidth, will increasingly rely upon specialized sites with high-capacity and high-bandwidth connections close to Internet backbones.

Collaboration: Collaboration in the domain of GI motility will drive improvements in language tools, with many GI specialists throughout the world who may need to collaborate via the Internet or phone services. Currently, Altavista and Google offer Web-based text translation, but more intelligence or domain-specific knowledge may be required. One common example in GI translation is translating the world "oral" to mean "verbal", when in fact, the correct reference is to the mouth cavity. Overall, however, text translation between languages is generally adequate for initial communication.

Oral translation is a high value direction to pursue, but at the same time, is a very difficult task. Psycholinguistics research is still unraveling the complexities of language parsing, and the ability of current artificial intelligence to understand language is severely limited. Nevertheless, translators in this area will prove essential and highly desirable for the next generation of online collaborators.

Similarly, one of the holy grails of artificial intelligence development is the creation of an artificial system capable of interpreting human language. Within specialized domains with limited vocabularies, artificial readers become more feasible, but the medical domain is particularly difficult, due to its large specialized vocabulary. Development in this area would provide rewards for medical researchers, allowing the creation of agents, which would allow researchers to process more information by selecting and even summarizing articles.

Organizationally, scientific research in domains such as medicine would benefit tremendously from the creation of a centralized authority to monitor, synthesize and rate research. The current system of research funding in America places research at the whim of special-interest private funding and sometimes misdirected public funding in overly popular or extremely esoteric areas. Regulating and directing research might also help avoid repeating inconclusive research, which does not get published (and thus may be repeated).

CONCLUSION

Gastrointestinal Motility Online is an evolving knowledge base related to Gastrointestinal Motility disorders. The current phase of the endeavor focuses on the collection and organization of knowledge from many different sources. Knowledge-mining tools are being developed to utilize this information as it becomes available to add fast, relevant access and other utilities to the information repository.

The continuous change in standards of care and knowledge due to rapid discoveries in the basic and clinical sciences prompts for a system that is more flexible than a textbook, while demanding thoroughness and accuracy. Knowledge mining tools and other advanced technologies to aid in the conversion and integration of articles and research into the mainstream science are being integrated into Gastrointestinal Motility Online, and look to impact the breadth and speed of knowledge-base upgrades. Gastrointestinal Motility Online serves to balance the needs of its user base while embracing the academic rigor in a novel application of technology to the science of medicine.

ACKNOWLEDGMENT

The authors would like to thank Dr. Raj Goyal for his invaluable input and contribution to GI Motility Online and this article. The authors also thank Richard Martin for his helpful comments and proofreading of the text.

REFERENCES

Alexa Internet. (2007a). *Three-month traffic statistics for wikipedia.org*. Retrieved from http://www.alexa.com/data/details/main?q=&url=wikipedia.org.

Alexa Internet. (2007b). *Top Sites United States*. Retrieved from http://www.alexa.com/site/ds/top_sites?cc=US&ts_mode=country&lang=none.

American Library Association. (1989). *Presidential Committee on Information Literacy - Final Report*. Retrieved from http://www.ala.org/ala/acrl/acrlpubs/whitepapers/presidential.htm.

Antman, E., Lau, J., Kupelnick, B., Mosteller, F., & Chalmers, T. (1992). A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. *Journal of the American Medical Association, 268*, 240-248.

Aronson A. (2001). Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program. *Proceedings of the 2001 AMIA Symposium*, pp. 17-21.

Carlson, C. (2003). Information overload, retrieval strategies and Internet user empowerment. In: L., Haddon (Ed.), *The Good, the Bad and the Irrelevant (COST 269) 1*(1) (pp. 169-173). Helsinki, Finland.

CEBM. (2007). *Why the sudden interest in EBM?*. Retrieved from http://www.cebm.utoronto.ca/intro/interest.htm.

Computing in the Physician's Practice. (2000). Retrieved from http://www.harrisinteractive.com/harris_poll/index.asp?PID=58.

Crawford, W. (2006). *Why aren't ebooks more successful?*. Retrieved from http://www.econtentmag.com/Articles/ArticleReader.aspx?ArticleID=18144.

Davis, D., Thomson, M., Oxman, A., & Haynes, R. (1997). Changing physician performance: A systematic review of the effect of continuing medical education strategies. *Journal of the American Medical Association, 274*, 700-705.

Fridman, S. (2000). *Doctors lag when it comes to computer use - Industry trend or event.* Retrieved from http://www.findarticles.com/p/articles/mi_m0HDN/is_2000_March_28/ai_60907019.

Giles, J. (2005). *Internet encyclopedias go head to head.* Retrieved from http://www.nature.com/news/2005/051212/full/438900a.html.

Google. (2007). Search returned http://www.google.com/search?hl=en&q=GI+Motility.

Gore, A. (1992). Infrastructure for the global village. *Scientific American*, *265*,150-153.

Guyatt, G. Cairns, J., Churchill, D., Cook, D. Haynes, B., Hirsch, J., & et al. (1992). Evidence-Based Medicine Working Group: Evidence-based medicine. A new approach to teaching the practice of medicine. *Journal of the American Medical Association, 268,* 2420-2425.

Hawking, S. (2001). *The universe in a nutshell*. New York, NY: Bantam Books.

Haynes, R. (1993). Where's the meat in clinical journals [editorial]? *ACP Journal Club*, *119*, A22-A23.

IMDB. (2007). *How/where do you get your information? How accurate/reliable is it?* Retrieved from http://imdb.com/help/show_leaf?infosource.

Kasper, D., Fauci, A., Longo, D., Braunwald, E., Hauser, S., & Jameson, J. (2005). *Harrison's principles of internal medicine*, 16th ed. New York, NY: McGraw-Hill.

Kumar, A. (2005). *Ontology-driven access to biomedical information (ODABI).* Undergraduate Thesis, University of Arizona.

Lindberg, D. (2003). *NIH: Moving research from the bench to the bedside.* Presentation to the Subcommittee on Health.

Lyman, P. & Varian, H. (2003). *How much information?*. Retrieved from http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/index.htm.

Mechaca, L. (2007). *Goodbye, Axim*. Message posted to http://direct2dell.com/one2one/archive/2007/04/11/11397.aspx.

Sackett, D. & Straus, S. (1998). Finding and applying evidence during clinical rounds: The evidence cart. *Journal of the American Medical Association, 28*0, 1336-1338.

Sackett, D. (1997). Using evidence-based medicine to help physicians keep up-to-date. *The Journal for the Serials Community, 9,* 178-181.

Sanger, L. (2001). *Let's make a wiki*. Message posted to http://web.archive.org/web/20030414014355/http://www.nupedia.com/pipermail/nupedia-l/2001-January/000676.html.

Sarnikar, S., Zhao, J., & Gupta, A. (2005). Medical information filtering using content-based and rule-based profiles. *Proceedings of the AIS Americas Conference on Information Systems (AMCIS 2005),* Omaha, NE.

Sarnikar, S. & Gupta, A. (2007). *A context-specific mediating schema approach for information exchange between heterogeneous hospital systems.* Forthcoming in International Journal of Healthcare Technology and Management.

Schwartz, B. (2004). *The paradox of choice.* New York, NY: HarperCollins Publishers.

Senge, P. (1994). The fifth discipline: The art and practice of the learning organization. New York, NY: Doubleday/Currency.

Sharma, R. (2005). *Automatic integration of text documents in the medical domain.* Undergraduate Thesis, University of Arizona.

Slawson, D., Hauck, F., Strayer, S., & Rollins, L. (2007). *What is information master?*. Retrieved from http://www.healthsystem.virginia.edu/internet/familymed/docs/info_mastery.cfm#Information.

Tufte, E. (2001). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.

Vanhecke, T., Barnes, M., Zimmerman, J., & Shoichet, S. (2006). PubMed vs. HighWire Press: A head-to-head comparison of two medical literature search engines. [Epub ahead of print] *Computers in Biology and Medicine*.

Wikipedia. (2007a). *Information retrieval.* Retrieved from http://en.wikipedia.org/wiki/Information_retrieval.

Wikipedia. (2007b). *Wikipedia.* Retrieved from http://en.wikipedia.org/wiki/Wikipedia.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 3, Issue 1 J. Tan , pp. 24-43, copyright 2008 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 5.11
Web Services in National Healthcare:
The Impact of Public and Private Collaboration

Matthew W. Guah
University of Warwick, UK

Wendy L. Currie
University of Warwick, UK

ABSTRACT

The implementation of a national programme for information technology into the complex environment of the UK's National Health Service (NHS) system is only the first step in a system modernisation journey for this multifaceted organisation. This article reports the findings of a two-year research study on the decline of the ASP (application service provision) industry, from which the current move to Web services was born. It combines the case of the NHS with existing literature on disparate research perspectives to explore the effects of the 'Not Invented Here' syndrome on an IS implementation journey. The article also suggests ways that project leaders can redirect such strong feelings about a new system to increase the chances for a successful outcome. Information systems (IS) is "an instantiation of information technology (IT), where the same information technology is instantiated in different ways" (Lee, 1999). A rich organizational and political process is required for a given set of IT to be instantiated, relying greatly upon the continual managing, maintaining, and changing of technology to sustain the instantiation. Within a rather diverse NHS environment, IS may include relational aspects like the effectiveness of system design, the timely delivery of such systems, an appropriately obtained usability training by all users, and future impact of IT in the organisational and entire society.

BACKGROUND

The National Health Service for the United Kingdom (NHS) has been responsible for the provision of health care and services in the UK for the past 56 years, on a "free for all, at the point of delivery" basis. The traditional perception of the NHS is one of a healthcare system organised as a professional guild, with unlimited finance from the government. This type of NHS is experiencing an irrevocable change, as taxpayers are no longer complaisant, and paternalistic employers are reacting against inflating costs and escalating complaints from the patients. The employer is reacting to the continuous massive flow of subsidies for inefficient physician practices, fragmented delivery systems, and cost-unconscious consumer demand. The patients are increasingly assertive as to their preferences, and few have expressed their willingness to make additional contributions for particular health benefits and medical interventions.

Web services (WS) are technologies with roots in the application service provision (ASP) business model which are used mostly to automate linkages among applications (Hagel, 2002). They are generally anticipated to make critical systems connections not only possible, but also easy and cheap (Kreger, 2003; Sleeper & Robins, 2001). One of the perceived benefits of WS is that organisations would be able to concentrate on their core competencies (Perseid, 2003). Service providers argued that the remote delivery of software applications would release managers from the perennial problems of running in-house IT departments, allowing more time to develop IT and e-business strategy rather than the day-to-day operations (Currie, Desai, & Kahn, 2004). This justification has been used in traditional forms of outsourcing over many years (Willcocks & Lacity, 1998).

The NHS is experiencing massive changes in the structure of information systems provision markets and organisations. The local service provision (LSP) and national service provision/provider (NSP) models, in use by the National Programme for Information Technology (NPfIT), are in a state of ferment. The payment methods borrow from both capitation and 'fee-for-service', and methods of utilisation management that compromise between arm's-length review and full delegation. LSP and NSP consist of large and more complex entities. These are the result of merger, acquisition, and product diversification. The service providers involved have had to take on a visible feature of ceaseless acquisition and divestiture, integration and outsourcing, and combination and recombination. Providers of medical systems, hospital administration systems, and health plans are coming together and then coming apart. They are substituting contract for joint ownership, creating diversified conglomerates and refocused facilities, and experimenting with ever-new structures of ownership, finance, governance, and management (Robinson, 2000). These would give the NHSIA the benefits not only of a middle ground between the extremes of vertical integration and spot contracting, but also a balance of coordinated and autonomous adaptation in the face of its ever-new challenges.

The general assumption is that expenditures in the nation's health will outpace the overall growth in the economy (Collins, 2003; Pencheon, 1998). This is reflected in the percentage of gross domestic products (GDPs) of the U.S. (13%), Germany (10.7%), France (9.6%), and the UK (7.6%) being devoted to the total cost of healthcare resources (Brown, 2002). Unlike the UK, however, some of these countries are faced with limitations in social willingness to pay. It has been documented that millions of U.S. residents currently lack the most basic insurance coverage (Institute of Medicine, 2002).

Response to Emerging Technologies in the NHS

Over the years, non-technologists in the NHS have managed to muddle through one powerful new system after another. 'Generational strategy' is one continuously being used to deal with some of the pressures induced by IS. Adopting such innovations as PCs and the Internet requires the personal and organizational costs of "unfreezing" deeply ingrained old habits. Many workforces ignore, deny, or deal awkwardly with such technologies.

A common response to new systems is the 'Not Invented Here' (NIH) syndrome (Collins, 2003; Haines, 2002; Guah & Currie, 2002). This often leads to certain organisations rejecting a perfectly useful system based on an implicit assumption that the system does not fully recognise or accommodate their own needs and idiosyncrasies (Brown & Venkatesh, 2003; Davis, 1989). Davis (1989) sees this as a likely result of a decline in communication with external sources. NIH syndrome could also result from competences that can be proven to be outdated and inefficient in comparison to an existing technology. One Trust, which plays a central role in the direction of regional IS strategy, had to reject a Department of Health-promoted system because the system was not as familiar as another bespoke system (Haines, 2002).

The common characteristics of new systems in the NHS are uniformity in products and prices in the face of great variability in consumer preferences and the actual costs of providing service (Collins, 2003). This one-size-fits-all approach usually leads to services that are of excessive costs for some users and insufficient quality for others, impeding the use of price flexibility to enhance capacity utilization (Robinson, 2000). Also of concern is a combination of over-capacity and low load factors in some regional trusts with under-capacity and shortages elsewhere. Concerns are growing in the NHS that this may generate cross-subsidies from trusts for which the cost of service will be low to trusts for which the cost of service will be high (McGauran, 2002). Additionally, deregulation of healthcare costs has spurred an outpouring of new services. Consequently, several of these services are the following (Collins, 2003; Pencheon, 1998):

- A different cost structure
- An impact on IS budgets
- A better match between supply and demand

Incomplete information has been a fascinating attribute of the NHS unusual systems organisational and normative characteristics. The asymmetry of NHS information between patients and medical practitioners has changed in an exogenous fashion over its 56 years. The amount of healthcare information available to patients is usually the result rather than the cause of changes in the economic and political environment (Robinson, 2000).

PROJECT DESCRIPTION: NATIONAL PROGRAMME FOR INFORMATION TECHNOLOGY (NPFIT)

The NPfIT is an initiative by the National Health Service Information Authority (NHSIA), born as a result of several plans to devise a workable IS strategy for the NHS (NHSIA, 2003; Wanless, 2002). The NPfIT was designed to connect the capabilities of modern IT to the delivery of the 'NHS Plan' devised in 1998. The core of this strategy is to take greater control of the specification, procurement, resource management, performance management, and delivery of the information and IT agenda (NHSIA, 2003).

The NPfIT is an essential element in delivering the NHS Plan. It has created £6 billion budget for information infrastructure, which could improve patient care by increasing the efficiency and effectiveness of clinicians and other NHS staff. The intention of the plan is to address the following (www.npfit.nhs.uk):

- Create an NHS Care Records Service to improve the sharing of consenting patients' records across the NHS

- Make it easier and faster for GPs and other primary care staff to book hospital appointments for patients
- Provide a system for electronic transmission of prescriptions
- Ensure that the IT infrastructure can meet NHS needs now and in the future

The decision to implement a national programme for IT into the NHS system complexity is only the first step in an IS modernisation journey for a multifaceted organisation. There are many examples of new technologies disrupting organizational routines and relationships in the NHS (Majeed, 2003; Atkinson & Peel, 1998; Metters et al., 1997). These usually require both medical professionals and NHS Regional Trusts Managers to re-learn how to work together. Orlikowski (1994) and Edmondson (2003) suggest that one technology can be seen differently by two groups of people in an organisation. Findings from Barney and Griffin (1992) and Orlikowski (1994) have showed how this could result in the elicitation of different responses for members of that organisation.

SCOPE OF PROJECT WORK

This article takes a more in-depth look at the role of the NHSIA (seen as the project leader for the NPfIT), currently the most visible spokesperson and translator for the potential implications of the resulting new technologies. Research has shown NHS IS staff to pay particular attention to what the NHSIA says and does in regards to information systems (Collins, 2003; Ferlie & Shortell, 2001). This research builds on a framework that identifies the key dimensions of the NHSIA tactic that is situation-specific for NPfIT as assumptions. The work looks at the NHSIA goal and roles for the NPfIT, as well as the role of the private sector service providers in the implementation of NPfIT.

Here are a few objectives the NPfIT hopes to accomplish:

- To have a series of tightly specified and priced framework contracts on a short list (of about five) primary service providers (PSPs) who can work at the regional/local Strategic Health Authority (StHA) level. This should enforce the integration and implementation partnership — at a national level — during all aspects of the NPfIT project. Each PSP will have an aligned consortium of service providers and vendors for the integrated care resource service element of the NPfIT, and will be mandated to work with the domain PSP for electronic booking the infrastructure providers and healthcare providers. StHA PSPs may not make their products exclusive or mandatory to their StHA.
- To create priced packages of national services and applications that the PSPs and StHAs can together implement locally. This activity will include managing the creation of a single HRIS and other national services, to access and move health record information as required.
- To create Service-Level Agreements for the national services and other services out of the scope of the PSP consortium which the PSPs can work towards in providing an integrated service to the StHA.
- To develop and maintain the national standards and specifications that all vendors must use. It is also anticipated to create the national business cases required for the Department of Health governance (required by the National Treasury), and to support the local decision-making business cases required at the StHA level.
- To procure, under national contracts, a backbone network infrastructure.

Such an arrangement provides the greatest clarity in respect to the appropriate allocation of responsibilities, and should be well understood in the public and private sectors (see Table 1). Services will be procured on a long-term basis so the combination of local and central funding will be required for at least five, and preferably 10, years at guaranteed levels.

THE RESEARCH STUDY

The findings reported in this article are part of a larger, five-year research study that was developed to investigate the deployment, hosting, and integration of the ASP and WS technologies from both a supply-side and demand-side perspective. The overall research was in two phases. The first phase, comprising a pilot study, was conducted in the U.S. and UK (Currie & Seltsikas, 2001). An exploratory-descriptive case study methodology (Yin, 1994) was used to investigate 28 ASP vendors and seven customer sites in the UK. The dual focus upon supply-side and demand-side was critical for obtaining a balanced view between vendor aspirations about the value of their business models, and customer experiences, which may suggest a less optimistic picture. The unit of analysis was the business model (Amit & Zott, 2001), not the firm or industry level, so a case study methodology was anticipated to provide a rich data set for analysing firm activities and behaviour (Currie, Desai, & Kahn, 2004).

The result from the pilot study led to the funding of two additional research studies by the Engineering and Physical Sciences Research Council (EPSRC) and the Economic and Social Research Council (ESRC), respectively. Industrial collaborators were selected for the roles of technology partner, service providers, and potential or existing customers. These studies were concerned with identifying sources of value creation from the ASP business model and WS technologies in different vertical sectors (including health).

Research Methodology

The research followed a number of stages involving the use of both qualitative and quantitative data collection techniques and approaches (Walsham, 1993). A questionnaire survey was distributed by e-mail to businesses and healthcare organisations all over the UK. These organisations were listed on a national database maintained by the NHSIA,

Table 1. PSP implementation timetable (as of July 2002)

Activity / output	***Target Date***
Agree Procurement Strategy (DoH & local health authorities)	End Jul 2002
Service requirement finalized and approved	End Sep 2002
Outline Business Case developed and approved	End Sep 2002
OJEC Advert	Oct 2002
Procurement of systems and implementation services for electronic booking begins	Oct 2002
National long list PSPs created	Dec 2002
Invitation to Negotiate Issued	Jan 2003
National shortlist PSPs created	Apr 2003
First local health authorities begin detailed planning with PSPs	Aug 2003
PSP Framework Contract finalized	Oct 2003
Infrastructure provider(s) contract agreed	Oct 2003
First local health authorities begin implementation	Nov 2003
Infrastructure migration begins	Mar 2004

plus those maintained by the university. To ensure relevant managers and practitioners responded, the cover letter clearly stated the purpose of the questionnaire and requested that it be passed on to the person(s) with responsibility for managing healthcare e-business strategy.

The Questionnaire

Scales to address the research questions were not available from the literature, so the questionnaire was developed based on the theory of strategic value (Banker & Kauffman, 1988). It included a checklist, open-ended questions, and a section seeking organisational data. Research questions under Part I required respondents to answer yes/no if the application of Internet technologies in healthcare was bringing value to patients. Data in Part II of the questionnaire was collected by open-ended questions seeking respondents' views on the best approach to healthcare performance improvement and WS value creation. This line of questioning was used to increase the reliability of data, since all respondents were asked the same questions, but some added additional information. The purpose was to impose uniformity across the sample of representation, rather than to replicate the data obtained from each participant (Yin, 1994).

ANALYSIS

As a result of the first phase of the research carried out in the U.S. and UK, a database was developed of 700 international firms, all of which had developed an ASP business model. The data served to build-up a 'data-bank' of market intelligence on a variety of ASP firms and their offerings. Many of these firms were tracked over a four-year period to identify changes in their business models. From this sample of 700 ASPs, about 55% continued in application provisioning (excluding infrastructure) by the time Phase II of the research began. A good number of the original sample had ceased to exist (about 24%), and some had been taken over by other firms (about 6%). The remaining (about 15%) had changed their business model, moving away from software-as-a-service (Kakabadse & Kakabadse, 2002) into data storage and managed services provisioning. The ASP database proved to be a very useful source of market intelligence for the research study since it helped to develop a questionnaire survey targeted at the healthcare sector.

In Phase II of the research, a total of 350 questionnaires were distributed to NHSIA directors, current and prospective suppliers of IS to the NHS, and medical practitioners from general practices in major cities as well as rural towns, in an effort to avoid bias. The vendors were drawn from a list of businesses negotiating for contract under a PSP scheme (Guah & Currie, 2004). After persistent follow-up (telephone calls), 225 questionnaires were completed and returned, representing a 64% total return rate. As this study was exploratory in nature and was designed to develop a framework to inform practice and guide future research, rather than testing hypotheses (Avison & Fitzgerald, 2003), the 225 organisations' situations on which it is based provided useful preliminary data.

Respondents were mainly middle-level business managers (60%) offering various Internet-related services in WS to the global healthcare industry, while the rest (40%) consisted of medical practitioners working in the NHS. The 40% included IT coordinators at different institutions nationwide, some already implementing WS architecture locally. The vendor organisations were predominantly American and British, consistent with the business structure in the UK. Seventy-eight percent of the companies were large ones with an annual turnover of over £1 billion, while 13% had between £25 million to £1 billion, and 9% were considered small with an annual turnover of less than £25 million.

Technology for gathering information is not a new thing in the NHS. Nevertheless, a few questions remain such as:

- Has IS accelerated to a point at which long-practiced denial is no longer possible?
- Has IS become so important to the NHS that a single project (NPfIT) implementation can no longer be ignored by any part of the NHS?

Nearly all our interviewees (79%) believed the cacophony of IS is approaching a critical threshold and they don't know how to prevent the volume from becoming even more deafening, let alone how to turn it off.

DISCUSSION

Earlier findings suggest that most of the software applications offered by vendors were loosely defined as horizontal business applications (i.e., accounting, human resources, fixed asset management software, etc.). Even where the NHS was deploying the Web services for more complex applications, there was little evidence that the scale and scope of usage amounted to extra time for managers to engage in other activities. Other benefits to the NHS were defined in terms of cost savings from total costs of operation (TCO). Many vendors used TCO models to show how IT costs would be reduced using a remote software delivery model. Our findings pointed to a lack of relevance in TCO models, particularly for the NHS with increasing demands on low IT budgets. One NHS IT manager said:

"Our overall IT spending is only about £150 k per annum. I don't think using a Web service solution will necessarily save us much money. But if it means fewer headaches with physicians using IT, then I would be prepared to pay more for Web services! "

Integrating Healthcare Applications with WS

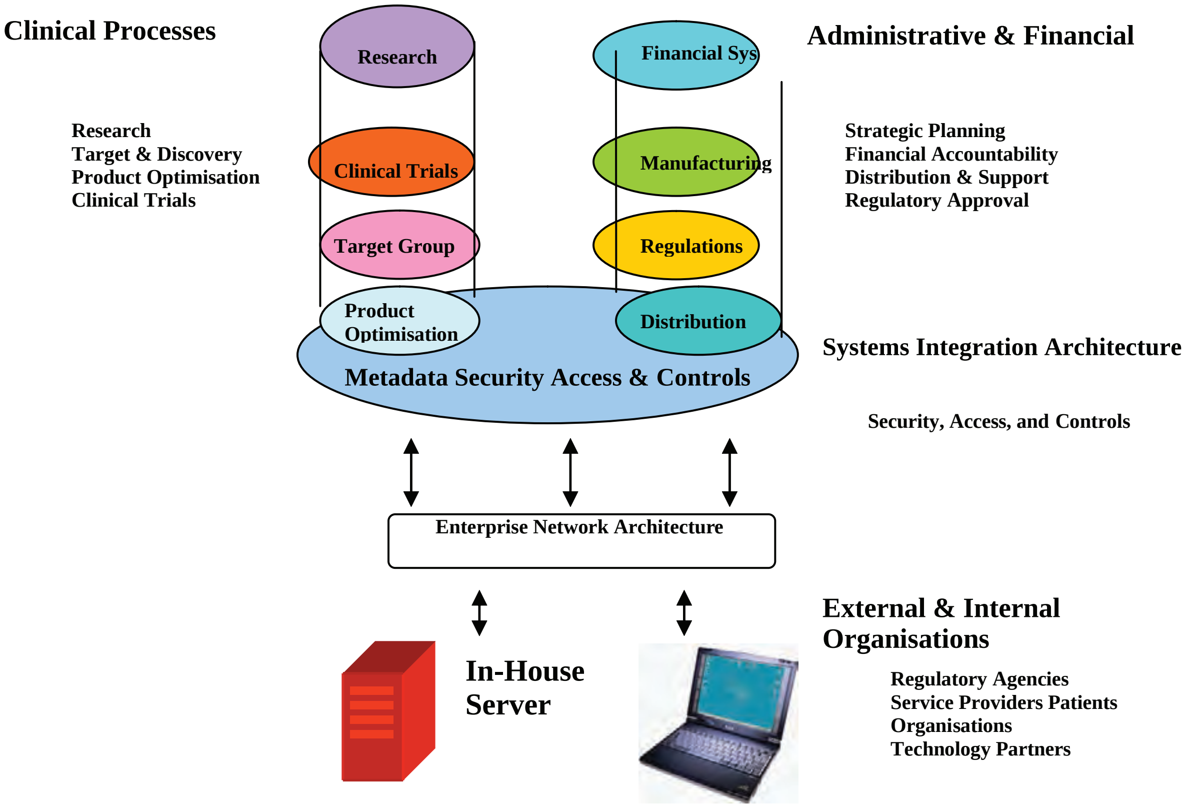
Medical practitioners among our interviewees failed to see the integration of software applications as an immediate priority. Some didn't see it as a priority they were prepared to pay for. What then is the NHS interest in WS? Many organisations in the NHS were deploying WS solutions that were non-mission-critical (e.g., e-mail, rather than patient records). One medical consultant, using an electronic diary application, commented:

We were not particularly looking for a WS solution to organise our time management. However, when a vendor approached us with a product for diary scheduling, we thought to give it a trial. The main advantage of this is that you can make changes to patient appointments from a remote location and submit the data to the surgery. It is particularly useful for staff working away from the surgery for long periods, but the current shortage of GPs in the NHS does not encourage that.

Such comments show the complexity of the outsourcing relationship between supplier and customer. It is largely effected by the number and scope of such partnerships, the sophistication of the products/services, and the type of outsourcing contract. Differentiation in products/services further suggests a more complex procedure for benchmarking service providers' performance, with measures designed to evaluate outcomes at every stage of the system development and implementation process (Ahn, Leem, & Yang, 2001).

One area where WS could potentially make a significant difference is the integration of software applications across multiple platforms, sites, and departments of healthcare organizations (Laroia, 2002). A commonly cited benefit from service providers was that software applications across business and IT functions could be integrated to fulfil the goal of enterprise application integration. Many service providers planned to develop a wide portfolio of business applications (including ERP, CRM, accounting and financials, logistics, etc.), which could eventually be integrated to provide the customer with a comprehensive enterprise solution.

Figure 1. Complex data generated by healthcare enterprises value chain



Majeed (2003) questioned if the employment of such features is sufficient to successfully deploy and manage IS in healthcare, where patient data and processes must be integrated, managed, and deployed strategically. To manage such real-time procedure requires the integration of complex enterprise information (see Figure 1) including the following:

- **Research:** Molecular, genetic, proteomic, and pharmacological data
- **Manufacturing:** Integrated global pharmaceutical information on drugs manufacturing process
- **Metadata:** Information about what is contained in these phases of development
- **Clinical:** Clinical trial, side effect, outcomes, and "real-life" effects of drugs
- *Validation*: Strategic planning data on the validation of the drug for the target patients
- **Administrative:** Healthcare claims, membership, diagnostic, and treatment data
- **Financial:** Design, development, pricing, deployment, and patient intelligence data

These have been articulated in two "stacks" of data presented in Figure 1:

1. **Clinical:** The data and information supporting medical and research and development (R&D) of new drug entities
2. **Administrative and Financial:** The data and information supporting the processes of drug approval, production, distribution, and patient care monitoring

Figure 1 describes the data model for the integration of information across the value chain of a healthcare organization that is multi-vendor and multi-source. Such an organization encourages the definition and sharing of data — within the organization and among partners and regulators.

Eisenhardt and Martin (2000) highlighted that credibility in identifying benefits from an IS is the top constraint, followed by cost justification. The NHS is aware of the potential benefits private service providers can bring to healthcare. On the other hand, it is unable to quantify those benefits and is therefore being sceptical of the scale. This scepticism may be strengthened by a degree of distrainment engendered with past experiences with IT projects.

The following are some of the factors that will promote implementation success:

- Support from trust managers
- Availability of sufficient resources
- Prime users' prior experience with similar technology
- Technology introductory procedure
- Reputation of technology in other industries (preferably other government departments)

Expressing concern about IT's great leaps forward is more than politically incorrect when a national IT project is being fully deployed (Williams, Dobson, & Walters, 1990). Most of the documentations on current discussions of the future of emerging technologies in the NHS come from insiders (scientists and technologists on scientific and technological grounds). The intention of this article is to present a view from outside the technological world, and on social and psychological grounds. The argument in this article is based on this unsettling proposition:

Information technology in the NHS has accelerated, is accelerating, and must continue to accelerate, but both medical staff and patients must become key players in that process.

The above proposition is both inevitable and scary. However, several authors have given a moderately rational basis for the position that information technology in the NHS has accelerated (Eccles et al., 2002; Jacklin et al., 2003; Keen, 1994) and must continue to accelerate (Doherty, King, & Marples, 2000; Majeed, 2003). Such basis extends well beyond the now familiar truth that more knowledge has been accumulated in the past 15 years than in the previous four decades of the NHS existence.

New systems are usually integrated into particular sub-cultures by different routes. In most medical systems, technological change is often mediated through opinion leaders, often respected doctors (Doherty, King, & Marples, 2000). Although some systems (i.e., diabetes monitoring systems) are initially adopted out of necessity, sometimes systems are adapted over time without conscious awareness that they are doing so.

On the other hand, IS — which the NHSIA seeks to promote and control — has also ensnared the NHS. Unless they keep up with global healthcare technological development, the NHSIA cannot hope to survive the violent, ever-changing mix of inter-organisational warfare and collaboration that has become their normal milieu (Leavitt, 2002). Some of our non-technologist interviewees saw the situation with IT in the NHS this way:

Between IS and the NHS, it is now no longer clear which is master and which is servant. What is clear is that both have spiralled up to positions of enormous social power.

IS's rising status and budgetary allocations signals much more than a mere shift in the NHS hierarchy. It also signals a shift in the thought-to-action ratio. IS becomes a drive shaft that turns the wheels of change in an organisation (Leavitt, 2002; Scott et al., 2003).

Few in the NHSIA hide their irritation at the frequent requests for real deliverables from past national IS initiatives. But is this enough justification for a huge project, like NPfIT, which stands to really change procedure for delivering health to the NHS users. One executive claimed that the identification of problems is the first step toward its solution. This project could presently deliver one modest deliverable that might help dampen users' resistance based on past projects. The NHSIA should learn to look further ahead at the possible effects of new technologies that would emerge from this project. Research findings haven't done a very good job of such forecasting (Gerowitz et al, 1996).

To deal with such an important part of the NPfIT, the NHSIA needs dialogue, not just among the IS departments, but among other staff and prospective users of IS in the NHS. This approach would handle the project's possible long-term social impact on its end-users. As most of the phase implementation goes ahead, the users await the transition from this achievement to its forthcoming implementation. The fact that this is one of those issues being debated vigorously sends a positive but inchoate signal to the staff and patients of the NHS (Marshall, 2003; McGauran, 2002; Robinson, 2002). However, a nationwide project costing £6 billion has become far too important to be left to one NHS agency and their private sector collaborators. There ought to be something more formal pushing the look-ahead idea further. Users are impatiently awaiting news of real actions being taken to avoid problems in the future and the consequences of things that may go wrong within the patient care process as a result of a failed NPfIT. The NHSIA should identify, at an early stage of the developmental process, the most likely social, educational, psychological, moral, and other probable fallouts of emerging technologies of the NPfIT, and propose ways of coping with them.

The Treasury sees economic rewards in promoting cost-decreasing healthcare services, given the government's own pressure to restrain inflation (Wanless, 2002). This affects the NHSIA emphasis on making each process in the NPfIT face continued pressure towards evolution or extinction. However, similar projects in the past have exemplified the process of organisational experimentation that has been unleashed by the transition to unmanaged internal competition in a NHS (Keen, 1994; Robinson, 2002). One anticipated benefit of the NPfIT is the nationwide integration of patient information, expected to moderate cost inflation. This must be an improvement on the traditional system of professional dominance and partially implemented systems of utility regulation and internally managed competition.

CONCLUSION

The original demand placed on IS in the NHS was to reduce the cost of healthcare, and not necessarily to improve quality and service. To this cause, most existing systems have responded accordingly. Patients, however, are now increasingly worried about the quality of care they receive from the NHS as a result of such cost-control measures. The NHSIA needs to shift its IS strategic emphasis to developing methods for measuring rather than improving service in a manner analogous to the process pursued in the NHS after deregulation. NPfIT should provide the NHS with a system that is subjected to systematic monitoring of quality and service levels. The intent should be of promoting clinical comparisons and quality-conscious patient choice. NHSIA is in no position to expect that a salient feature of the NPfIT will be reliance on unmonitored trust and opposition to quantitative, validated measures of performance. LSPs and NSPs should ensure satisfaction surveys as indicators of preventive services utilization. There must also be a means of tracking appropriate clinical processes and risk-adjusted measures of patient outcomes. If these are taken seriously by the NHSIA, the NPfIT will prove to hold great potential to enhance as well as simply measure the quality of care, since statistical and epidemiological methods always outperform the pick-and-choose approaches to quality improvement (Robinson, 2000).

The emerging technologies from NPfIT are expected to pioneer new methods for the collection, dissemination, and comparison of data on patients and clinical quality. While the progress to date has been frustratingly slow, it has laid the foundation for more specific, severity-adjusted, and outcomes-oriented measures in the future.

Some are afraid that the NPfIT, like other IS innovations in the NHS, it will not progress smoothly, nor always at full throttle. The message that seems to be coming from the NHSIA is that this project will surely go forward.

Despite some pessimistic views in this article, it is clear that the NHSIA strategy of private partnership has brought some benefiting systems to the NHS, which have outweighed its costs (NHSIA, 2003; Robinson, 2002). While some users look upon further IS acceleration quite positively, others have visualised an enormous technology-driven economic expansion ahead for the NHS. Most staff continues to envision a technological paradise awaiting the NHS at the end of the NPfIT. Some may even dismiss this article as just another illusion.

REFERENCES

Ahn, J.G., Leem, C.S., & Yang, J.H. (2001). *Jae Geun AhnA framework for certification and audit of application service provider. ASP, 10*(3), 239-252.

Amit, R. & Zott, C. (2001). Value creation in e-business. *Strategic Management Journal, 22,* 493-520.

Atkinson, C.J. & Peel, V.J. (1998). Transforming a hospital through growing, not building, an electronic patient record system. *Methods of Information in Medicine, 37,* 285-293.

Avison, D.E. & Fitzgerald, G. (2003). Where now for development methodologies? *Communications of the ACM, 46*(1), 79-82.

Banker, R. & Kauffman, R. (1988). Strategic contributions of Information Technology: An empirical study of ATM networks. *Proceedings of the Ninth International Conference on Information Systems.* Minneapolis.

Barney, J.B. & Griffin, R.W. (1992). *The management of organizations: strategy, structure, behavior.* Boston: Houghton Mifflin.

Brown, S. (2002). NHS finance: The issue explained. *The Guardian,* (30 May).

Brown, S.A. & Venkatesh, V. (2003). Bringing non-adopter along: The challenge facing the PC industry. *Communications of the ACM, 46*(4), 76-80.

Collins, T. (2003). Doctors attack health IT codes. *ComputerWeekly,* (6 February).

Currie, W., Desai, B., & Khan, N. (2004). Customer evaluation of application services provisioning in five vertical sectors. *Journal of Information Technology,* (forthcoming).

Currie, W.L. & Seltsikas, P. (2001). Exploring the supply side of IT outsourcing: Evaluating the emerging role of application service providers. *European Journal of Information Systems, 10,* 123-134.

Davis, F.D. (1989). Perceived usefulness, perceived ease of use and user acceptance of information technology. *MIS Quarterly, 13*(4), 982-1003.

Doherty, N.F., King, M., & Marples, C.G. (2000). The impact of hospital information support systems on the operation & performance of hospitals. *Information Systems Review, 1*(1), 97-107.

Eccles, M., McColl, E., Steen, N., Rousseau, N., Grimshaw, J., Parkin, D., & Purves, I. (2002). Effect of computerised evidence based guidelines on management of asthma and angina in adults in primary care: Cluster randomised controlled trial. *British Medical Journal, 325*(26 October), 941-944.

Eisenhardt, K. & Martin, J.A. (2000). Dynamic capabilities: What are they? *Strategic Management Journal, 21,* 1105-1121.

Ferlie, E.B. & Shortell, S.M. (2001). Improving the quality of health care in the United Kingdom and the United States: A framework for change. *Milbank Quarterly, 79,* 281-315.

Gerowitz, M., Lemieux-Charles, L., Heginbothan, C., & Johnson, B. (1996). Top management culture and performance in Canadian, UK and U.S. hospitals. *Health Services Management Research, 6,* 69-78.

Guah, M.W. & Currie, W.L. (2002). Evaluation of NHS information systems strategy: Exploring the ASP model. *Issues of Information Systems Journal, III,* 222-228.

Guah, M.W. & Currie, W.L. (2004). Application service provision: A technology and working tool for healthcare organisations in the knowledge age. *International Journal of Healthcare Technology and Management, 6*(1/2), 84-98.

Hagel, J. (2002). *Out of the box: Strategies for achieving profits today and growth tomorrow through Web services.* Boston: Harvard Business School Press.

Haines, M. (2002). *Knowledge management in the NHS—platform for change.* Retrieved November 2002 from www.healthknowledge.org.uk.

Institute of Medicine. (2002). *Crossing the quality chasm: A new health system for the 21st century.* Committee on Quality Health Care in America, National Academy Press, Washington, DC.

Jacklin, P.B., Roberts, J.A., Wallace, P., Haines, A., Harrison, R., Barber, J.A., Thompson, S.G., Lewis, L., Currell, R., Parker, S., & Wainwright, P. (2003). Virtual outreach: Economic evaluation of joint teleconsultations for patients referred by their general practitioner for a specialist opinion. *British Medical Journal, 327* (12 July), 84-92.

Kakabadse, N. & Kakabadse, A. (2002). Software as a service via application service providers (ASPs) model of sourcing: An exploratory study. *Journal of Information Technology Cases and Applications*, *4*(2), 26-44.

Keen, J. (1994). *Information management in health services.* Buckingham: The Open University Press.

Kreger, H. (2003). Fulfilling the Web services promise. *Communications of the ACM, 46*(6), 29-34.

Laroia, A. (2002). *Leveraging Web services to connect the healthcare enterprise.* Retrieved February 2002 from http://e-serv.ebizq. net/wbs/larois_1.html.

Leavitt, H.J. (2002). Technology and organization: Where's the off button? *California Management Review, 44*(2), 126-140.

Lee, A.S. (1990). Researching MIS. In W.L. Currie & G. Galliers (Eds.), *Rethinking management information systems* (pp. 7-27). Oxford University.

Majeed, A. (2003). Ten ways to improve information technology in the NHS. *British Medical Journal, 326,* 202-206.

Marshall, M.N., Mannion, R., Nelson, E. & Davies, H.T.O. (2003). Managing change in the culture of general practice: Qualitative case studies in primary care trusts. *British Medical Journal, 327,* 599-602.

McGuaran, A. (2002, June). Foundation hospitals: Freeing the best or dividing the NHS? *British Medical Journal, 324*(1), 1298.

Metters, J., Abrams, M., Greenfield, P.R., Parmar, J.M., & Venn, C.E. (1997). *Report to the Secretary of State for Health of the professional committee* *on the appeal of Mr. D. R. Walker under paragraph 190 of the terms and conditions of service of hospital medical and dental staff (England and Wales).* London: Department of Health.

NHSIA. (2003). *Annual operating plan: To be the national provider of information and infrastructure services.*

Orlikowski, W.J. & Tyre, M.J. (1994). Windows of opportunity: Temporal patterns of technological adaptation in organisations. *Organisation Science*, (May), 98-118.

Pencheon, D. (1998). Matching demand and supply fairly and efficiently. *British Medical Journal, 316*, 1665-1667.

Perseid Software Limited. (2003). *The strategic value of Web services for healthcare and the life sciences.* Accessed August 2003 from www.perseudsiftware.com.

Robinson, J.C. (2000). Deregulation and regulatory backlash in health care. *California Management Review, 43*(1), 13-33.

Robinson, R. (2002). Gold for the NHS: Good news that raises questions on consistency and sustainability. *British Medical Journal*, 324(April), 987-998.

Scott, J., Mannion, R., Marshall, M., & Davis, H. (2003). Does organisational culture influence health care performance? *Journal of Health Service Research Policy*, *8*, 105-117.

Sleeper, B. & Robins, B. (2001). *Defining Web services.* Retrieved April 2002 from www.stencilgroup.com.

Wanless, D. (2002, April). *Securing our future health: Taking a long-term view.* Final Report of an Independent Review of the Long-Term Resource Requirement for the NHS, London.

Walsham, G. (1993). *Interpreting information systems in organisations.* Chichester: John Wiley & Sons.

Willcocks, L. & Lacity, M. (1998). *Strategic sourcing of information systems.* Chichester: John Wiley & Sons.

Williams, A., Dobson, P., & Walters, M. (1990). *Changing culture* (2nd edition). London: IPM.

Yin, RK. (1994). *Case study research: Design and methods.* CA: Sage Publications.

This work was previously published in International Journal of Technology and Human Interaction, Vol. 1 Issue 2, edited by B. Stahl, pp. 48-61, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 5.12
Using Hospital Web Sites to Enhance Communication

Sherrie D. Cannoy
The University of North Carolina at Greensboro, USA

Lakshmi Iyer
The University of North Carolina at Greensboro, USA

INTRODUCTION

A large number of patients currently utilize the Internet to access healthcare-related information (Tobin, 2002). Many physician and health-related Web sites have been information portals lacking interactive services that could benefit healthcare partners through decreased costs, increased convenience, and communication. Patients typically visit Web portals to learn more about medical topics, often discussing this information with physicians.

Emerging Internet technologies can be a strategic asset for hospitals to impact physician bonding, patient self-service, and overall enterprise performance efforts. We conducted an investigation of Web sites of 10 hospitals listed in *U.S. News and World Report*'s Best Hospitals of 2004 Honor Roll, as well as a random selection of seven other hospital sites. An examination of each hospital's site was performed to determine what features were provided to enhance communication between the partners in healthcare. Partners are defined as patients and their families, referring physicians, insurance companies, vendors, pharmacies, job seekers, and the media. Communication-enhancing features are any features that have the potential to increase communication between the hospital and its partners. We focused on patient communication-enhancing features, since patients are the primary partners of healthcare entities.

BACKGROUND

Healthcare Basis

A medical Internet Usage Survey conducted by the Health on the Net Foundation (2002) found that 57.95% of the respondents had used the Internet for more than four years. Of the patients who responded, 21.62% correspond with their providers through e-mail and 75.52% have used online medical consultation services; 69.47% of physicians that responded stated that patients discuss information they found on the Internet with them, while 62.75% of those providers recommend specific information-based sites to their patients.

Healthcare can be improved through e-health services such as online patient pre-registration for admission, access to test results and medical records, insurance referrals and eligibility, access to reputable links for accurate healthcare information, patient forms and brochures, online support groups, access to clinical trial information, appointment scheduling and reminders, refill requests and authorizations, and e-mail capabilities. In implementing e-health "there is an urgent need for healthcare organizations to re-engineer their processes" (Wickramasinghe, Fadlalla, Geisler, & Schaffer, 2004), and physicians face obstacles such as "technologic barriers, resource priorities, and privacy issues" (Zingmond, Weilim, Ettner, & Carslile, 2001) that are secondary to providing excellent healthcare. Hospital and medical office sites can be enhanced to offer patient-centered services, while informing patients of their strongest services/specialties, convincing the patients that their organization is better than any general Web portal for accessing health-related information (Anonymous, 2001).

The drive for e-health initiatives can be patient driven, physician driven, or government mandated. "What is less clear is whether or not the services offered by healthcare organizations and the services that patients desire are the same" (Wilson & Gustafson, 2003). Government-driven initiatives include electronic health record implementation within the next 10 years which serves a dual-purpose: to allow hospitals and healthcare providers to access patient records in a standard format, and to decrease medical care errors. Veterans Affairs Secretary Anthony Principi stated that "one in every seven hospital admissions and 20% of lab tests occur because health records are not available to the clinician. More than one of every seven hospitalizations is complicated by medical prescription errors" (CNN, 2004). Thus, comprehensive electronic medical records which the patient, physician, or pharmacist can access are critical to reducing such errors. Access to these records may also be offered online so that patients can access them and check for accuracy or take them to a physician's office to which they have been referred for further tests.

In a related initiative, the National Quality Forum (NFQ) sponsored a National Summit on Information Technology and Healthcare Quality in March, 2002, to examine a national healthcare infrastructure. The design principles stated by NFQ to aid in the infrastructure development (NFQ, 2002) are:

1. Care based on continuous health relationships
2. Customization based on patient needs and values
3. The patient as the source of control—encourage shared decision making
4. Shared knowledge and free flow of information
5. Anticipation of patient needs

Theoretical Basis

The Internet can be considered a mass medium, and therefore, communications theories have been applied most frequently in this area (Merrill et al., 1996). Such theories applied include the learner-as-a-bucket theory in which the user searches for information and information is "poured" into the brain via a Web portal (Morris & Ogan, 1996). Another theory applied frequently is the critical mass theory, which states that the diffusion of innovation and adoption by about 20% of the population results in critical mass being achieved (Morris et al., 1996). Because use of the Internet, as well as other forms of electronic communication, is widespread now, it can be assumed that these are forms of media utilized by a critical mass.

In an article regarding physician communication skills, Kurtz (2002) outlines the domain of physician communication with patients. This could be expanded to any healthcare partners, and involves goals and approaches. Goals promote collaboration and partnership, and ensure increased accuracy of medical information, supportiveness, improving patient and physician satisfaction, and quality of healthcare. Approaches depend on the type of communication and can involve a well-conceived and delivered message versus communication for interaction, feedback, confirmation, and relationship. Kurtz (2002) also states that effective communication requires planning and should reduce unnecessary uncertainty—a critical factor in healthcare communication. Examples of Internet communication include e-mail, listserv, discussion forums/support groups, chat, and interactive site features (December, 1996). Features that enhance the communication between the doctor and the patient have not been explored. Prior studies analyzed sites of specialty practices (Smith-Barbaro, Licciardone, Clarke, & Coleridge, 2001), one specific type of communication such as e-mail (Moyer, Stern, Dobias, Cox, & Katz, 2002), Web portal information (Zhang, Zambrowicz, Zhou, & Roderer, 2004), or healthcare support groups or forums (Zrebiec & Jacobson, 2001). Our study specifically examines communication features offered on hospital Web sites.

WEB SITE ANALYSIS

Seventeen hospitals were chosen, with 10 of the Top Honor Roll hospitals (Group 1) selected from *U.S. News and World Report's* Best of Hospitals 2004 Report. Our premise was that the top 10 hospitals will be more likely to utilize technology to communicate better with patients, vendors, and insurance companies. Seven additional hospitals (Group 2) were randomly chosen from the *American Hospital Directory* (2004) for comparison of size and specialty, and site technology. Tables 1 and 2 display the selected hospitals.

Background information regarding bed sizes, hospital type, specialty areas, and accreditation were verified through the *American Hospital Directory*, the individual hospital sites, and the Directory of America's Hospitals (*U.S. News*, 2004b). Web site addresses listed in the *Directory of America's Hospitals* were selected as the URL locations to be examined.

Sites were examined in August 2004. Criteria chosen were adopted from related surveys by Zingmond et al. (2001) and Kind, Wheeler, Robinson, and Cabana (2004), and supplemented with features determined to be communication enhancing. The Zingmond et al. (2001) and Kind et al. (2004) studies examined only the type of content and quality of information in the sites, not specific features offered.

The criteria for Web site communications are listed in Table 3. Weighted features are noted by * in the descriptions of pertinent features discussed as follows.

1. **HIPAA* (Health Insurance Portability and Accountability Act):** Coding: 1 if there was a HIPAA statement only on the homepage, and 2 if the HIPAA statement was found on more than one page.
2. **International Patient Information/Languages*:** Some hospitals have an international presence and include such information as traveling, hotel information, specialty information, and so forth, as well as offering Web pages or forms in different languages. Coding: 1 if international patient information page was offered in different languages, 2 if additional pages or forms were offered in different languages.
3. **Support Groups*:** 56.34% of doctors that responded to the Medical Internet Usage HON survey recommend support groups

Table 1.

TABLE 1		
	Top 10 from Best Hospitals 2004 Honor Roll List (GROUP 1)	**Bed Size**
1	Johns Hopkins Hospital, Baltimore, MD	886
2	Mayo Clinic/St. Mary's Hospital, Rochester, MN	797
3	Massachusetts General Hospital, Boston, MA	875
4	Cleveland Clinic, OH	999
5	UCLA Medical Center, Los Angeles, CA	552
6 TIED	Duke University Medical Center, Durham, NC	758
	University of California, San Francisco Medical, CA	544
8	Barnes-Jewish Hospital, St. Louis, MO	906
9 TIED	New York-Presbyterian Hospital, NY	2,163
	University of Washington Medical Center, Seattle, WA	410

to their patients (Health on the Net, 2004). Coding: 1 if support groups were listed and phone numbers given for a contact, and 2 if a support group was offered online, even if through a linked affiliated site.

4. **External Web Sites:** The availability of links to external sites, further information, or discussion groups was analyzed.
5. **Insurance Accepted:** Physicians' offices are frequently asked about the insurance companies with which they participate. Having this information available online could enhance communication by giving patients a list to refer to if their insurance companies change.
6. **Appointment Scheduling*/Fill Out Patient Information Online*:** These features would allow for great efficiencies in the check-in process for patients. If an appointment has been scheduled online or patient information has been filled out online, the check-in process for a patient is more efficient. Patients may be allowed to fill in their information online (Code 2) or they may be able to download a form to print (Code 1), fill out, and take to the physician's office at the appointment time. Appointments were coded as 1 if only requests for appointments were offered, and 2 if appointments could be made online.
7. **Patient Access to Records:** With mandates for Electronic Medical Records evolving, accessing patient records online will be important for partners. Often, patient records and lab test results are copied and/or faxed to a physician, and this may be more

Table 2.

TABLE 2	
Randomly Selected Hospitals (GROUP 2)	**Bed Size**
Loma Linda University Med Ctr, CA	658
Miami Valley Hospital, Dayton, OH	848
Moses Cone Health Center, NC	1,141
High Point Regional Health System, NC	368
Univ of Maryland Med Center, MD	601
St Vincent Hosp & Hlth Ctr, Indianapolis, IN	806
Yale-New Haven Hospital, CT	808

Table 3. Feature coding

TABLE 3 *Feature Coding*		
	Coding	
Feature not present	0	No link to form to complete
Basic feature was present	1	Downloadable or printable form to complete manually
Weighted Feature*--Increased functionality of feature	2	Ability to fill out information online

time consuming than allowing appropriate partners access.

8. **E-Mail Doctor:** Disadvantages of offering the physician's e-mail are privacy issues, timeliness of the response, liability issues, filing insurance on e-consults, and volume of e-mail that could be received, especially for large hospitals with international presence. Advantages are that the doctor can answer at his convenience and provide patients an avenue to ask embarrassing questions or clarify instructions. The limitations are most likely from the disadvantages of offering widespread access to physicians, including a patient's inappropriate use of e-mail (in lieu of office visits), security issues, and that physicians may be inundated with e-mails from not only their patients, but others seeking advice (Mandl, Kohane, & Brandt, 1998).
9. **Referring Physician Online Access*:** Allowing physicians to fill out a form online for patients who have been referred for tests, x-rays, or admission would save time and money, as well as decrease medical errors and medication interactions. Coding: 1 if contact information was given for the referring physician to call, and 2 if the referring physician could fill the patient information online.
10. **Intranets:** These serve the purpose of communicating between internal partners of the hospital. Policies, procedures, and other hospital information can be available through the intranet. Intranets may possibly include the capability for physicians to enter prescriptions directly to the hospital pharmacy.
11. **Interactive Content:** This includes features such as virtual tours, interactive maps, and videos.

One issue to be considered was that hospitals often operate underneath an umbrella of a parent organization, which also includes other medical centers and often educational organizations. Generally, there is a seamless navigation between hospital entities, therefore the extensiveness of sites was an advantage to the user, and no differentiation between the parent organization and its entities was made. All sites examined had search features which would search anywhere within the site for requested information.

Findings

Results of the site evaluation are in Table 4. All sites included general information and features such as a search feature, information about the organization, mission, services, privacy policy, parking and directions, news, employment, support groups, and general phone contacts. The following is a summary of the findings:

1. **HIPAA (Health Insurance Portability and Accountability Act):** It was expected

that the sites would offer a main link, probably in the privacy policy, that explained HIPAA and its regulations regarding patient confidentiality. However, only five offered a direct link to HIPAA information through a HIPAA link or through the privacy policy. Searches of some sites found no results of "HIPAA."

2. **International Patient Information/Languages:** Group 1 offered more information in different languages. This is understandable, since hospitals on the Honor Roll list are larger and have an international presence. The top four of the Honor Roll list (Johns Hopkins, Mayo Clinic, Massachusetts General, and Cleveland Clinic) allowed a user to select a language in which to view the International Patient Information page.
3. **Support Groups:** Of the Honor Roll hospitals, 40% provided a main link for access to support group information, while 85.71% of the Group 2 hospitals provided a main link. However, all sites provided at least a link to support groups in each specialty area (cancer, etc.).
4. **External Web Sites:** 82% of the sites provided external links. Links to outside sites were generally to government or health organization sites. Sometimes the information given by hospital sites was generated by a purchased knowledge database, and it was difficult to determine if the information was generated by an outside site or a knowledge database. Knowledge database information was considered as internal information and not as a hyperlink to an external site. Only one hospital, New York Presbyterian, required the user to read a disclaimer and click to accept before accessing external health information.
5. **Insurance Accepted:** 70% in Group 1 provided insurance participation information as compared to 28.57% in Group 2. Many sites allowed searching for a doctor according to the insurance with which they participated.
6. **Appointment Scheduling:** Since the intricacies of appointment scheduling are vast, and result from specialty variations as well as individual physician preferences, some hospitals allowed requests for appointments in which someone would contact the patient to set up the appointment. Hospitals from Group 1 were more likely to offer appointment requests online, with 80% providing the service, as compared to 28.5% in Group 2.
7. **Patient Access to Records:** Only three hospitals (Cleveland Clinic, Duke University Medical Center, and Barnes-Jewish Hospital) allowed some type of access to the patient's records, including prescription refill requests, appointment requests, patient information updates, appointment reminders, online payments of balances, and self-assessments of healthcare.
8. **E-Mail Doctor:** None of the hospitals in Group 1 and only three in Group 2 had the capability to e-mail physicians directly.
9. **Referring Physician Online Access:** 71% of all sites allow referring physicians to fill out patient information online. This would allow the hospital time to verify insurance information and send materials and brochures to patients before being admitted.
10. **Intranets:** The presence of an intranet was difficult to determine, since the intranet may be a different URL than the hospital URL. Of all hospitals, 41% had some type of intranet available. The intranet at Cleveland Clinic was confirmed to have physician access to medical records online.
11. **Interactive Content:** The two groups had equal participation in offering interactive content, which was in the form of audio, video, or Web tours of facilities.

Table 5 shows some of the interesting interactive and weighted features.

FUTURE TRENDS

Availability of comprehensive patient information and other hospital services will improve the efficiency and effectiveness of patient care through better communication among healthcare partners. Some hospitals which were contacted regarding intranet availability were in the process of electronic record conversion. EMRs (electronic medical records) would provide a concise map of a patient's medical history, and, if accessible by various healthcare partners, including the patient, could offer a means of communication by clarifying questions or documenting previous medical problems and medications.

Blogs such as those offered on the Cleveland Clinic and High Point Regional sites are being tested as a type of forum for patients to support one another. Blogs allow personal expression and diary-type entries for those who want to share health-related experiences. Ideally, communication could be customized by allowing the user to enter a username and password which will allow specific access to that person's information and personal preferences—a patient portal. Messages and reminders from physicians may appear, as well as charts showing, for example, one's diabetic sugar level readings for the past few months. Very few hospital Web sites offer the capability to pay bills online, and access to one's medical bill history online may decrease calls to the billing department regarding balances owed. Patients may opt to receive automatic e-mails from physicians and healthcare partners for appointment reminders, test results, prescriptions, or to inform the patient if the pharmacy has the medication in

Table 4. Analysis of results

	TOP 10 HOSPITALS (GROUP 1)		SELECTED HOSPITALS (GROUP 2)		
FEATURE	% with features	% with weighted features	% with features	% with weighted features	% of ALL hospitals with features (1 or 2)
Patient Services					
HIPAA Information	80%	40%	57%	14%	71%
International patient information	90%		29%		65%
Different languages	70%	40%	29%	0%	53%
Health-related information (Diagnosis, Procedures, etc.)	90%		100%		94%
Links to online resources/outside web sites	80%		86%		82%
Access to support groups	100%	40%	100%	86%	100%
Events and classes	90%		86%		88%
Find a doctor	90%		86%		88%
Information about the physicians and staff	90%		71%		82%
Clinical Trial Information	100%		71%		88%
Insurance Accepted	70%		29%		53%
Check-in information	90%		57%		76%
Fill out patient information online/Download form to fill out	20%	10%	29%	0%	24%
Appointment Scheduling--Phone/OnlineRequest/OnlineForm	80%		29%		59%
Access to Test Results	10%		0%		6%
Access to medical records	30%		0%		18%
Access to billing information	20%		14%		18%
Email a Patient	20%		43%		29%
See pictures of newborns	10%	0%	57%	43%	29%
Contact Information					
Email availability					
Doctor	0%		43%		18%
Specialty/Dept.	10%		29%		18%
Billing	10%		29%		18%
General	70%		71%		71%
Phone Contact					
Doctor/Dept.	100%		86%		94%
Billing	100%		86%		94%
General	100%		100%		100%
Physician Services					
Referring Dr	70%	30%	71%	57%	71%
Physician/Nurse Training (CME)	100%		71%		88%
Intranet	40%		43%		41%
Access to patient records via web	0%				6%
On-call schedules (hospitals)	0%				
Special Features:					
Interactive content	70%		100%		82%
Web tours of facilities	10%		71%		35%
Videos	20%		57%		35%
Extranet	0%		0%		0%

Table 5. Sample of interactive features

TABLE 5
Sample of Interactive Features

Hospital	Features
Massachusetts General	Live surgery webcasts and e-consults for second opinions
Cleveland Clinic	e-clinic for second opinions
University of California, SF Medical	Sample bills available online
Loma Linda	e-cards and radio shows
Moses Cone Health System	Excellent library sources
High Point Regional Health System	Request callback; Extensive search feature--can search for a physician by name, specialty, gender, insurance accepted
University of Maryland Medical Center	Webcasts
St. Vincent Hospital	MyStVincent--access to personal information; can update insurance information online
Yale-New Haven	Call Me Later feature

stock. Telemedicine and videoconferencing are popular in other countries such as Australia and New Zealand, and could provide healthcare to those patients in remote areas, if the technology were provided.

CONCLUSION

Communication between physicians and healthcare partners, specifically patients, is critical to the quality of care provided. Forms of online communication need to be identified and researched further so that efficiencies and effectiveness can be gained for healthcare organizations' quality of care goals. Because patients are active in their medical care, healthcare organizations need to be aware of patient needs and preferences in their communications. What needs to be determined is how users decide to search for information on a specific hospital site. Which features are preferred by both physician and patient? As Bohn (1999) stated, "As sites grow ever more central to successful healthcare business, analysis of the way they are used becomes a fundamental issue." In this respect, the functionality of sites needs to be addressed—is it enough to offer such interactive features and anticipate that they will be found and utilized, or does this interactive functionality become part of emerging culture of healthcare? The opportunity to utilize Web technology to communicate with partners allows hospitals to become more strategically competitive while also increasing efficiency and effectiveness, and ultimately improving patient care.

REFERENCES

American Hospital Directory. (2004). Retrieved August 2004, from http://www.ahd.com/

Anonymous. (2001, April 10). *Expert: Hospital Web sites require sophisticated expectations*. Retrieved August 7, 2004, from http://www.healthdatamanagement.com/html/PortalStory.cfm?type=trend&DID=5426

Bohn, L. (1999). What's your hospital's Web site worth? Methods for hospitals to build an effective Web site. *Health Management Technology, 20*(8), 22.

CNN. (2004, April 28). *Bush calls for electronic medical records*. Retrieved from http://www.cnn.com/2004/ALLPOLITICS/04/27/bush.healthcare.ap/

December, J. (1996). Units of analysis for Internet communication. *Journal of Communication, 46*(1), 14-38.

Health on the Net Foundation. (2002). Evolution *of Internet use for health purposes (survey).* Retrieved July 2004 from http://www.hon.ch/Survey/Spring2002/res.html

Kind, T., Wheeler, K., Robinson, B., & Cabana, M. (2004). Do the leading children's hospitals have quality Web sites? A description of children's hospital Web sites. *Journal of Medical Internet Research, 6*(2), 20.

Kurtz, S. (2002). Doctor-patient communication: Principles and practices. *Canadian Journal of Neurological Science, 29*(Supplement 2), S23-S29.

Mandl, K., Kohane, I., & Brandt, A. (1998). Electronic patient-physician communication. *Annals of Internal Medicine, 29*(6), 495-500.

Morris, M., & Ogan, C. (1996) The Internet as mass medium. *Journal of Communication, 46*(1), 39-50.

Moyer, C., Stern, D., Dobias, K., Cox, D., & Katz, S. (2002). Bridging the electronic divide: Patient and provider perspectives on e-mail communication in primary care. *The American Journal of Managed Care, 8*(5), 427-433.

National Quality Forum. (2002). *Information technology and healthcare quality: A national summit.* Retrieved from http://www.qualityforum.org/ITvision.pdf

Smith-Barbaro, P., Licciardone, J., Clarke, H., & Coleridge, S. (2001). Factors associated with intended use of a Web site among family practice patients. *Journal of Medical Internet Research, 3*(2), 17.

Tobin, M. (2002). *Survey of respondent use of the medical Internet and the impact on the patient/physician relationship.* Retrieved April 2005 from http://www.mikety.net/Articles/InternetSurvey/IS10y.html#Tables

U.S. News and World Report. (2004a). *Best hospitals of 2004: Honor roll.* Retrieved August 2004 from http://www.usnews.com/usnews/health/hosptl/honorroll.htm

U.S. News and World Report. (2004b). *Directory of America's hospitals.* Retrieved August 2004 from http://www.usnews.com/usnews/health/hospitals/hosp_home.htm

Wickramasinghe, N., Fadlalla, A., Geisler, E., & Schaffer, J. (2004). A framework for assessing e-health preparedness. *Proceedings of the 10th Americas Conference on Information Systems.*

Wilson, E., & Gustafson, N. (2003). Modeling patient's acceptance of e-health. *Proceedings of the 9th Americas Conference on Information Systems.*

Zhang, D., Zambrowicz, C., Zhou, H., & Roderer, N. (2004). User information seeking behavior in a medical Web portal environment: A preliminary study. *Journal of the American Society for Information Science and Technology, 55*(8), 670-684.

Zingmond, D., Weilim, Y., Ettner, S., & Carslile, D. (2001). Information superhighway or billboards by the roadside? An analysis of hospital Web sites. *The Western Journal of Medicine, 175*(6), 385-391.

Zrebiec, J. F., & Jacobson, A. M. (2001). What attracts patients with diabetes to an Internet-support group? A 21-month longitudinal Web site study. *Diabetic Medicine, 18,* 154-158.

KEY TERMS

Blog: A Web log which allows for a diary-style forum for posting one's personal thoughts and experiences.

E-Health: A concept in which Internet-based technologies are used to enhance access, delivery, quality, and effectiveness of health services, as well as information utilized by patients, physicians, healthcare organizations, and related partners such as pharmacies, vendors, and insurance providers.

Extranet: Allowing an organization's external partners (i.e., pharmacies, insurance companies, vendors) to access a computer system via a username and password.

HIPAA: The Health Insurance Portability and Accountability Act (1996) regulates the electronic exchange, privacy, and security of one's health information.

Intranet: Allowing access to a hospital's computer system to internal users via a username and password.

Privacy Policy: A disclaimer placed on Web sites informing users about how the organization handles one's personal information.

Telemedicine: The delivery of healthcare through the Internet or other computerized means such as videoconferencing.

This work was previously published in Encyclopedia of E-Commerce, E-Government, and Mobile Commerce, edited by M. Khosrow-Pour, pp. 1139-1145, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 5.13
The Internet, Health Information, and Managing Health:
An Examination of Boomers and Seniors

Christopher G. Reddick
The University of Texas at San Antonio, USA

ABSTRACT

This article examines the use of the Internet for gathering health information by boomers and seniors. This study attempts to determine whether online health seekers (individuals that have Internet access and have searched for health information online) have changed their behaviors from the information they found online. Essentially, has online health information helped them to manage their health more effectively? This research analyzes the Kaiser Family Foundation e-Health and the Elderly public opinion dataset of access by boomers and seniors to online health information. The major results indicate that boomers marginally use online health information more than seniors for the management of their health. The most significant results indicated that boomers and seniors who are more aware and have positive feelings toward online health information would use it more to manage their health.

INTRODUCTION AND BACKGROUND

For baby boomers, the Internet has become the most important source of health information other than consultation with their family doctor (Kaiser Family Foundation, 2005). The focus of this article is on both baby boomers, those in the age range of 50 to 64, and seniors, or those 65 and older.[1] This study examines the use of online health information by baby boomers and seniors and how they use the information for managing their health. The primary objectives of this article are to examine the differences in behavior between boomers and seniors and to test for the presence of a variety of associations among their characteristics and a number of management of health variables.

This study explores five specific questions. First, are there any differences between boomers and seniors and their access to health information for managing health? Second, will healthier boomers and seniors rely less on online health information in order to manage their health because they would have less need? Third, will the presence of boomers and seniors that have more experience and familiarity with the Internet lead to greater use of online health information to manage health? Fourth, will individuals who are in a lower sociodemographic status rely less on online health information because of lack of resources to access this information? Finally, will avid Internet users use online health information more often to manage their health because they would have greater access to and familiarity with the Internet?

The American health care system is different from many Western countries, since it is administered primarily by the private marketplace. The majority of the United States population contracts with a private provider for his or her health insurance coverage. Medicare is a federal health insurance program for people age 65 and older. In addition, Medicaid, a program sponsored by the federal government and administered by states, is intended to provide health care and health-related services to low-income individuals. However, there are millions of Americans who do not fit into either the Medicare or Medicaid plans and, essentially, remain uninsured. Online health information is especially important, given the millions of uninsured Americans trying to get information on their health situation. Individuals can use this online health information to make informed choices on their health care needs. They potentially can use information on the Internet to better manage their health.

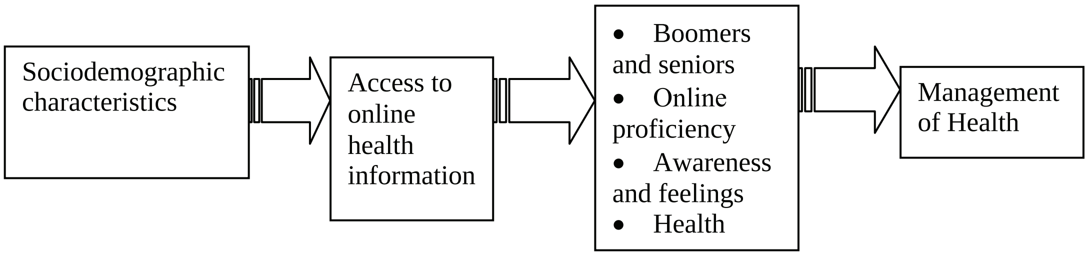
Essentially, has online health information influenced the behaviors of boomers and seniors with respect to their health care needs? This influence could be as extensive as visiting a doctor or simply talking to family or friends about health information that a boomer or senior found online.

Access to timely and reliable information on health and health care long has been a goal for seniors, who face a greater number of health conditions and use prescription drugs and health care services at a higher rate than younger adults (Kaiser Family Foundation, 2005). However, the online behavior of seniors has not been studied as closely as that of health information searches of adolescents (Gray, Klein, Noyce, Sesselberg, & Cantrill, 2005), women (Pandey, Hart, & Tiwary, 2003), cancer patients (Eysenbach, 2003; Ziebland, 2004), those affected by the digital divide (Skinner, Biscope, & Poland, 2003), and those that compare online and off-line behavior (Cotton & Gupta, 2004). There is little empirical research that examines whether online health searches affect the management of health (Lueg, Moore, & Warkentin, 2003; Nicholas, Huntington, Williams, & Blackburn, 2001), one of the two objectives of this study. This study measures whether Internet health information changed the self-reporting behavior of boomers and seniors and does not specifically address change in health outcomes.

There are two reasons why this study does a comparison of both boomers and seniors. First, baby boomers represent future seniors, and by examining this age group, this study can provide some indication about what the future holds for the Internet and health information. Second, both boomers and seniors are in the greatest need of health information, since they are more prone to have health problems than other age groups.

This study is different from existing works of Nicholas et al. (2001), Lueg et al. (2003), and Huntington et al. (2004), since it focuses on the use of online health information in the management of health. This study focuses especially on comparing two groups, boomers and seniors, while the existing empirical work examines the entire Internet population. This study is different from studies that conduct a meta-analysis, which combine published results from different sources

Figure 1. Conceptual framework of access to online health information by boomers and seniors



(Eysenbach, 2003). This research performs a statistical analysis that leads to conclusions that are different from the original dataset (Kaiser Family Foundation, 2005). The aim of this study is not just to learn about the differences between boomers and seniors and access to online health information; it is to discern the magnitude of differences between these groups and the impact of factors such as awareness and feelings on health management.

In order to accomplish the goal of examining online health information and the management of boomers' and seniors' health, this article is divided into several sections. First, this research examines the literature on the use of the Internet as a health information source. Second, this article outlines how the literature can be summarized into hypotheses that model the most probable impacts on management of boomers' and seniors' health. Third, this research provides details of the Kaiser Family Foundation's e-Health and the Elderly dataset that is used to model public opinion data of online health information (Kaiser Family Foundation, 2005). The fourth and fifth sections discuss the models and results of tests on the use of online health information for health care management. The sixth section provides a discussion that outlines how the test results confirm or deny the specified hypotheses and shows the broader significance of this work. The last section provides avenues for future research and presents limitations of this study.

LITERATURE REVIEW

The following section outlines the common themes found in the literature on the Internet and health information and the management of health. They can be divided into the factors of differences between age groups, the health of the individual, online proficiency, sociodemographic characteristics, and awareness and feelings about online health information. Existing research shows that little is known about how Internet usage, health status, and sociodemographic characteristics affect health information seeking (Cotton & Gupta, 2004).

Eysenbach (2003) provides a conceptual framework of the possible link between Internet use and cancer. Some of the important factors, according to that author's meta-analysis, indicate that Internet use is related to communication, community, and content, leading to an impact on cancer outcomes. In a similar line of inquiry, a study by Lueg et al. (2003) provided a conceptual framework of Internet searches for online health information. These authors examine the situations with which individuals find themselves confronted in terms of health needs and frequency of use, predicting access to health information. Eysenbach's (2003) conceptual framework is different from Lueg et al. (2003), in that the former examines the social aspects of Internet use and health information, while the latter study focuses on the situation involvement and the frequency of Internet use. The conceptual framework of this study is similar to Eysenbach (2003) and Lueg et al. (2003) but differs, in that it examines boomers and seniors and factors such as frequency and satisfaction having an influence on the management of health (Figure 1).

Online Health Information and Managing Boomers' and Seniors' Health

If the Internet can be used to change the behavior of individuals, this is one assessment of the long-term utility of this information resource. If individuals just look at information online and do not use it in any substantial way, it does not make much sense to invest in Internet health resources. The Internet suggests a remarkable change from the traditional doctor-knows-best approach (Eysenbach & Jadad, 2001). The Internet can be seen as challenging hierarchical models of information sharing, in which the provider of the information decides how the information should be delivered (Ziebland, 2004).

For example, existing research in a 2001 survey showed that 44% of online health information seekers said that the information they found online affected a decision about how to treat an illness or to cope with a medical condition (Fox & Rainie, 2002). A majority of respondents to a survey of adolescents and use of Internet health information reported that it helped them to start a conversation with a lay or professional medical person (Gray et al., 2005). Online health information seekers mostly are going online to look for specific answers to targeted questions (Fox & Rainie, 2000). In addition, four out of 10 young people say that they have changed their personal behaviors because of health information that they obtained online (Rideout, 2002). In an Internet survey, more than one-third of the respondents said that their conditions had improved after having visited a Web site, and more than one in four said that the Web information had resulted in a deferred visit or had actually replaced a visit to the doctor (Nicholas et al., 2001). Therefore, the existing research has examined adolescents' and all age groups' behavioral changes but has not focused on seniors and their use of the Internet for managing health. There are five factors outlined in the literature, which are differences between boomers and seniors, the health of boomers and seniors, their online proficiency, their sociodemographic characteristics, and awareness and feelings toward online health information that are predicted to have an impact on whether online health information is used.

Boomers' and Seniors' Differences

Existing empirical evidence shows that health seekers are proportionately more middle-aged than very young or old, with the highest proportion of usage witnessed in those between the ages of 30 and 64 (Fox & Rainie, 2000). A more recent survey indicates that 70% of baby boomers has gone online in 2004, while only 31% of seniors has gone online (Kaiser Family Foundation, 2005). Boomers will retire shortly, and the amount of online health information for which they will search should increase dramatically compared to what seniors are currently consuming. In order to explore both the present and what the future will hold for online health information and seniors, it is important to compare both age groups. In addition, individuals over the age of 50 may have a greater need for more information on health care than someone much younger because of the greater chance of facing health problems (Brodie et al., 2000).

Health of Boomers and Seniors

The literature also mentions that individuals who are in worse health will want to search more for online health information. The Internet becomes an additional tool in order for them to search for health information. Empirical evidence shows that there is a link between an individual's health and his or her need for online health information. Less healthy individuals are more likely to explore different aspects of a Web site and to use more health-related interactive features and, in doing so, improve their well being (Lueg et al., 2003). Individuals who were suffering with an illness were two and half times as likely, compared to respondents without a standing illness, to say that they had used information from the Internet as an alternative to seeing their general practitioner (Nicholas et al., 2001). In many cases, information seekers were acting on behalf of others, such as family and friends. However, access to online health information also should be related to the consumer's ability to use the Internet, not just on whether they are healthy.

Online Proficiency

The ability to use the Internet also should have an impact on whether boomers and seniors use online health information to manage their health. Individuals who use Internet information more to manage their health have broadband Internet access, are frequently online, spend many hours online, and search for information on many different topics. Research shows that individuals using a Web site regularly were more likely to have said that the information was helpful (Nicholas et al., 2001). A survey of adolescents shows that there are issues of the disparity of Internet access and quality of Internet access such as dial-up vs. broadband connection (Skinner et al., 2003).

Sociodemographic Characteristics of Online Health Seekers

Another factor explored in the literature that should have an impact on access to online health information is the sociodemographic characteristics of the individual. There is research on the digital divide, between the haves and the have-nots of Internet access. This research predicts that those who have greater access to the Internet would have more resources in society. For instance, those groups of individuals who are more disadvantaged economically in the United States would have less access to the Internet and online health resources. Hispanics, the largest minority group in the United States, traditionally have had less Internet access (Fox & Rainie, 2000). Those with medium to high family incomes should be able to access the Internet more for health information because of greater resources.

Existing research shows that individuals who are older, have lower incomes, are minorities, are less educated, and are males will be less likely to use the Internet for health information seeking (Cotton & Gupta, 2004; Anderson, 2004). In contrast, women increasingly rely on the Internet to supplement health information received from traditional sources (Pandey et al., 2003) and are more likely than men to seek online health information (Fox & Rainie, 2000; Nicholas et al., 2001). Awareness and feelings toward online health information also should have an impact on using this information to manage consumers' health.

Awareness and Feelings Toward Online Health Resources

A final factor that should explain access to Internet health information is the awareness and feelings of the individual toward online health information. If boomers and seniors have more positive feelings about the Internet as a health information resource, they will utilize it more often than someone who harbors more negative feelings toward the Internet. In addition, individuals who go online for health information frequently should use it more to manage their health. If a boomer's or senior's doctor or medical professional recommends or uses the Internet as a communication device, the patient is more likely to use it to manage his or her health. In summary, the prediction is that boomers and seniors who are more aware of online health resources should use these resources more to manage their health. In addition, boomers and seniors who have positive feelings about the benefits of online health information will use this resource more to manage their health.

Empirical evidence shows that there is a relationship between using the Internet more often and accessing health information (Lueg et al., 2003). Those using the Web once a day were twice as likely to report that it helped a lot in terms of being better informed from health information found on the Web (Nicholas et al., 2001). E-mail is still a new medium for obtaining access to consumer health information and also is explored in the research as a way to manage a consumer's health (Huntington et al., 2004). The literature just outlined can be formally specified with the following hypotheses that demonstrate the relationship between boomers and seniors, the Internet, and the management of health.

HYPOTHESES

In order to examine whether online health information has affected the choices that individuals make in managing their health and the differences between boomers and seniors, several hypotheses are tested in this article. These hypotheses are derived from the literature mentioned in the previous section and are divided into five areas:

Boomers' and Seniors' Differences

Hypothesis 1: Online health seekers who are baby boomers are more likely to believe that online health information has helped them to manage their health better compared with seniors.

Health of Boomers and Seniors

Hypothesis 2: Online health seekers who are healthy or who have family and friends that are healthy will rely less on online health information because of lack of need.

Online Proficiency

Hypothesis 3: Online health seekers who have broadband Internet access will go online more for health information to manage their health.

Hypothesis 4: Online health seekers who go online more often and conduct many online activities will use Internet health information more to manage their health.

Sociodemographic Status

Hypothesis 5: Boomers and seniors who are females will rely more on online health information.

Hypothesis 6: Boomers and seniors who are college educated will rely more on online health information.

Hypothesis 7: Boomers and seniors who are Hispanics will go online less for health information.

Hypothesis 8: Boomers and seniors who have family income above $75,000 will go online more for health information.

Awareness and Feelings Toward Online Health Resources

Hypothesis 9: Online health seekers who most of the time and always look to see who provides medical information on the Internet will use online health information more to manage their health.

Hypothesis 10: Online health seekers who access health information online once or twice a month will have a greater likelihood of using online health information to manage their health.

Hypothesis 11: If a doctor has recommended a Web site to an online health seeker, he or she is more likely to use health information to manage his or her health.

Hypothesis 12: If an online health seeker has communicated with his or her doctor via e-mail, he or she is more likely to use online health information to manage his or her health.

Hypothesis 13: Online health seekers that have more positive feelings about looking for health information on the Internet are more likely to use this information to manage their health.

These hypotheses are examined with a dataset that surveyed public opinion of both baby boomers and seniors on their use and acceptance of online health information.

DATASET AND METHODS

The e-Health and the Elderly dataset is a nationally representative random digit dial telephone survey of 1,450 adults age 50 and older.[2] Included in this sample were 583 respondents age 65 and older. The survey was designed by Kaiser Family Foundation (KFF) (2005) in consultation with Princeton Survey Research Associates (PSRA), and the survey was administered in the field by PSRA. The survey interviews were conducted between March 5 and April 8, 2004. The entire dataset of 1,450 respondents was first examined to determine the characteristics of boomers and seniors and access to online health information.

Out of the 1,450 responses to the survey, this study also has taken a subsample of 628 respondents, of which there were 464 boomers and 164 seniors surveyed. Therefore, the original dataset was split, and the sample sizes differ for both age groups. The 628 boomers and seniors represent those individuals who are called online health seekers. They both have Internet access and have looked for online health information. This group is of interest, since in this study, there is a comparison of the characteristics of those that actually look up online health information.

Table 1. Boomers and seniors who go online for health information

Go online for health information (Yes or No)	Age group	Frequency	Percent
No	50-64	335	42.5
	65+	454	57.5
	Total	789	100
Yes	50-64	464	78.2
	65+	129	21.8
	Total	593	100

In this study, we use a consumer survey to explore the differences between boomers and seniors and their use of online health information to manage health. This research uses both descriptive statistics and logistic regression to explore differences in access to online health information between boomers and seniors.[3]

DESCRIPTIVE STATISTICS OF BOOMERS AND SENIORS AND ONLINE HEALTH INFORMATION

In order to model the relationship between seniors and boomers, online health information, and its impact on managing health care needs, this study has specified the following variables that will comprise the models tested.

Table 1 provides information on boomers and seniors who go online for health information. Boomers that go online for health information represent 78.2%, while seniors that go online for health information represent just over 21% of those surveyed in this category. This table generally supports the notion that boomers tend to go online more for health information than seniors. Boomers that do not go online for health information represent 43%, and seniors that do not go online represent 58% of those surveyed in this category.

Table 2. Demographic information of boomers and seniors and going online for health information

Go online for health information (Yes or No)		N	Mean	Standard Deviations
No	College educated	822	0.15	0.36
	Gender is female	822	0.65	0.48
	Race is Hispanic	822	0.03	0.18
	Family income 2003 above $75,000	822	0.07	0.25
	Age	822	68.79	12.41
Yes	College educated	628	0.43	0.50
	Gender is female	628	0.61	0.49
	Race is Hispanic	628	0.03	0.17
	Family income 2003 above $75,000	628	0.27	0.45

Table 2 outlines demographic information of boomers and seniors that go online and do not go online for health information. The digital divide is very evident with the data presented in this table. For instance, 43% of college-educated individuals go online for health information compared with only 15% who are college-educated that do not go online for health information. Among females and Hispanics, there is not much of a difference in the percentage who go online and do not go online for health information. However, boomers and seniors that have a family income above $75,000 in 2003 were more likely to go online for health information. Finally, age seems to have an impact on accessing online health information. The mean age was 61 years for individuals that go online and 69 years for consumers who do not go online for health information. Higher income implies greater use of online health information, and having a college education means a greater likelihood of going online for health information. This finding also indicates that boomers are more likely to go online for health information, since the average age range was just over 61 years old.

Logistic regression is used to test whether sociodemographic variables predict whether boomers or seniors go online for health information. Logistic regression was used, since this study models dependent variables that are binary, represented by either a 1 or 0 (Nicholas et al., 2001; Lueg et al., 2003). The odds ratio can be used to interpret the relative impact of the observance of a 1 in the dependent variable. Table 3 shows that almost all of the sociodemographic variables help to explain whether someone goes online for health information, with the only exception being Hispanic. For instance, having a college education means that a boomer or senior is four times more likely to go online for health information. Having a higher income indicates that boomers and seniors are two times more likely to go online for health information. However, as the age of the respondent increases, this marginally decreases the likelihood of someone going online for health information.

Dependent Variables

Table 4 provides a list of the dependent and predictor variables and also demonstrates whether there were differences between boomers and seniors in these variables. Perhaps the most important dependent variable is whether "somewhat" or "a lot" of information on the Internet has helped to take care of a senior's or a boomer's health. The mean score indicates that 59% of boomers and 46% of

Table 3. Logistic regression results of sociodemographic variables predicting going online for health information

Dependent Variable	Go Online for Health Information		
Predictor Variables	Odds Ratio	Wald Statistic	Prob. Sig.
Age	0.95	(301.47)***	0.00
College educated	4.05	(300.72)***	0.00
Gender is female	1.25	(11.49)***	0.00
Race is Hispanic	1.06	0.26	0.61
Family income 2003 above $75,000	2.25	(63.38)***	0.00
Constant	11.34	(139.67)***	0.00
Nagelkerke R-Square	0.25		

Note: significant at the 0.01 level

seniors believed that the Internet has helped them to take care of their health, demonstrating some impact on the management of their health.

The second dependent variable measures whether online health seekers had a conversation with family or friends about health information that they found online (Table 4). Family and friends who go online for health information may guide someone else as to whether they should see a doctor because of this information (Eysenbach, 2003). The results indicate that 66% of boomers said that they had a conversation with family members or friends, and only 48% of seniors said that they had this conversation about the information they saw online. There were statistically significant differences between seniors and boomers for this question with the reported F-statistic being significant at the 0.01 level, meaning that boomers were more likely to have a conversation with family and friends about health information that they found online.

The third dependent variable measures whether online health information changed the behavior of boomers and seniors (Table 4). Thirty-six percent of boomers' behaviors changed as a result of online health information, compared with 25% of seniors. This result also was shown to have a statistically significant difference between boomers and seniors at the 0.01 level. Around one-third of boomers changed their behaviors, which is a good indication that the information that they are finding is affecting their health.

A fourth management of health issue was whether boomers or seniors made a decision on treatment of an illness as a result of the information they found online (Table 4). The results showed that 34% of boomers believed that they made a decision about how to treat an illness because of information they found online, while only 26% of seniors made a decision on treatment. This result also showed a statistically significant difference between the two age groups at the 0.01 level.

Another dependent variable was visiting a doctor as a result of the health information found online (Table 4). Only 16% of boomers visited a doctor as a result of health information they found online, while 13% of seniors visited a doctor. Visiting a doctor was the least utilized change in behavior as a result of online health information.

Referring back to Hypothesis 1 on whether online health information has been used to manage a boomer's or a senior's health, this study has found that overall, there were differences between both groups of online health seekers (Table 4). The mean values for all five dependent variables were higher for boomers compared with seniors. In addition, three out of the five dependent

Table 4. Difference of means tests of dependent and predictor variables for online health seekers; boomers are significantly different from seniors

Variable Name	Mean of Boomers	Standard Deviations Boomers	Mean of Seniors	Standard Deviations Seniors	Probability Significantly Different Boomers and Seniors
Dependent Variables					
Somewhat and a lot of information on Internet helped take care health	0.59	0.49	0.46	0.50	0.07
Had a conversation family or friend about online health information	0.66	0.47	0.48	0.50	0.00
Online health information changed behavior	0.36	0.48	0.25	0.43	0.00
Made a decision about how to treat an illness because of online health information	0.34	0.48	0.26	0.44	0.00
Visited a doctor because of information found online	0.16	0.37	0.13	0.34	0.07
Predictor Variables: Health					
Excellent or very good health	0.58	0.49	0.52	0.50	0.07
Health problems index	3.95	2.45	3.80	2.41	0.48
Predictor Variables: Online Proficiency					
Broadband Internet access	0.43	0.50	0.29	0.45	0.00
Online more than 10 hrs week	0.28	0.45	0.23	0.42	0.00
Online every day	0.57	0.50	0.51	0.50	0.08
Online activities index	2.86	0.97	2.45	1.14	0.00
Predictor Variables: Online Health Information					
Most of the time and always look to see who provides medical information on Internet	0.40	0.49	0.24	0.43	0.00
Access health information online once or twice a month or greater	0.38	0.49	0.34	0.48	0.04
Doctor recommended a health or medical Website	0.06	0.23	0.04	0.20	0.13
Communicated with doctor or other health care provider through email	0.12	0.32	0.11	0.31	0.55
Positive feelings about looking for health information on the Internet Index	2.59	0.73	2.23	0.94	0.00
Negative feelings about looking for health information on the Internet	0.68	0.78	0.71	0.81	0.25

Notes: The number of observations are 464 for boomers and 164 for seniors.

variables showed statistically significant differences between boomers and seniors at the 0.01 level. With these dependent variables outlined, this research also should describe the predictor variables and their characteristics.

Predictor Variables

The predictor variables used to explain how the Internet has managed the health care of boomers and seniors also are presented in Table 4. Many of the predictor variables are represented in terms of binary numbers in order to capture the specific impacts on the dependent variables. As previously noted, this study has divided the hypotheses into the differences between boomers and seniors, the relative health of the individual, his or her online proficiency, and how active he or she is at seeking online health information. This study discerns the impact that these factors have on the management of the health care of boomers and seniors.

To see all of the predictor variables, refer to Table 4. We will only mention a few of them in this section. For instance, an index was created of the health problems that boomers and seniors or someone they know have faced in the past year. An individual who has more health problems or is concerned with someone else's health problems would score higher on the index. The health problems index indicates less than four issues that they or someone they know faced, indicated by online health seekers (out of nine possible health problems). The nine possible health problems listed were cancer, heart disease, obesity and weight loss, arthritis, diabetes, Alzheimer's, high cholesterol, osteoporosis, and mental health.

The online activities index measures the amount of activities that boomers and seniors conduct online, and the average is around two activities (Table 4). The prediction is that health seekers who conduct more online activities have a greater likelihood of using health information to manage their health because of their familiarity and comfort with the Internet. The four online activities that comprised the index were using instant messaging, reading news, buying a product, and checking the weather.

The online health information predictor variables also show the capacity of the individual to look up health information on the Internet (Table 4). Seniors are more trusting of the health information that they read online, with only 24% of seniors "most of the time" and "always" looking to see who provides medical information on the Internet. On the other hand, 40% of boomers are looking to see who provides the online health information. This difference was also statistically significant at the 0.01 level. In addition, boomers are more frequent consumers of online health information, using it at least once or twice a month, as represented by 38% of the sample. Seniors consume online health information marginally less frequently with 34% doing so.

With regard to overall positive feelings toward the Internet, there was an average score of two on an index scaled from zero to three, indicating that boomers and seniors have overall positive feelings toward the Internet as a source of health information. The index was calculated by adding up the specific responses to whether the online health seeker agreed that online health information gave them information quickly, whether it helped them feel more informed when they go to the doctor, and whether it allows them to get information from a lot of different sources.

On an index of zero being the lowest and two being the highest, less than one was found, indicating that very few online health seekers harbor negative feelings toward the Internet as a source of health information. Having positive feelings about online health information also showed a statistically significant difference between boomers and seniors at the 0.01 significance level. Similarly, this negative feelings index was calculated by adding the individual responses, if they agreed that online health information was frustrating because it is hard to find what they were searching for and if it is confusing because there is too much information. The following section tests the relationship between accessing online health information and managing a boomer's and a senior's health.

RESULTS OF LOGISTIC REGRESSION MODELS OF ONLINE HEALTH INFORMATION MANAGING HEALTH

This study uses logistic regression with five separate management-of-health dependent vari-

Table 5. Logistic regression of factors predicting whether online health information has managed a boomer's or a senior's health

Dependent Variables	Somewhat and a lot of information on Internet helped take care health		Had a conversation family or friend about online health information		Online health information changed behavior		Made a decision about how to treat an illness because of online health information		Visited a doctor because of information found online	
Independent Variables	Odds Ratio	Wald Statistic	Odds Ratio	Wald Statistic	Odds Ratio	Wald Statistic	Odds Ratio	Wald Statistic	Odds Ratio	Wald Statistic
Boomers = 1 (age between 50 to 64)	1.31	(4.49)**	1.49	(10.67)***	1.71	(14.99)***	1.66	(12.63)***	0.90	(0.36)
Health										
Excellent or very good health	1.11	(0.88)	1.64	(21.39)***	1.21	(3.06)	1.36	(7.09)***	1.10	(0.41)
Health problems index	0.99	(0.36)	1.14	(34.51)***	1.04	(3.40)	1.00	(0.00)	1.04	(1.35)
Online Proficiency										
Broadband Internet access	0.82	(2.79)	0.63	(15.51)***	0.90	(0.80)	0.68	(9.37)***	0.89	(0.51)
Online every day	1.55	(13.13)***	1.31	(5.01)**	1.82	(22.22)***	1.43	(7.15)***	1.41	(3.99)**
Online more than 10 hrs week	0.65	(9.62)***	0.89	(0.71)	0.97	(0.05)	1.46	(7.26)***	0.81	(1.39)
Online activities index	1.01	(0.01)	0.93	(1.50)	0.95	(0.55)	0.92	(1.83)	0.80	(7.75)***
Online Health Information										
Access health information online once or twice a month or greater	3.51	(115.95)***	1.59	(17.35)***	2.29	(54.99)***	3.11	(95.83)***	1.82	(16.15)***
Communicated with doctor or other health care provider through email	0.91	(0.28)	1.95	(12.45)***	1.58	(7.28)***	2.83	(35.37)***	1.86	(9.94)***
Doctor recommended a health or medical Website	0.76	(0.97)	0.91	(0.10)	1.17	(0.31)	2.13	(7.10)***	3.68	(22.13)***
Positive feelings about looking for health information on the Internet Index	2.46	(119.44)***	1.63	(46.32)***	1.85	(42.97)***	1.70	(31.68)***	1.84	(19.73)***
Most of the time and always look to see who provides medical information on Internet	1.31	(4.86)**	2.80	(63.13)***	1.61	(15.80)***	1.57	(13.23)***	1.69	(11.21)***
Negative feelings about looking for health information on the Internet	0.75	(16.82)***	0.81	(10.16)***	0.71	(21.06)***	0.70	(20.83)***	0.77	(6.21)***
Constant	0.08	(89.93)***	0.13	(65.14)***	0.03	(122.43)***	0.04	(103.03)***	0.03	(66.25)***
Nagelkerke R-Square		0.27		0.22		0.22		0.26		0.13

*Notes: ** significant at the 0.05 level and *** significant at the 0.01 level.*

ables. A "1" was recorded for each of the five dependent variables if (1) the online health seeker said "somewhat" or "a lot" of information on the Internet helped them to take care of their health; (2) they had a conversation with family or friend about online health information; (3) online health information changed their behavior; (4) they made a decision about how to treat an illness because of online health information; and (5) they visited a doctor because of information found online. A "0" was recorded for each of the five dependent variables if this was not the case.

The results in Table 5 indicate that for four of the five dependent variables, boomers were slightly more likely to use the Internet to manage their health. For instance, an odds ratio of 1.31 for the dependent variable of "somewhat" or "a lot" of information on the Internet has helped to take care of the online health seekers' problems implies that consumers are around one and one-third times more likely to say that this is the case, if they are a boomer rather than a senior.

Changing behavior for boomers because of information found online registered an odds ratio of 1.71, having a conversation with family or friend about online health information had an odds ratio of 1.49, and making a decision about how to treat an illness had an odds ratio of 1.66 (Table 5). Overall, the results showed that boomers are around one and one-half times more likely than seniors to change their behavior and to use online health information to manage their health, which is not that high, given the attention placed on the differences between these age groups (Kaiser Family Foundation, 2005).

In terms of online proficiency, those who had broadband Internet access were less likely to have a conversation with a family member or a friend about what they saw online and less likely to make a decision about how to treat an illness. However, for those online health seekers who are online every day, this had a consistent impact across all five dependent variables, if they used online health information to manage their health. For instance, there was a 1.82 times greater chance that daily online users changed their behavior because of online health information. In addition, individuals who go online every day were 1.41 times more likely to visit a doctor because of information that they found online. In terms of being online more than 10 hours a week, this had a negative likelihood for the dependent variables "somewhat" or "a lot" of information on the Internet helped to take care of their health, but had a positive impact with an odds ratio of 1.46 for making a decision on how to treat an illness. Overall, there was no overwhelming support that being more proficient with the Internet had a substantial impact on using health information to manage an online health seeker's health. The only consistently significant variable that had an impact on managing health was being online every day.

Boomers and seniors who had excellent or very good health were 1.64 times more likely to have a conversation with a family member or a friend about online health information. In addition, the boomer or senior who had more health problems or who knew someone who was experiencing health problems was 1.14 times more likely to have a conversation with family or friends about online health information. Individuals who were in excellent and very good health were 1.36 times more likely to use online health information to make a decision about how to treat an illness because of online health information. Generally, the health of the individual was not a strong predictor of using health information to manage a boomers' or seniors' health.

The strongest predictors of using health information to manage health were for the online health information awareness and feelings variables. Frequent consumers of online health information were 3.51 times more likely to use the Internet to take care of their health. It has become a valuable tool for their health care management needs. There was also a 3.11 greater likelihood of someone making a decision about treating an illness to use the Internet more than once a month for health information. In fact, frequently accessing health information registered an impact for all of the dependent variables.

Individuals who had positive feelings about online health information were more likely to use online health information to manage their health. Boomers and seniors who had more negative feelings toward online health information were less likely to use it for health management. This finding was consistently found across all five of the dependent variables. Online health seekers who usually looked to see who provided the health information were more likely to use this to manage their health. For instance, individuals who looked to see who provided the health information were 2.80 times more likely to have a conversation with family or friends about the information that they saw online. If a doctor recommended a health or medical Web site, online health seekers were 2.13 times more likely to make a decision about how to treat an illness because of online health information, and they were 3.68 times more likely to visit a doctor because of information they found online. In addition, individuals who communicated with a doctor or a health care provider via e-mail were 1.95 times more likely to have a conversation with a family member or a friend about health information that they found online. Overall, the logistic regression results indicate the most consistent and highest support for increased awareness and positive feelings toward online health information as a driver for helping to manage boomers' and seniors' health.

CLASSIFICATION TREES ANALYSIS OF TAKING CARE OF BOOMERS' AND SENIORS' HEALTH

Another way to examine the relationship between boomers and seniors and online health information is a classification tree analysis. Classification trees are used to predict membership of cases or objects in the classes of a categorical dependent variable from their measurements on one or more predictor variables. Classification tree analysis is a common data mining technique. Figure 2 shows that the taking care of health variable is related to boomers and seniors having positive feelings toward looking for online health information. In addition, having positive feelings about online health information is related to accessing health information online once or twice a month or more. These findings reinforce the logistic regression results that awareness and feeling toward online health information helps in the management of a boomer's or a senior's health. How do these findings relate to the hypotheses outlined in the beginning of this article?

DISCUSSION OF RESULTS AND HYPOTHESES

This section will discuss how the empirical results of this study confirm or deny the hypotheses (see Table 6). First, the evidence shows through the difference of means tests that boomers and seniors are different in terms of their use of online health information in the management of their health. The mean values scored higher for boomers in using online health information to take care of their health, having a conversation with a family or a friend about the health information found online, changing their behavior because of online health information, and making a decision about treating an illness because of online health information. There is some evidence that boomers will use more online health information to manage their health, supporting Hypothesis 1. However, there is no overwhelming support for differences between boomers and seniors and using health information to manage their health, with boomers only utilizing this information one and one-half times more than seniors.

The health of the individual, or Hypotheses 2, only predicted the use of online health information to manage a boomer's and a senior's health when

Figure 2. Classification tree of impact of online health information taking care of a boomer's and a senior's health

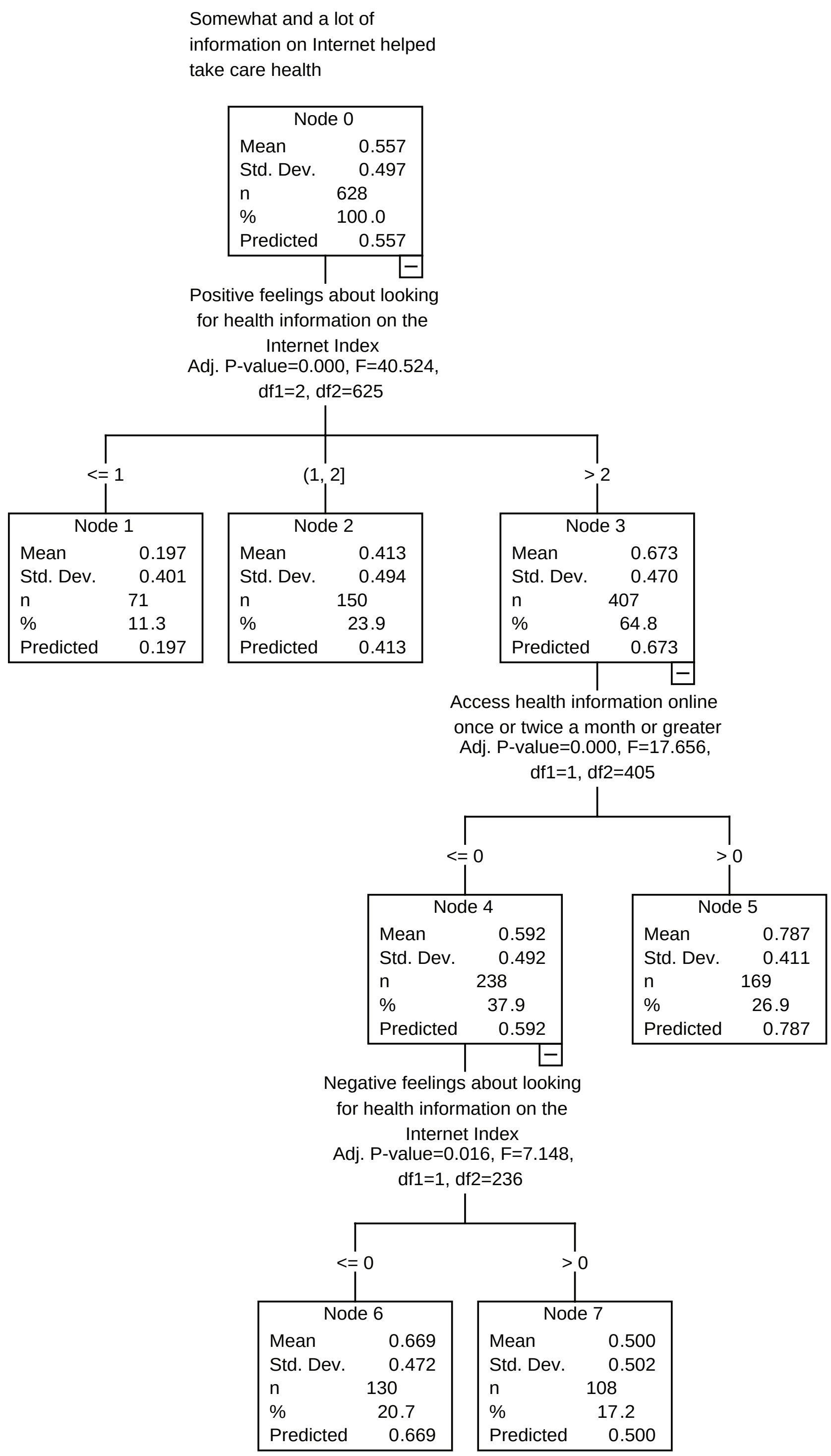


Table 6. Support for hypotheses of boomers and seniors and online health information

Hypotheses	Supported? (Yes, No, or Partially)	Major Test(s) Performed
Hypothesis 1: Online health seekers who are baby boomers are more likely to believe that online health information has helped them manage their health better compared with seniors.	Yes	Descriptive statistics, difference of means tests, and logistic regression
Hypothesis 2: Online health seekers who are healthy or who have family and friends that are healthy will rely less on online health information because of lack of need.	Partially	Logistic regression
Hypothesis 3: Online health seekers who have broadband Internet access will go online more for health information to manage their health	No	Logistic regression - Evidence found in the opposite direction
Hypothesis 4: Online health seekers who go online more often and conduct many online activities will use Internet health information more to manage their health.	Yes	Logistic regression
Hypothesis 5: Boomers and seniors who are females will rely more on online health information.	Yes	Logistic regression
Hypothesis 6: Boomers and seniors who are college educated will rely more on online health information.	Yes	Logistic regression
Hypothesis 7: Boomers and seniors who are Hispanics will go online less for health information.	No	Logistic regression -No support found in logistic regression
Hypothesis 8: Boomers and seniors who have family income above $75,000 will go online more for health information.	Yes	Logistic regression
Hypothesis 9: Online health seekers who most of the time and always look to see who provides medical information on the Internet would use online health information more to manage their health.	Yes	Logistic regression
Hypothesis 10: Online health seekers who access health information online once or twice a month would have a greater likelihood of using online health information to manage their health.	Yes	Logistic regression
Hypothesis 11: If a doctor has recommended a Website to an online health seeker he or she is more likely to use health information to manage their health.	Partially	Logistic regression
Hypothesis 12: If an online health seeker has communicated with his or her doctor via email he or she is more likely to use online health information to manage their health.	Yes	Logistic regression
Hypothesis 13: Online health seekers that have more positive feelings about looking for health information on the Internet are more likely to use this information to manage their health.	Yes	Logistic regression

they had a conversation with a family member or a friend about online health information and when they made a decision about how to treat an illness. In addition, the health variable predicted a boomer's and a senior's behavior when he or she talked to his or her doctor about information that he or she found online. Overall, there was not overwhelming support that the health of the boomer or senior had an impact on use of online health information to manage his or her health.

Hypotheses 3 and 4 examine whether being more proficient online means that the online health seeker will use the Internet more to manage his or her health. The results consistently showed that those who go online every day were more likely to use the Internet to manage their health. There was not much support that being more proficient with the Internet meant that boomers and seniors would use it more often for health information, since many of the other independent variables in this category were not statistically significant. This is most likely explained by what these online health seekers are doing on the Internet; they are looking for information, which does not require, for instance, broadband Internet access, since a standard dial-up connection will suffice. Therefore, this research cannot confirm that being more proficient with the Internet means that online health seekers will use this communication media to manage their health more than those who are not as proficient. However, this could change with greater availability of streaming video health information, which is much more suited to a broadband Internet connection.

Boomers and seniors of higher sociodemographic status use the Internet for health information much more than lower sociodemographic status individuals (Hypotheses 5-8). Therefore, this research confirms that there is a digital divide in access to online health information, and public policy should attempt to address this issue. It should be noted that one-third of the United States adult population has not gone online and, therefore, would not be able to benefit from online health information (Pew Internet & American Life, 2005).

Hypothesis 10 was confirmed in the logistic regression results that those who are frequent patrons of online health information actually will use it more often to manage their health. The results showed that health seekers who are accessing health information online once or twice a month or more will be more likely to actually use this information to manage their health. Therefore, these individuals are not just searching for information; they actually are using some of what they find online. Individuals who have positive feelings about online health information also will use it more often to manage their health, and individuals that harbor more negative feelings will use it less often (Hypothesis 13). Online health seekers who are very aware of who provides the medical information on the Internet are more likely to use this information to manage their health (Hypothesis 9). Finally, where there is communication with their doctor via e-mail or at the doctor's office about online health information, boomers and seniors are more likely to use online health information to manage their health (Hypothesis 12). Awareness and feelings toward online health information generally were well-supported predictor variables (Hypotheses 9-13), having an impact on a boomer's or a senior's use of this information to manage his or her health care needs.

RECOMMENDATIONS, LIMITATIONS, AND FUTURE RESEARCH

This article examined the use of the Internet for accessing health information by boomers (age 50 to 64) and seniors (age 65 and over). Boomers generally use Internet health information to manage their health more than seniors. However, there was no overwhelming differences between boomers and seniors, which is the main difference in the finding from another study (Kaiser Family Foundation, 2005). For instance, boomers are much more likely to talk to a doctor about health information that they saw online. Boomers are around one and one-half times more likely than seniors to use online health information to manage their health. This study found that awareness and feelings toward online health information provided the best explanation of health information for management of boomers' and seniors' health.

Since boomers were found to use online health information marginally more than seniors, what are the implications of this observation? Will seniors of tomorrow be similar to seniors of today? Perhaps boomers will continue to seek online health information as they get older. The implication that boomers and seniors may be in the greatest need of health information may not be true in the future with the growing obesity epidemic in the United States, which affects all age groups.

Some policy recommendations should be noted in order to bring more seniors online and to enhance the quality of Internet health resources. Health care professionals should recommend Web sites, promote more effective search and evaluation techniques, and be more involved in developing and promoting uniform standards for health Web sites (Morahan-Martin, 2004). Since only a minority of seniors has ever gone online, this represents a significant digital divide. These findings confirm that for the foreseeable future, the Internet is less likely to be a primary source of information for most seniors, which suggests a need to invest more heavily in education and outreach strategies. This is especially the case for seniors with low or modest incomes, who are least likely to go online for this information. These recommendations could make seniors more aware and could create a positive experience when going online for health information.

In the near future, the Internet will become a decision-making tool for seniors, who will need to make choices about the Medicare prescription drug benefits. They will need to decide which plan has the most attractive premium and to determine whether it will cover the medications they take and will work with the pharmacy they use. Seniors also will need to manage the Internet to make these important decisions. Web site design is part of the solution, since seniors have problems scrolling on Web sites and remembering Web pages (Voelker, 2005).

There are some limitations of this research. With any type of public opinion data, especially when asking subjective questions about sensitive topics of consumers' health, respondents may not be as forthcoming with information. Another limitation is that of the general applicability of the results, given that the proportion of the sample is different for seniors and boomers. In addition, there is no question that specifically addressed whether there was an improvement in the health outcome, just that people felt better informed. Future research could do a longitudinal follow up of this dataset, which might reveal shifts in the use of Internet health information for managing health with boomers and seniors, looking at other measures to see if there is an impact on change in the person's health.

REFERENCES

Anderson, J. G. (2004). Consumers of e-health: patterns of use and barriers. *Social Science Computer Review, 22*(2), 242-248.

Brodie, M., Flournoy, R. E., Altman, D. E., Blendon, R. J., Benson, J. M., & Rosenbaum, M. D. (2000). Health information, the Internet, and the digital divide. *Health Affairs, 19*(6), 255-265.

Cotton, S. R., & Gupta, S. S. (2004). Characteristics of online and offline health information seekers and factors that discriminate between them. *Social Science & Medicine, 59*(9), 1795-1806.

Eysenbach, G. (2003). The impact of the Internet on cancer outcomes. *CA A Cancer Journal for Clinicians, 53*(6), 356-371.

Eysenbach, G., & Jadad, A. R. (2001). Evidence-based patient choice and consumer health informatics in the Internet age. *Journal of Medical Internet Research, 3*(2), e35.

Fox, S., & Rainie, L. (2000). *The online health care revolution: How the Web helps Americans take better care of themselves.* Washington, DC: Pew Internet & American Life.

Fox, S., & Rainie, L. (2002). *Vital decisions: How Internet users decide what information to trust when they or their loved ones are sick*. Washington, DC: Pew Internet & American Life.

Gray, N. J., Klein, J. D., Noyce, P. R., Sesselberg, T. S., & Cantrill, J. A. (2005). Health information-seeking behavior in adolescence: The place of the Internet. *Social Science & Medicine, 60*(7), 1467-1478.

Huntington, P., Nicholas, D., Homewood, J., Polydoratou, P., Gunter, B., Russell, C., et al. (2004). The general public's use of (and attitudes towards) interactive, personal digital health information and advisory services. *Journal of Documentation, 60*(3), 245-265.

Kaiser Family Foundation. (2005). *E-health and the elderly: How seniors use the Internet for health information*. Washington, DC: Kaiser Family Foundation.

Lueg, J. E., Moore, R. S., & Warkentin, M. (2003). Patient health information search: An explanatory model of Web-based search behavior. *Journal of End User Computing, 15*(4), 49-61.

Morahah-Martin, J. M. (2004). How Internet users find, evaluate, and use online health information: A cross-cultural review. *CyberPsychology & Behavior, 7*(5), 497-510.

Nicholas, D., Huntington, P., Williams, P., & Blackburn, P. (2001). Digital health information provision and health outcomes. *Journal of Information Science, 27*(4), 265-276.

Pandey, S. K., Hart, J. J., & Tiwary, S. (2003). Women's health and the Internet: Understanding emerging trends and implications. *Social Science & Medicine, 56*(1), 179-191.

Pew Internet & American Life. (2005). *January 2005 Internet tracking survey*. Retrieved May 10, 2005, from http://www.pewinternet.org

Rideout, V. (2002). Generation Rx.com: What are young people really doing online? *Marketing Health Services, 22*(1), 27-30.

Skinner, H., Biscope, S., & Poland, B. (2003). Quality of Internet access: Barrier behind Internet use statistics. *Social Science & Medicine, 57*(5), 875-880.

Voelker, R. (2005). Seniors seeking health information need help crossing the "digital divide." *JAMA, 293*(11), 1310-1312.

Ziebland, S. (2004). The importance of being expert: The quest for cancer information on the Internet. *Social Science & Medicine, 59*(9), 1783-1793.

ENDNOTES

[1] In this study, for simplicity, baby boomers are classified as those individuals between the ages of 50 and 64, and seniors are classified as 65 and older.

[2] This author would like to thank Victoria Rideout, M.A., Vice President and Director, Program for the Study of Entertainment Media and Health, Kaiser Family Foundation (KFF), for the dataset and documentation used in the statistical analysis of this study. I would also like to thank Virginia Rodgers for her editorial assistance.

[3] For the data analysis, the software package used was SPSS version 13.0.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 1, Issue 2, edited by J. Tan , pp. 20-38, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 5.14
The Role of Internet Self-Efficacy in the Acceptance of Web-Based Electronic Medical Records

Qingxiong Ma
Central Missouri State University, USA

Liping Liu
University of Akron, USA

ABSTRACT

The technology acceptance model (TAM) stipulates that both perceived ease of use (PEOU) and perceived usefulness (PU) directly influence the end user's behavioral intention (BI) to accept a technology. Studies have found that self-efficacy is an important determinant of PEOU. However, there has been no research examining the relationship between self-efficacy and BI. The studies on the effect of self-efficacy on PU are also rare, and findings are inconsistent. In this study, we incorporate Internet self-efficacy (ISE) into the TAM as an antecedent to PU, PEOU, and BI. We conducted a controlled experiment involving a Web-based medical record system and 86 health care participants. We analyzed both direct and indirect effects of ISE on PEOU, PU, and BI using hierarchical regressions. We found that ISE explained 48% of the variation in PEOU. We also found that ISE and PEOU together explained 50% of the variation in PU, and the full model explained 80% of the variance in BI.

INTRODUCTION

Application service provision (ASP) — a model of distributing software services over the Internet — has shown its advantages over the traditional model of information technology (IT) deployment. The expected benefits include the reduced cost of technology ownership, the reduced time to market, and the reduced risks with software deployment.

Nevertheless, the growth of the ASP business has been comparatively slow. In response to the situation, many researchers (Jayatilaka, Schwarz, & Hirschheim, 2002; Peterson & Fairchild, 2003; Susarla, Barua, & Whinston, 2003) examined its ensuring factors. Along the same line of inquires, this study attempts to understand the acceptance issue from the end-user perspective and searches for guidance on methods and effective interventions to promote the adoption of the ASP model.

Understanding user acceptance behavior is important for several reasons. First, it is the end users who use the technology in their work on a daily basis. Any decision that changes their work behavior should consider their willingness to adopt the change. Empirical evidence has shown that the technology adoptions involving end users were more successful than those without (Chau & Hu, 2002; Lederer, Maupin, Sena, & Zhuang, 2000). Second, only the end-user acceptance can ensure a potential long-term continuous adoption (Bhattacherjee, 2001). This is particularly crucial to the ASP adoption since most ASPs are operated on short-term renewable contracts.

In the technology adoption literature, the technology acceptance model (TAM) by Davis (1989) is one of the most widely applied models (see Ma & Liu, 2004, for a meta-analytical survey). It has received extensive empirical support through validations, applications, and replications. Compared with competing models, the TAM is believed to be more parsimonious, predicative, and robust (Venkatesh, 2000). However, the TAM has been criticized for being less informative in understanding usage behavior (Taylor & Todd, 1995). Accordingly, researchers have attempted to extend the TAM by embedding it into a nomological network of other antecedents and consequences. To this end, a few researchers appeal to cognition theories and emphasize the importance of self-efficacy.

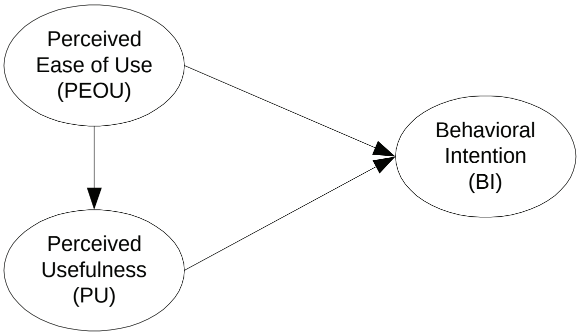
The notion of self-efficacy refers to beliefs about individuals' capabilities of performing a certain task (Bandura, 1977). Numerous studies have found that self-efficacy is an important determinant of perceived ease of use. For example, Venkatesh (2000) found that computer self-efficacy was one of the main factors that affect ease of use. Agarwal, Sambamurthy, and Stair (2000) defined software-specific self-efficacy and had a similar finding. However, currently there is no research, to the best of our knowledge, testing the relationship between self-efficacy and behavioral intention (to use an information technology). The impact of self-efficacy on perceived usefulness is also less known, and findings are inconsistent. The goal of this paper is to fill in this gap. In particular, we extend the TAM by considering Internet self-efficacy (ISE) as an antecedent to perceived ease of use, usefulness, as well as behavioral intention in the context of accepting Web-based electronic medical records, and investigate how such an extension affects our understanding about end-user acceptance behavior.

The rest of this paper is outlined as follows. In the next section, we review the literature related to the TAM. In "Research Hypotheses," we present our research model and develop research hypotheses. In the section after that, we describe our research design, including research methodology, experiment procedure, and instrument validation. In "Data Analysis," we perform data analysis and show test results. Finally, we discuss the results and draw conclusions.

TECHNOLOGY ACCEPTANCE MODEL

There are numerous perspectives from which one studies user acceptance and usage behavior of information technologies. Among them, the TAM by Davis (1989) is a more popular one. The TAM is grounded in the theory of reasoned action (Fisherbein & Ajzen, 1975). It stipulates a nomological network of three constructs (Figure 1) — perceived usefulness (PU), perceived ease

Figure 1. Technology acceptance model



of use (PEOU), and behavioral intention (BI) — connected by the causal links that both PEOU and PU directly influence BI, and PEOU influences BI indirectly through PU.

The TAM has been applied to a wide range of technologies, including e-mail, fax, word processors, spreadsheets, and workgroup applications. It has been recently applied to the adoption of e-commerce and Internet technologies (Gefen & Straub, 1997, 2000; Lederer et al., 2000). According to Ma and Liu (2004), there have been over 100 studies applying or validating the TAM. Most of these studies confirmed the reliability and validity of PU and PEOU in predicting BI or technology usage, although conflicting evidence exists.

Compared with other competing models, the TAM is believed to be more accurate and parsimonious when it is used to predict adoption. However, its parsimony results in being less informative in understanding usage behavior (Taylor & Todd, 1995) or providing usable results. What IT managers desire to have are implementable prescriptions (Benbasat & Zmud, 1999). It is imperative that researchers equip them with some useful guidance on methods and effective interventions, for example, by augmenting PU and PEOU through manipulating their causal antecedents to achieve greater technology acceptance or usage. As Gefen and Keil (1998) noted, without a better understanding of the antecedents to PU and PEOU, managers and developers are unable to know which levers to pull in order to affect these beliefs and, through them, technology usage.

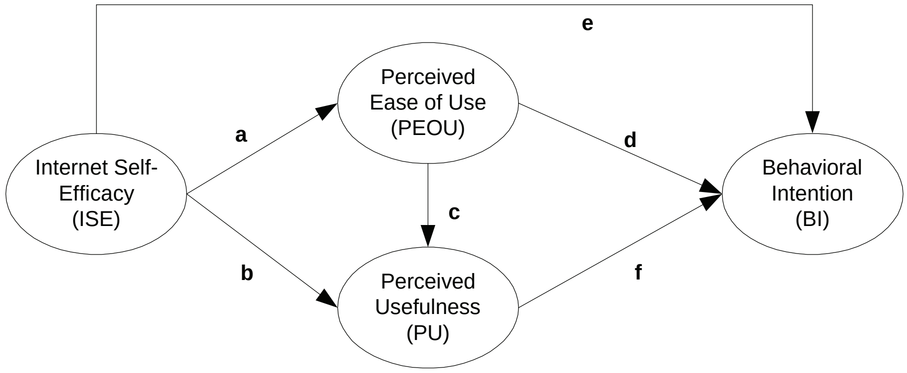
Recognizing the necessity of identifying the antecedents to user acceptance, researchers have attempted to extend the TAM by encompassing various constructs such as gender, culture, trust, experience, and social influence (Chircu, Davis, & Kauffman, 2000; Gefen & Straub, 1997; Straub, 1994). Among the constructs, self-efficacy is recognized to be a more important one.

Self-efficacy has been documented to be an important determinant of PEOU (Agarwal et al., 2000; Venkatesh, 2000; Venkatesh & Davis, 1996). However, there is no research examining the relationship between self-efficacy and BI. Also, very few studies have been done to understand the impact of self-efficacy on PU, and their findings are inconsistent. In a study with a sample of 58 participants from the California Network Engineering Center, Lopez and Manson (1997) found that computer self-efficacy was positively related to the PU of the Empowered Desktop Information System. Using 288 junior business students, Agarwal et al. (2000) conducted a similar test within the context of Web technologies. They found a positive effect of self-efficacy on both PU and PEOU, and a stronger effect on PEOU than on PU. In contrast, Igbaria, Iivari, and Maragahh (1995) found that self-efficacy had an insignificant direct effect on PU. Chau (2000) also found that computer self-efficacy had a relatively small, but negative, impact on PU. Therefore, to understand the exact effect of self-efficacy on PU, additional evidence seems necessary.

RESEARCH HYPOTHESES

The aim of this study is to examine the role of ISE in accepting Web-based electronic medical records. In addition, we want to validate existing findings in the new context. To this end, we extend the TAM by including ISE as an antecedent to the three theoretical constructs: PEOU, PU, as well

Figure 2. Research model



as BI. We hypothesize that ISE directly influences PEOU and PU, which in turn influence BI (Figure 2). We also hypothesize that ISE directly influences BI. We justify these hypotheses based on existing theories and findings.

The notion of self-efficacy is due to Bandura (1977). It refers to beliefs about individuals' capabilities to produce designated levels of performance that exercise influence over events that affect their lives. Since its inception, the construct has been widely tested in organizational behavior, education, and human resource management.

In the context of end-user computing, Compeau and Higgins (1995) proposed the concept of computer self-efficacy to refer to the judgment of one's capability of using the computer. Marakas, Yi, and Johnson (1998) made a further distinction between computer self-efficacy and task-specific self-efficacy. Following the same distinction, in this study we define ISE as the estimate of one's capability to perform Internet tasks.

ISE is different from general computer self-efficacy in that ISE focuses on what an individual believes he or she can accomplish online now or in the future — the belief that one can establish, maintain, and utilize the Internet effectively over and above basic personal computer (Eastin & LaRose, 2000). It is the judgment of one's capability of applying Internet skills rather than the measure of his or her specific skills in using an Internet browser. People who have little confidence in their ability to use the Internet or who are uncomfortable using the Internet may be said to have weak self-efficacy beliefs.

Venkatesh (2000) studied factors that affected PEOU and classified them into two categories: anchors and adjustments. Anchors include computer self-efficacy, perception of external control, computer anxiety, and computer playfulness. Adjustments include perceived enjoyment and objective usability. The general computer self-efficacy turns out to be a strong determinant of PEOU before hands-on experience (Venkatesh & Davis, 1996). Based on this evidence, we expect that ISE has a positive effect on the PEOU of Web-based applications.

Hypothesis 1 (path a): Internet self-efficacy has a positive effect on the perceived ease of use of Web-based electronic medical records.

Self-efficacy is significantly related to end-user competence. Munro, Huff, Marcolin, and Compeau (1997) suggested that the construct of user competence is multifaceted, composed of both *breadth* and *depth* of knowledge on end-user technologies, and one's ability to creatively apply these technologies (*finesse*). Thus, it is reasonable to infer that self-efficacy is also a multidimensional construct, with some capability beliefs coming from the breadth of knowledge, some from the depth of knowledge, and some from the finesse of knowledge. Based on this logic, we assume that some aspects of competence or knowledge give a user confidence in using a technology, which contributes to the perception of its ease of use, whereas others give the user confidence in understanding the design logic or objectives of the technology, which contributes to the perception of its usefulness. Therefore, we anticipate that ISE has a direct positive effect on PU.

Hypothesis 2 (path b): Internet self-efficacy has a positive effect on the perceived usefulness of Web-based electronic medical records.

Research has found that self-efficacy is an important determinant of user perception besides PEOU and PU. For example, Levine and Donitsa-Schmidt (1998) found that as individuals expressed stronger computer confidence, they demonstrated more positive attitudes toward computers and higher levels of computer-related knowledge. According to Bandura (2001), self-efficacy is also a major determinant of intention. Although there is no IT research testing the relationship, studies in other areas confirm its significance. For example, Armitage and Conner (2001) conducted a study testing a model based on the theory of planned behavior in blood donation. They found that self-efficacy accounted for unique variance in behavioral intention. In another extended study with a sample of 172 participants, they also found that self-efficacy and several other factors were all important independent predictors of behavioral intention. Based on this body of theoretical and empirical evidence, we propose the following.

Hypothesis 3 (path e): Internet self-efficacy has a positive effect on behavioral intention to use Web-based electronic medical records.

In addition to Hypotheses 1 to 3, we present the original three hypotheses stipulated by the TAM as Hypotheses 4 to 6 as follows.

Hypothesis 4 (path c): Perceived ease of use has a positive impact on the perceived usefulness of Web-based electronic medical records.

Hypothesis 5 (path d): Perceived ease of use has a positive impact on behavioral intention to use Web-based electronic medical records.

Hypothesis 6 (path f): Perceived usefulness has a positive impact on behavioral intention to use Web-based electronic medical records.

RESEARCH DESIGN

We conducted a controlled experiment using senior clinic trainees and staff members who are responsible for managing patient records. We collected data by means of a survey instrument. We distributed a questionnaire and a cover letter with instructions to selected participants. In the cover letter, we informed the participants that the experiment was volunteer-based and they could refuse to participate or quit at any time. In the instruction sheet, we provided a URL to the target system under testing, a test account for using the system, and a link to a quick user guide that details the functionalities of the system.

Target System

According to our research objective, we used a commercial Web-based electronic medical records (EMR) system. We used two criteria in selecting the system. First, the system should be truly Web-based; a participant merely needs a Web browser to load the system and perform record management tasks. Second, the system must be a full-fledged EMR system. It must allow a participant to create, update, delete, and retrieve patient records from anywhere at anytime. It must support electronic communication with labs, pharmacies, hospitals, and other service providers. For example, participants could create and retrieve prescriptions or add and modify medications and pharmacies. The system must support multiple file types including sounds and videos so that a participant can add quick notes to a document, and store and maintain dictation files on the server.

Literally there were 100s of application service providers who claimed to provide EMR applications. Among them, a few delivered applications that truly worked within a Web browser. Many others employ client-server architecture, that is, a user has to download and install a Windows client in order to work with a remote server. After a long process of searching, eliminating, pondering, and many runs of back-and-forth communications with application providers, we decided that HyperCharts™ was the most appropriate for this study.

Subjects

The goal of this study is to understand the end-user behavior in accepting Web-based medical record systems. Thus, the ideal target population should be those health care workers who are responsible for managing patient records using the computer. Since these people are typically supportive staff members in clinics, their contact information is often not accessible to outsiders. We assessed that a mail survey was going to be difficult to find its way to a target audience and the overall response rate would be low. Therefore, in this study, we identified 90 senior health care trainees in a large mid-west university as surrogates. These students were majoring in dental hygiene, physician assistant, and radiology. These four-year programs were designed to prepare the graduates to successfully enter health care organizations such as hospitals and clinics. Their program curricula include intense problem-based learning modules combined with clinical experience such as internships and externships. All the selected participants had training and experience in managing patient records and were familiar with the daily operations in hospitals and clinics.

We also obtained 85 clinic e-mail addresses through a medical worker association. We sent e-mails to the candidates requesting their participation. We provided a link to the HyperCharts™ system in the message. We assumed they agreed to participate if they did not indicate otherwise. In the following week, we did a Web-based survey. Eventually, we received 11 completed responses; two candidates refused to participate and 13 messages were undeliverable.

In total, 86 participants completed the experiment and returned usable responses. Among them, 75 were senior health care trainees and 11 were staff workers in clinics. The overall response rate was 49%. We did t-tests on gender, experience with Web-based medical records, and attitudes toward the Internet, and there was no evidence showing any difference between these two groups. Among all the subjects, almost all of them were female and over 85% of them were at the age of 20 to 25. Of the participants, 13% reported awareness of some EMR systems and only 8% of them had hands-on experience with Web-based medical records. The rest of them had used traditional paper-based systems. Based on the five-point Likert scale from very negative to very positive, 69% of them had reported neutral attitudes to their current paper-based systems. With regard to Internet experience, 84% of them had over three years of using the Internet and 69% of them had positive feelings toward surfing the Internet.

Instrument

All constructs are measured using multiple scale items. The items underlying each construct were carefully developed according to its respective definition. Each item is measured using a five-point Likert scale ranging from *low* to *high*.

Internet self-efficacy. Research on self-efficacy prescribes that self-efficacy assessment should reflect different facets of the task domain, types of capabilities required, and situational circumstances in which those capabilities are exercised

Table 1. Scale reliability

Construct	Number of items	Cronbach's Alpha
Perceived Usefulness (PU)	5	0.94
Perceived ease of Use (PEOU)	5	0.91
Behavior Intention (BI)	5	0.89
Internet Self-efficacy (ISE)	6	0.93

(Bandura, 1997). However, the original computer self-efficacy scales were developed prior to the rise of ubiquitous Internet computing. They do not capture many important skills unique to the Internet (Torkzadeh & Van Dyke, 2002). Existing studies on ISE (Nahl, 1996; Nahl & Meer, 1997) focused on the operations of specific tasks such as entering a Web address (URL), creating folders, adding and removing bookmarks, mailing Web pages, and using file transfer protocol. Accordingly, Joo, Bong, and Choi (2000) measured ISE based on specific tasks in Internet search such as navigation, printing, and closing the browser. On the other hand, Eastin and LaRose (2000) proposed an instrument based on overall performance of Internet users. In the current study, we took a balanced stand in between these two positions and adapted most of the items from the studies (Dinev & Koufteros, 2002; Eastin & LaRose; Joo et al.; Torkzadeh & Van Dyke). We created an additional item for an important skill, which was not addressed in previous studies. Finally we had eight items based on the skills that are required to run Web-based applications (see Appendix A).

Other Constructs. Items for measuring PEOU and PU were originally proposed by Davis (1989) and successively tested and validated in many other studies (Chau, 1996; Gefen & Straub, 1997; Hendrickson, Massey, & Cronan, 1993; Mathieson, 1991; Ridings & Gefen, 2000). We tailored the items so that they are applicable to Web-based technologies. The items for measuring BI were adapted from the studies of Chau (1996) and Venkatesh and Davis (1996).

Task

Due to time limitation, we did not require the participants to test all functionalities available in HyperCharts™. Instead, we focused on the primary EMR functions as suggested by Waegemann (2002) and required participants to accomplish the following tasks: (a) Log into the system with the provided user name and password; (b) create, update, and delete a patient record, search for a patient record, and add quick notes to a patient record; (c) search for prescriptions for a patient, add and sign off a new prescription, and add and modify medications on a prescription; and (d) upload and download a dictation audio file. These primary functions were also described in the introduction to the system on the HyperCharts™ Web site. All test data are similar to what the subjects are dealing with in their everyday work.

DATA ANALYSIS

Analytical Techniques

When deciding on analytical techniques, we considered hierarchical multiple regressions versus structural equation modeling (SEM) techniques like LISREL. SEM has its advantages in assessing both the efficacy of a measurement model and the significance of a structural model in one test, and is becoming popular in IT research in recent years. However, one is advised that SEM has some limitations compared with other more traditional techniques such as multiple regressions. First, SEM has a stringent requirement on sample sizes. Ding, Velicer, and Harlow (1995) noted that 100 to 150 participants is the minimum sample size for SEM while Hu, Bentler, and Kano (1992) indicated that sometimes 5,000 participants is insufficient. Second, the goodness-of-fit indices for SEM have no single statistical test of significance that identifies a correct model (Schumacker & Lomax, 1996). The widely cited criteria often falsely accept bad models (Liu, Li, & Karau, 2003). The problem compounds when the sample is too small or too large.

Therefore, in this study we choose hierarchical multiple regressions as the primary technique for data analysis. The use of hierarchical regression allows us to test recursive models, where a predictor in one model may become the dependent variable in another, and therefore an antecedent variable may have impacts on a consequent variable in multiple distinct ways, with some direct and some indirect. Our research model is a case in point; each hypothesis stipulates the total effect of a predictor on a dependent variable, including both direct and indirect effects. For example, according to Figure 2, ISE can affect behavioral intention in four distinct ways, and the total effect of ISE on BI is the sum of all these direct and indirect effects.

1. ISE → BI (path e)
2. ISE → PU → BI (path b + path f)
3. ISE → PEOU → BI (path a + path d)
4. ISE → PEOU → PU → BI (path a + path c + path f)

Following the guideline proposed by Hair, Anderson, Tatham, and Black (1998), we assessed the assumptions of hierarchical regressions: the linearity of a phenomenon, the constant variance of the error term, and the normality of the error term distribution. We chose studentized residuals as the measure of the prediction error of each dependent variable. We examined residual scatter plots and partial regression plots. We did not detect any violations of these assumptions.

Scale Assessment

Before the survey was conducted, three PhD students and two faculty members reviewed the survey questions regarding their clarity and face and content validity. Two scale items for ISE were dropped because they were not directly related to Web-based medical record systems. As a result, six items are used to measure the construct of ISE while the scales for other constructs remain unchanged (Appendix B).

Since all research constructs are latent, each one is reflected by a group of measurable items and indirectly measured as the average of their scores. Therefore, before we conduct data analysis on the constructs, we need to ensure the reliabil-

Table 2. Factor loadings

	Factor 1	Factor 2	Factor 3
PEOU2	.835		
PEOU1	.813		
PEOU3	.785		
PEOU4	.760		
PEOU5	.627		
PU4		.851	
PU2		.847	
PU3		.840	
PU5		.756	
PU1		.734	
ISE6			.858
ISE1			.802
ISE5			.764
ISE3			.728
ISE2			.697
ISE4			.693
Eigenvalues	9.478	1.540	1.339
% Variance	59.24	68.91	77.28

Note: Factor loadings lower than .50 are not shown.

ity of their scales, which are usually measured by Cronbach's alpha coefficients. The reliability coefficient of each construct is reported in Table 1. In general, a coefficient value of 0.70 or higher is considered acceptable (Hair et al., 1998). As per this standard, the reliabilities of our constructs are very high and comparable to or even better than those in other similar studies (Appendix A).

To ensure that the items for the same construct measure a single trait while items for different constructs measure distinct traits, we conducted a principal factor analysis with varimax rotation on 16 items for all independent constructs: ISE, PEOU, and PU. Using the Kaiser eigenvalues criterion, we extracted three factors that collectively explained 77.28% of the variance in all items. Statistically, to obtain a power level of 80% at significance level 0.05, the significant value of factor loadings with a sample size of 100 is 0.50 (Hair et al., 1998). The rotated factor matrix in Table 2 shows that all the items cleanly load on the correct latent constructs, supporting the factorial validity of the measurement instrument.

Results

In hierarchical multiple regressions, the researcher decides on not only how many predictors to enter, but also the order in which they enter. Usually, the order of entry is based on logical or theoretical considerations. We enter the most exogenous predictors first and the most endogenous predictor last. In our case, because ISE is modeled as having a direct influence on PEOU, it is important to statistically control the direct influence of ISE on BI before evaluating the independent contribution of PEOU to BI. Failing to control the direct influence of ISE could result in the relationship between PEOU and BI being artificially inflated due to the indirect influences through the perceptual variable. In the same vein, it is important to control the direct influence of PEOU on BI before evaluating the independent contribution of PU. Thus, ISE was entered into the regression model first, followed by PEOU and PU in the second and third step. In this way, we are able to tease out the influence of ISE before considering that of PEOU and PU. The regression results are presented in Table 3. We did similar regressions when PU and PEOU are dependent variables, and the results are shown in Table 4 and 5, respectively.

R^2 value — the coefficient of determination — represents the percentage of the variance that can be explained by the predictors in a relationship. According to the guideline provided by Hair et al. (1998), the threshold of R^2, for example, with a power of .80 at significance level 0.01, is about 15% for a sample of 100 participants with four independent variables. In other words, if the R^2 value for a relationship is greater than 15%, we will be confident that the effect of the independent variables on the dependent variable is statistically significant. According to this standard, all R^2 values are statistically significant.

An R^2 change is probably more important than an R^2 value itself. A significant R^2 change, say at level $\alpha \leq 0.01$, means that an additional predictor can explain a significant amount of the variance in the dependent variable and hence add additional explanation power to the model. For example, according to Table 3, about 80% of the total variance in BI is explained by three predictors: ISE, PEOU, and PU. Besides, all R^2 changes are significant, meaning all three predictors exhibited significant positive influences on BI. The three predictors explained respectively about 55% (ISE), 15% (PEOU), and 10% (PU) of the variance in BI. Similarly, according to Table 4, about 50% of the variance in PU was explained by ISE (45%) and PEOU (5%).

In order to test our research hypotheses, we need to calculate the coefficient of each path, called direct effect, as well as the total effect of each predictor. A direct effect represents the change in the dependent variable directly attributable to a standard deviation change in a predictor. A total effect represents the total change in the dependent variable attributable to the direct

Table 3. Hierarchical regressions for dependent variable BI

Model	Coefficients			Model Summary		
	Beta	T	Sig.	R^2	R^2 Change	Sig.
1				.548	.548	.000
ISE	.741	10.100	.000			
2				.704	.156	.000
ISE	.363	4.391	.000			
PEOU	.546	6.613	.000			
3				.801	.097	.000
ISE	.162	2.157	.034			
PEOU	.408	5.699	.000			
PU	.441	6.306	.000			

Table 4. Hierarchical regressions for dependent variable PU

Model	Coefficients			Model Summary		
	Beta	T	Sig.	R^2	R^2 Change	Sig.
1				.451	.451	.000
ISE	.672	8.310	.000			
2				.502	.051	.004
ISE	.455	4.246	.000			
PEOU	.313	2.925	.004			

Table 5. Hierarchical regressions for dependent variable PEOU

Model	Coefficients			Model Summary		
	Beta	T	Sig.	R^2	R^2 Change	Sig.
1				.478	.478	.000
ISE	.691	8.770	.000			

effect of the predictor, as well as its effects that are mediated through other predictors. Thus, we need to accumulate both direct and indirect effects to compute a total effect. Direct effects are the standardized beta coefficients in the full model. For example, according to Table 3, the beta coefficients in the full model are respectively 0.16 for ISE, 0.41 for PEOU, and 0.44 for PU. Thus, the direct effect of ISE on BI is 0.16. Similarly, the direct effect of PEOU on BI is 0.41 and that of PU is 0.44. Figure 3 shows all direct effects taken from Tables 3 to 5.

To compute the total effect, we consider all the paths that link a predictor to the dependent variable with or without mediators in between. If a path has no mediator, it carries a direct effect. If a path consists of one or more mediators, it carries an indirect effect, which is the product of all direct effects on the path. For example, the indirect effect of ISE on BI through the path ISE $\rightarrow$ PEOU $\rightarrow$ PU $\rightarrow$ BI is $0.69 \times 0.31 \times 0.44 = 0.09$. The total effect is simply the sum of the direct and indirect effects carried by all the paths. In Table 6, we summarized the total effect for each pair of a

Figure 3. Direct effects of predictors

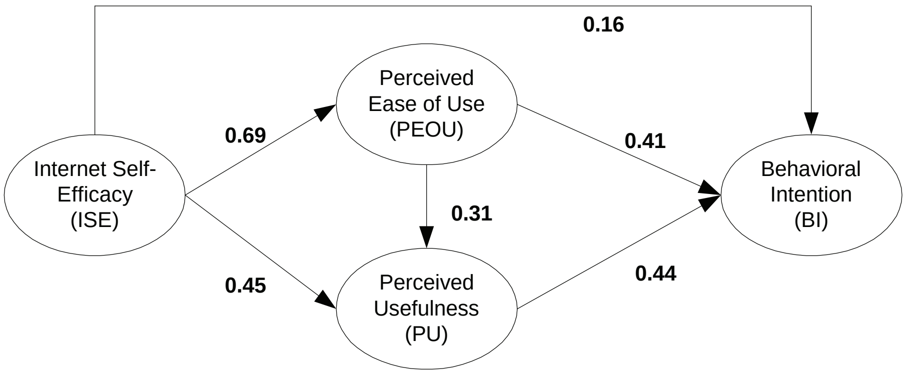


Table 6. Total effects summary

Hypothesis	Relationship	Total Effect	Supported?
H1	ISE → BI	.741	Yes (p<.001)
H2	ISE → PEOU	.691	Yes (p<.001)
H3	PEOU → PU	.313	Yes (p=.004)
H4	PU → BI	.441	Yes (p<.001)
H5	PEOU → BI	.546	Yes (p<.001)

predictor and a dependent variable, and the level of support for each research hypothesis. As shown, all hypotheses are strongly supported by the data at the significance level 0.005 or higher.

DISCUSSION AND CONCLUSION

The results of this study indicated that Internet self-efficacy has a significant impact on perceived ease of use, perceived usefulness, as well as behavioral intention to use Web-based electronic medical records. The total effect of Internet self-efficacy on perceived usefulness is slightly weaker than that on perceived ease of use. It indicates that users perceived the system to be both easier to use and more useful when their Internet self-efficacies are higher. This finding is consistent with that of previous studies (Agarwal et al., 2000; Lopez & Manson, 1997). By synthesizing this and other self-efficacy studies (Agarwal et al.; Igbaria et al., 1995; Lopez & Manson), we found that the impact of self-efficacy on usefulness is significant in Internet- or Web-related settings, but not significant in traditional non-Web-based contexts. Of course, this finding needs further investigation before jumping to a firm conclusion.

Perceived ease of use influences both usefulness and intention. However, its total effect on behavioral intention is stronger than that on usefulness. This finding is interesting because the role of ease of use has been found unstable and controversial (Ma & Liu, 2004). Most studies indicated that the impact of ease of use on usefulness is stronger than that of ease of use on behavioral intention, although few other researchers found a much larger effect of ease of use on intention than usefulness (Lim, 2001). In a study with the Internet technology, Gefen and Straub (2000) found that PEOU influences BI when a Web site involves inquiries, but does not influence BI when a Web site is used for a purchasing task.

Implications

Theoretically, this study helps understand the antecedents to the key constructs in the technology acceptance model. It confirms the existing findings on the role of self-efficacy in perceiving the usability of an information technology. It adds additional evidence to the conflicting findings on the role of self-efficacy in determining perceived usefulness. Most importantly, it does all these in the context of accepting Web-based electronic medical records, which has not been done yet. Since the software industry is currently undergoing a revolutionary transition from software as products to software as services, a study of how end users respond to such a transition is very important. This paper represents a first attempt in applying the technology acceptance model to the adoption of application services. The result may be generalizable to other emerging e-business technologies such as Web and grid services (Liu & Ma, 2004).

The findings of this study have some important managerial implications in practice. First, cognitive factors have been recognized to be the biggest obstacle for the widespread adoption of electronic medical records (Waegemann, 2002). In fact, the medical informatics community blames physicians for not being ready for a change to the computer. Among others, three critical cognitive factors are recognized (Voelker, 2003): (a) confidence in computers amongst their users, and especially in the availability, privacy, and security of data made of the electronic medical records; (b) adoption of a positive attitude towards computers in the workplace; and (c) adequate skills and proficiency in the use of the computer application. Our study confirms these general beliefs and addresses these factors using three constructs: self-efficacy, perceived ease of use, and perceived usefulness. Further research is needed to address other aspects of these cognitive factors. Second, our findings imply that, as for other technologies, enhanced usability and usefulness are a very important step toward the acceptance of Web-based electronic medical records. Third, to achieve end-user acceptance of electronic medical records, health care organizations can introduce interventions aiming at improving usability and utility perceptions as well as the end-user's Internet self-efficacy. However, as Table 6 shows, the total effect of self-efficacy on behavior intention is much stronger than that of perceived ease of use and usefulness. The significance of Internet self-efficacy suggests that enhancing self-efficacy may be a more effective option toward technology acceptance than enhancing usability and utility perceptions. After all, enhancing perceived ease of use and usefulness often requires significant effort in modifying the services being provided and is thus more difficult to do.

Note that self-efficacy is not a measure of skills. Rather, it reflects what individuals believe they can do with the skills they possess. As such, it is easy to be manipulated. For example, optimistic outlook, self-deception, heuristics judgments, and excessive wishful thinking can all induce overconfidence (Metcalfe, 1998). Then, will overconfidence bias our findings? In metacognition, underestimation was found to be detrimental to learning outcomes while overconfidence was found to be beneficial, although individuals with overconfidence tend to overestimate their knowledge (Gravill, Compeau, & Marcolin, 2002). However, its impact on perceptual and attitudinal variables is not known yet. More research seems necessary to study this phenomenon and examine whether overconfidence enhances one's willingness to accept a technology.

REFERENCES

Agarwal, R., Sambamurthy, V., & Stair, R. (2000). The evolving relationship between general and specific computer literacy: An empirical assessment. *Information Systems Research, 11*(4), 418-430.

Armitage, C. J., & Conner, M. (2001). Social cognitive determinants of blood donation. *Journal of Applied Social Psychology, 31*, 1431-1457.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215.

Bandura, A. (2001). Guide for constructing self-efficacy scales. *Measurement and Evaluation in Counseling and Development*. Retrieved from http://www.emory.edu/EDUCATION/mfp/banseg.html

Benbasat, I., & Zmud, R. W. (1999). Empirical research in information systems: The practice of relevance. *MIS Quarterly, 23*(1), 3-16.

Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. *MIS Quarterly, 25*(3), 351-370.

Chau, P. K. (1996). An empirical assessment of a modified technology acceptance model. *Journal of Management Information Systems, 13*(2), 185-205.

Chau, P. K. (2000). Influence of computer attitude and self-efficacy on IT usage behavior. *Journal of End User Computing, 13*(1), 26-33.

Chau, P. Y. K., & Hu, P. J. (2002). Investigating healthcare professionals' decisions to accept telemedicine technology: An empirical test of competing theories. *Information & Management, 39*(4), 297-311.

Chircu, A. M., Davis, G. B., & Kauffman, R. J. (2000). *The role of trust and expertise in the adoption of electronic commerce intermediaries* (Tech. Rep. No. WP 00-07). Minneapolis: University of Minnesota, Carlson School of Management.

Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly, 19*(2), 189-211.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-339.

Dinev, T., & Koufteros, X. (2002). *Self-efficacy and Internet usage: Measurement and factorial validity*. Proceedings of DSI, San Diego, CA.

Ding, L., Velicer, W. F., & Harlow, L. L. (1995). Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. *Structural Equation Modeling, 2*, 119-143.

Eastin, M. S., & LaRose, R. (2000). Internet self-efficacy and the psychology of the Digital Divide. *Journal of Computer-Mediated Communication, 6*(1), Retrieved from http://www.asusc.org/jcmc/vol6/issuel/eastin.html

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Gefen, D., & Keil, M. (1998). The impact of developer responsiveness on perceptions of usefulness and ease of use: An extension of the technology acceptance model. *The Data Base for Advances in Information Systems, 29*(2), 35-49.

Gefen, D., & Straub, D. (1997). Gender differences in the perception and use of e-mail: An extension to the technology acceptance model. *MIS Quarterly*, *21*(4), 389-401.

Gefen, D., & Straub, D. (2000). The relative importance of perceived ease of use in IS adoption: A study of e-commerce adoption [Electronic version]. *Journal of the Association for Information Systems*, *1*(8).

Gravill, J. I., Compeau, D. R., & Marcolin, B. L. (2002). Metacognition and IT: The influence of self-efficacy and self-awareness. *Proceedings of Eighth Americas Conference on Information Systems*, 1055-1064.

Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). Multivariate data analysis with readings. Englewood Cliffs, NJ: Prentice-Hall.

Hendrickson, A. R., Massey, P. D., & Cronan, T. P. (1993). On the test-retest reliability of perceived usefulness and perceived ease of use scales. *MIS Quarterly, 17*(2), 227-230.

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112*, 325-344.

Igbaria, M., Iivari, J., & Maragahh, H. (1995). Why do individuals use computer technology? A Finnish case study. *Information and Management*, *29*(5), 227-38.

Jayatilaka, B., Schwarz, A., & Hirschheim, A. (2002). Determinants of ASP choice: An integrated perspective. *Proceedings of the 35th Hawaii International Conference on System Science.*

Joo, Y., Bong, M., & Choi, H. (2000). Self-efficacy for self-regulated learning, academic self-efficacy, and Internet self-efficacy in Web-based instruction. *Educational Technology Research and Developmen*t, *48*(2), 5-17.

Keil, M., Beranek, P. M., & Konsynski, B. R. (1995). Usefulness and ease of use: Field study evidence regarding task considerations. *Decision Support Systems, 13*(1), 75-91.

Ketler, K., Willems, J. R., & Srinivasan, M. (2002*). The need for training in information technology: A survey of healthcare*. Proceedings of Decision Sciences Institute 2002 Annual Meeting, San Diego, CA.

Lederer, A. L., Maupin, D. J., Sena, M. P., & Zhuang, Y. (2000). The technology acceptance model and the World Wide Web. *Decision Support Systems, 29*(3), 269-292.

Levine, T., & Donitsa-Schmidt, S. (1998). Computer use, confidence, attitudes, and knowledge: A causal analysis. *Computers in Human Behavior, 14*(1), 125-146.

Lim, K.-S. (2001). *An empirical test of the technology acceptance model.* Proceedings of Decision Science Institute Annual Meeting, San Francisco, CA.

Liu, L., Li, C. C., & Karau, S. J. (2003). *A measurement model of trust in Internet stores.* Akron, OH: University of Akron.

Liu, L., & Ma, Q. (2004). Emerging e-business technologies for electronic medical records. *International Journal of Healthcare Technology and Management*, *6*(1/2), 1-22.

Lopez, D., & Manson, D. (1997). A study of individual computer self-efficacy and perceived usefulness of the empowered desktop information system. *Journal of Interdisciplinary Studies, 10*, 83-92

Ma, Q., & Liu, L. (2004). The technology acceptance model: A meta-analysis of empirical findings. *Journal of End User Computing, 16*(1), 59-72.

Marakas, G. M., Yi, M. Y., & Johnson, R. D. (1998). The multilevel and multifaceted character of computer self-efficacy: Toward clarification of the construct and an integrative framework for research. *Information Systems Research, 9*(2), 126-163

Mathieson, K. (1991). Predicting user intentions: Comparing the technology acceptance model with the theory of planned behavior. *Information Systems Research*, *2*(3), 173-191.

Metcalfe, J. (1998). Cognitive optimism: Self-deception or memory-based processing heuristics? *Personality and Social Psychology Review, 2*(2), 100-110.

Munro, M. C., Huff, S. L., Marcolin, B. L., & Compeau, D. R. (1997). Understanding and measuring user competence. *Information & Management, 33*(1), 45-57.

Nahl, D. (1996). Affective monitoring of Internet learners: Perceived self-efficacy and success. *Proceedings of the 59th ASIS Annual Meeting* (Vol. 33), 100-109.

Nahl, D., & Meer, P. (1997). User-centered assessment of two Web browsers: Errors, perceived self-efficacy, and success. *Proceedings of the 60th ASIS Annual Meeting* (Vol. 34), 89-97.

Peterson R. R., & Fairchild, A. M. (2003). *Adoption trends in application service provisioning: An exploratory field study of small and medium-size enterprises*. Proceedings of the European Conference on Information Systems (ECIS), Naples, Italy.

Ridings, C. M., & Gefen, D. (2000). Applying TAM to a parallel systems conversion strategy. *Journal of Information Technology Theory & Application, 2*(2).

Schumacker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. Mahwal, NJ: Lawrence Erlbaum Associates Publishers.

Straub, D. W. (1994). The effect of culture on IT diffusion: E-mail and fax in Japan and the US. *Information Systems Research, 5*(1), 23-47.

Susarla, A., Barua, A., & Whinston, A. B. (2003). Understanding the service component of application service provision: An empirical analysis of satisfaction with ASP services. *MIS Quarterly, 27*(1), 91-123.

Taylor, S., & Todd, P. (1995). Assessing IT usage: The role of prior experience. *MIS Quarterly, 19*(4), 561-570.

Torkzadeh, G., & Van Dyke, T. P. (2002). Effects of training on Internet self-efficacy and computer user attitudes. *Computers in Human Behavior, 18*, 479-494.

Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research*, *11*(4), 342-365.

Venkatesh, V., & Davis, F. D. (1996). A model of the antecedents of perceived ease of use: Development and test. *Decision Sciences, 27*(3), 451-481.

Voelker, K. G. (2003). *Primer on electronic medical records*. Retrieved January 3, 2003, from http://www.elmr-electronic-medical-records-emr.com/electronic_medical_record_Primer.htm

Waegemann, P. C. (2002). *Status report 2002: Electronic health records*. Medical Records Institute.

This work was previously published in Journal of Organizational and End User Computing, Vol. 17, edited by J.N.D. Gupta, M. Adam Mmahmmood, S. Graciela, pp. 38-57, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

APPENDIX A

Initial Items for Internet Self-Efficacy

Item #:	Item	Source	Construct α	Loading
1	I feel confident to use search engines like Yahoo, Google, and AltaVista.	Joo et al.	.95	NA
2	I feel confident to download necessary material from Internet.	Joo et al.	.95	NA
		Dinev & Koufteros	.96	.78
		Torkzadeh & van Dyke	.94	.77
3	I feel confident to search for information on the Internet.	Joo et al.	.95	NA
		Eastin & LaRose	.93	0.65
		Torkzadeh & van Dyke	.93	.78
4	I feel confident to visit a web site if I am given a web address (URL).			
5	I feel confident to log on to a website if I have the account information.			
6	Overall, I feel comfortable when I am using Internet.			
7	I feel it is easy for me to learn to use an Internet feature.	Dinev & Koufteros	.96	.82
8	I feel confident to perform data transactions (e.g., buy a book) on the web.	Dinev & Koufteros	.96	.58

APPENDIX B

Survey Instrument

PU1	Using this system in my job would enable me to accomplish tasks more quickly
PU2	Using this system would improve my job performance
PU3	Using this system would enhance my effectiveness on the job
PU4	Using this system would make it easier to do my job
PU5	This system is useful in my job
PEOU1	I found it was easy to do whatever I want
PEOU2	The navigation on the site is easy
PEOU3	Learning to operate this system is easy for me.
PEOU4	This web site provides information content that is easy to understand
PEOU5	I found this system was flexible to interact with
ISE1	I feel confident to use search engines like Yahoo, Google, and AltaVista.
ISE2	I feel confident to log on to a website if I have the account information.
ISE3	I feel confident to download necessary material from Internet.
ISE4	I feel confident to search for information on the Internet.
ISE5	I feel confident to visit a web site if I am given a web address (URL).
ISE6	Overall, I feel comfortable when I am using Internet.
BI1	I will support it if my clinic decides to use this system.
BI2	I would like to come back to this site for a second look
BI3	I do not mind spending some time to learn how to use this system for my work
BI4	I am willing to use the system for my work
BI5	It would be efficient if we were going to adopt this system.

Chapter 5.15
Intelligent Portals for Supporting Medical Information Needs

Jane Moon
Monash University, Australia

Frada Burstein
Monash University, Australia

ABSTRACT

The aim of this chapter is to review the way portal technology can assist users seeking medical information. There has been an increase in health Internet usage, and better health-care delivery outcomes are predicted as users are better informed when making medical decisions. At the same time, there is much concern about the need for medical portals to meet community information needs. This chapter discusses what constitutes an intelligent portal, discusses desirable portal components and attributes of intelligent portal features, and how these can be implemented to meet the needs of diverse users. Seven Australian medical Web sites have been analysed according to intelligence features. The results and analysis are presented and discussed, in particular, with respect to their functionality as defined for intelligent portals. The discussion is focused on the extent to which these attributes help users with their information seeking and therefore support their decision-making processes.

INTRODUCTION

Internet technology provides a useful and easy information channel, which enables various content materials to be streamlined to users on request. The amount of information that is now available electronically is as large as that produced in non-electronic form since the development of print media. With the global proliferation of Internet use, obtaining information in this way has become the norm for many. This presents opportunities for providing the most relevant and current content to users when they need it. Information relating to health is no exception; more and more people are relying on the Internet as a way of accessing information relating to health. Research shows that one in four of 80 million users of the World Wide Web in 2000 visited at least one medical Web site, and the number of visitors to such sites grew faster than with other sites in general (Murray, 2002).

However, studies show that many Internet users looking for information are often frustrated by a large amount of irrelevant information retrieved by search engines and by the time it takes to obtain truly relevant information. Personalization and customization of the interface through portal technology are a timely response to these problems (Finkelstein & Aiken, 2000).

There are many definitions of portals. Merrill Lynch in November 1998 described portals as "applications that enable companies to … provide users with a single gateway to personalized information needed to make informed business decisions" (Quirk, 2001, p. 2). A very generic definition of a portal is as a single gateway connected by a server that connects people with information (Harvard Computing Group, 2002).

The limitations in portal technology and specific human behaviour are factors affecting the efficiency of information retrieval. Consequently, modern portals are built with the intent of implementing some intelligent ways of catering for different needs. Improving in portal technology and understanding the dynamics of human nature can improve information retrieval.

The aim of this chapter is to review the way portal technology can assist users in broader community contexts. In particular, we look at how portals are employed for meeting community medical information needs and ways to improve these medical portals.

Specifically, we explore the extent to which these portals behave "intelligently" in addressing users' needs. The analyses of medical portal intelligence issues are discussed in terms of search engines, spell checking, "sounds like" indexing, parsing, ontology, use of thesaurus, personalization, and decision facilities or expert system functionality. Finally, Australian medical portals are analysed in order to illustrate the problems and opportunities of intelligent community portals.

HEALTH INTERNET USAGE

Increasingly people are seeking medical information on the Internet. In America, Ehrenberger in 1991 predicted that by 2005, 88.5 million Americans will be using the Internet to seek health information (Ehrenberger, 2001). The current survey exceeds his prediction. According to the Harris Interactive consulting firm, health Internet users grew exponentially from 50 million in 1998 to 69 million in 1999, 97 million by year 2001, and reached 110 million in year 2002. The research concluded that:

The Internet continues to be used by huge, and growing, numbers of the public interested in getting information about particular diseases or treatments or about staying healthy. The results also demonstrate the critical importance to health care websites of the need to be quickly and easily accessible through search engines and portals. (Taylor, 2002, p. 1)

Brodie et al. report (2000) there is no significant difference in information-seeking habits due to age, and there is a direct correlation between computer usage and access to health information:

Once people gain access to the Internet, its use at home to get health information is similar across income, education, race and age. Therefore the number of persons using the Internet to access health information should rise along with computer use. (Brodie et al, 2000, p. 262)

A report by the Australian Bureau of Statistics in 2002 on Internet usage shows that there has been a gradual increase in the use of computers and the Internet at home. In 2000, 53% of households (3.8 million) had a computer, an 11% increase over 1999. About one-third had Internet access, a 49% increase over 1999. The Australian Government is taking a proactive role in developing medical portals to encourage general use of the Web for the dissemination of medical information (National Health Information Management Advisory Council, 2001). Australian government portals such as Health*Insite* and HealthFinder argue that the better informed the users are, the better their health outlook (National Health Information Management Advisory Council).

Because there is an increased usage of the Internet and because medical information is one of its most valuable resources, the control of this content is very important. Research shows that the degree of trust people have in the content of the information from medical portals is increasing (Ho & Tang, 2001). Anecdotal evidence suggests that some users prefer information from the Internet over advice from their consultants. For example, a woman who was recommended to have a splenectomy (removal of spleen) by her consultant refused surgery twice. However, she changed her mind after she read the same information on the Internet. She put more trust in the information she read from the Internet than her consultant. Ho and Tang showed that of their 1,232 respondents more than 60% believed that the Internet contained more in-depth information than either TV or newspapers. Whilst Ho and Tang's research reveals trust shown by the users, there are conflicting reports, however, suggesting that users are still confused about the material they read from medical sites (Wilson, 2002).

There is also an issue of retrieving vast, complex, and widely distributed medical information. How do you retrieve correct, relevant, and reliable information in a timely manner? How do you know when you have searched enough? It is extremely difficult for users to access the exact type of medical information they need (Shepherd, Zitner, & Watters, 2000). Whilst there is a growing number of Internet users and reliance on information from these sources, many users looking for information are often frustrated by the amount of irrelevant information retrieved via search engines, the time it takes to retrieve this information (Finkelstein & Aiken, 2000), and its perceived quality (Wilson, 2002).

CURRENT PORTAL TECHNOLOGY

In Latin, the word *porta* means a door, gate or gateway, entrance, or opening. A portal can be defined as a gateway to information (Clarke & Flaherty, 2002; Quirk, 2002; Rao, 2001). It is Web-based and connects a set of commonly used information or services via a link from a single page (Notes, 2002; O'Leary, 1999).

There are many types of portals. In general portals can largely be classified into four categories:

- Corporate or enterprise (intranet) portals;
- E-business portals;
- Personal portals; and
- Public portals. The public portals include general public portals such as Yahoo, AOL, AltaVista, and community portals.

The types of medical portals we will be looking at are community portals. Some of the functionalities that can be implemented into medical portals are as follows (Collins, 2001; Millen, 2000; Tushkar, 2000):

- **Browse/Navigate documents**—this allows users to search for and locate information manually by navigating through a directory structure.
- **Collaboration**—represents an important aspect of a portal and is a powerful tool. It

includes instant messages such as "chat," document sharing, videoconferencing, virtual conferencing, and discussion forums.

- **Content management**—a process of approving, authoring, delivering, maintaining, and publishing content integrated with or accessed from a portal or other Web site. This can be implemented as a separate component with a common database accessible through the portal.
- **Directory**—a directory within the portal's enterprise taxonomy. It is a collection of data structured into a hierarchy of categories.
- **Document management**—similar to content management but mainly dealing with control and management of an enterprise's documents. It involves managing electronic files, including scanned images of pictures and documents.
- **Personalization**—portal personalization can be achieved at different levels. An individual can have a personal setting; a group of people sharing the same function can have group settings. An organization can set up its own setting. This feature allows a portal to be customized according to needs.
- **Search engine**—the ability to search is the fundamental component of a portal. It allows users to browse the content, retrieve information on the content basis, and link to other repositories for information.
- **Subscribe/What's new**—allows users to register an interest in a particular aspect of a portal. Portals then notify the user when changes in content occur.
- **Taxonomy**—a classification scheme to organize information. Additional functionalities, such as metadata added to the taxonomy, could help to organize documents into different categories that could make the information easy to browse, search, or navigate.
- **Online community**—provides the opportunity for users with similar interests to share their medical conditions and experiences. This could range from sharing experiences of treatment to discussing medical conditions and alternative medicine.

For the purpose of this chapter, we assume that a portal is an Internet-based information system providing uniform access to different sources of information in an enterprise and maintaining dynamic links to information resources. A portal supports communication within the enterprise and connects people with information and applications they need for performing tasks. Unlike a conventional Web site, portals should support both push (subscription) and pull (search) functions in assisting users to gain access to essential contextual information (Probst, Raub, & Romhardt, 2002).

INTELLIGENT MEDICAL PORTALS

The term "intelligent medical portal" refers to a Web gateway environment that allows users with varied medical interests and diverse backgrounds to access medical resources and information to support their decision-making processes, which often involve critical medical problems.

Current portal technology fails to deliver the types of medical information needed by users (McKemmish, Burstein, Manaszewicz, & Fisher, 2002). In the following paragraph, some of the users' experiences are included to illustrate this deficiency of portal technology.

A 44-year-old female who wants to go on a Pap smear registry so that she can be notified automatically every two years could not get a satisfactory answer from the Australian government medical Web site, Health*Insite*. It provided her with hyperlinks but these were not operational—not much use to the user. She was inundated with general information on Pap smears instead. In another medical Web site, Health Network, users were encouraged to write to the consultants.

An enquirer's e-mail did not get an answer for 3 weeks, and in that time she had already found the answers through other channels.

A quality search engine with built-in intelligence features such as a thesaurus, sounds-like index, ontology, and parsing could assist users with difficult medical jargon, treatments, and the spelling of difficult drug names. However, current portal technology does not have many of these intelligence features. If a person looking for medical information on rheumatoid arthritis types phonetically *rumatoid arthritis*, he would not get any returns in current medical portals. Ideally some assistance should be provided as in Google: "Did you mean: …" This could help the users.

Current medical portals do not have provisions for differentiating users with different educational background, age, ethnicity, location (urban or rural), and stage of disease. For example, there is no distinct information for a 63-year-old female with early breast cancer or for a 35-year-old mother experiencing advanced breast cancer living in a rural area. Hyperlinks to the actual breast cancer information seem to be provided with a "one size fits all" category, which does not suit everyone (McKemmish et al., 2002).

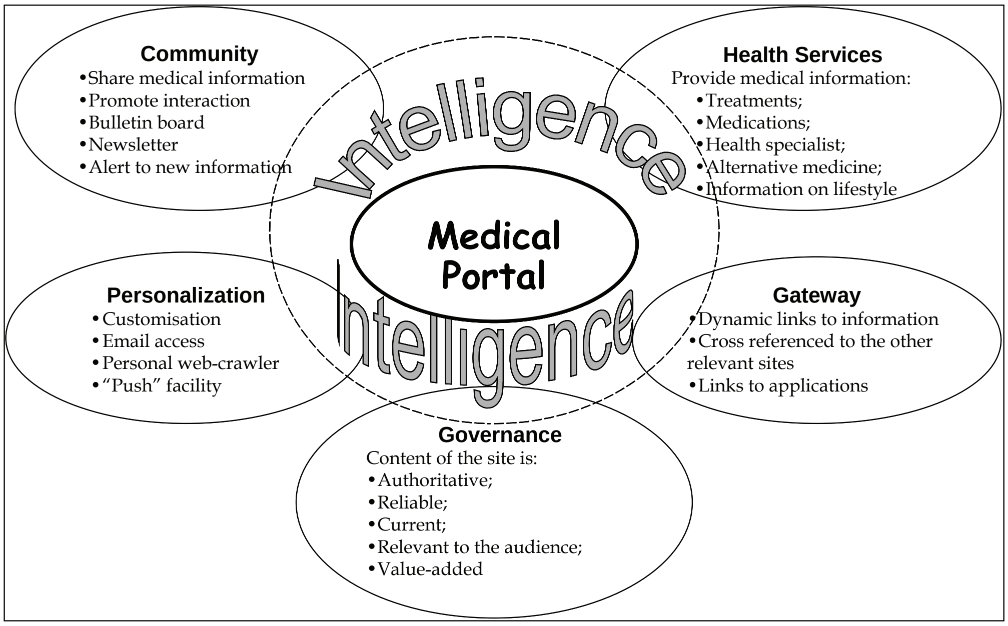
The problems lie in the limited portal technology, which is inflexible, and inadequate search engines, which do not allow for sophisticated searching facilities.

How can a portal be made to behave intelligently so that users can get access to vastly distributed medical information easily? McKemmish et al. (2002) describe how an intelligent portal "will improve the structure and manage the quantity of information presented" (p. 5). Indeed an intelligent portal should provide information that is of high-quality, relevant, and timely. It should address the needs of users by helping them to solve issues that are often critical. With the use of intelligent portal technology, combining knowledge management issues of personalization and decision support, the needs of the Internet users of medical portals could be better served.

Components of Medical Portals

The following diagram shows the components of medical portals which should be implemented

Figure 1. Proposed components of medical portals



to ensure their efficient information retrieval capabilities. Figure 1 depicts the major components expected from the medical portal. These components are:

- **Health services**—medical portals should provide foremost medical information pertinent to that site, i.e., if it is breast cancer, provide medical facts on treatments, as well as associated and often sought-for information on lifestyle and complementary medicine.
- **Gateway functionality**—should provide dynamic links to other sites, facilities for cross-referencing, and links to other applications.
- **Community support**—involves a space to share medical information and treatments, share personal experiences, offer a bulletin board to provide information about the latest technologies and medications; it could also be a forum for medical professionals to share their knowledge and peer reviews on aspects relevant to the medical portal.
- **Personalization**—should provide facilities to customize sites so users can get to them with ease and e-mail systems to receive the latest news and to send any queries or share any information at a personal level.
- **Governance**—the contents of the information can have significant influence on the livelihood of users. Some users could make important decisions about their treatments on the basis of what they read from the Internet. Thus it is paramount to ensure the quality, authority, currency, and relevance of the information.

Intelligence Features of Medical Portals

Intelligence is the necessary interface between the users and those components that make the portal a useful and unique information system to satisfy the potential dynamic needs of a user. As pointed out by Shepherd et al. (2000), medical information is vast and distributed. To retrieve these items of information effectively, there needs to be a quality search engine that "understands" the user's needs.

The following intelligence features can be incorporated into medical portals to improve information retrieval:

- **Search engine**—the ability to search is fundamental for the implementation of a portal. It allows users to browse the content and retrieve information on a content basis, and assists users by providing links to other relevant information repositories. Use of artifical agents such as artificial neural network (pattern matching) and fuzzy logic can be applied to specify the search (Negnevitsky, 2002).
- **Spell checking**—often medical terms and the drugs that are used for treatments are hard to spell, and there are both English and American variants. Current portal technology does not have provisions for misspelt words. A list of possibilities should be shown if the word entered is spelt incorrectly.
- **Sounds-like indexing**—this is to search for "sounds like" terms. Fuzzy logic can assist in refining the search. Medical jargon is extremely difficult to spell; thus, sounds-like indexing could add value to the portal.
- **Parsing** sentences or phrases to identify key words and then searching for the content. In general, search engines are based on key words and often inadequate. For example one could be searching for "evening primrose" for either dermatitis or menopausal symptoms. Parsing allows the user to enter more than one search option, and this could be a powerful tool if implemented.
- **Metadata**—a mechanism used for resource description and specification. It can be used for marking or tagging the repository content

for locating the resource according to parameters specified in the metadata. For example, the Australian Government Locator Service (AGLS) was introduced in 2002 for use by government departments and agencies to improve the visibility and accessibility of their services and information provision over the Internet. It uses a standard set of 19 descriptive elements (National Archives of Australia, 2000).

- **Ontology**—representing a domain discourse in a way that terms specific to this domain can be identified from synonyms. For example, the word *drug* can be expressed in many different ways (medicine, medication, tablets, pills, prescription). So the word retrieved should have relevance to the search. That is, if *drug* is typed into a breast cancer portal, the portal should return Tamoxifen, for example, a common anti-estrogen drug.
- **Thesaurus**—can be used together with ontology to describe a structure of the problem domain. In our context, it can be too broad if it includes outside knowledge not relevant to medicine. However, if complemented by a metadata repository, it can be narrowed.
- **Personalization**—special service to the specific users, which can be set up at various levels. The users can have personal settings in the portal (profile) such that they can share information and receive notifications when there is a change to the content. The user can set up a Web spider to collect information of interest and push it to the user as soon as it is collected. Users can tailor the "look and feel" of the portal interface to suit their preferences. They can also select personal services such as news, e-mail, and relevant service sites to be available within a portal page.
- **Decision-making facilities**—some analytical tools can be provided with the portal to help users with decision-making processes (loan calculators, for example). These can be more sophisticated if built on the basis of decision analysis algorithms, or implement if/then/else rule reasoning engines. The users' profile information can be utilised here to assist them with their enquiries.
- **Quality-rated information**—the information retrieved is presented to the user with some indicators relating to generally accepted criteria of quality, such as author, publisher, currency, purpose, etc. This should help the user to be in a position to make an informed choice.
- **Content management**—a process of approving, authoring, delivering, maintaining, and publishing content integrated within or accessed from a portal or other Web site. It can be implemented as a separate component with a common database accessible through the portal.
- **Document management**—similar to content management but deals with control and management of the enterprise documents. It involves managing electronic files, including scanned images of pictures and documents. It can also incorporate workflow facilities for supporting organizational processes. In medical portals this can be a facility to access insurance forms and claims.
- **Knowledge management**—this can be seen as a mechanism for capturing know-how and know-why in a knowledge repository.
- **Use of artificial intelligence**—such as artificial neural networks, fuzzy logic, expert systems for implementing learning capabilities, adaptation, and dynamic reasoning.
- **Other intelligent agents**—intelligent agents can assist in searching through the data to support "push" facilities of the portal (Probst et al., 2000). They can save users time by sifting the data, retrieving relevant information, and pushing it to the user on request.

The intelligence features listed above can be incorporated within portal components as shown in Figure 1. The following sections present the analysis done against the intelligence features identified above to see to what extent existing medical portals comprise those features.

MEDICAL PORTALS ANALYSIS

In this section, we will analyse Australian medical portals to assess the extent to which identified intelligence features are present. The features that will be investigated are as follows:

- Search engine;
- Spell check;
- Sounds-like index;
- Parsing;
- Ontology;
- Personalization; and
- Thesaurus.

The Australian medical Web sites were specifically chosen to assess to what extent medical information needs are met. The problems and types of information the users were seeking were pertinent to Australians. For example, recent user needs analysis regarding availability of breast-cancer-related information online showed that one third of the users live in rural areas. For those users, availability of medical specialists, treatments, and travelling and accommodation during the treatments are issues (Manaszewicz, Fisher, Williamson, & McKemmish, 2002). The current medical portals should address those needs.

The seven most popular Australian medical Web sites have been chosen and analysed according to intelligence features. We have analysed how these features are implemented in medical portals. The portals analyzed are listed in Table 1.

The next section provides an analysis of each intelligence feature with screen shots.

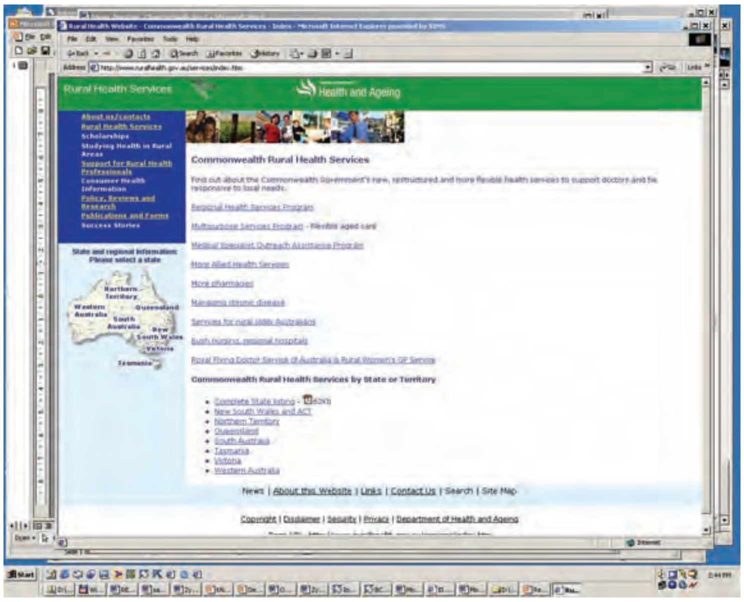
Search Engine

The first analysis was to look for the presence or absence of search engines in those seven medical Web sites. Not all of these Web sites were provided with search engines. For example, Rural Health does not provide a search capability, either internal or external. In fact, it operates as a gateway to the various resources available from other medical or health Web sites and draws heavily on content from Health*Insite* (see Figure 2). This Web site provides access to a knowledge repository selected mostly from a review of existing resources, "personalizing" them to a specific audience—the Australian rural population.

Table 1. Medical portals analyzed in this study

Australian Medical Portal	Web address	Managing organisation
BetterHealthChannel	www.betterhealth.vic.gov.au	Victorian Government's Department of Human Services
Health*Insite*	www.healthinsite.gov.au	Commonwealth Government of Australia
RuralHealth	www.ruralhealth.gov.au	Office of Rural Health (ORH), Australian Commonwealth (Federal) Department of Health and Ageing
Australian Indigenous Health-*InfoNet*	www.healthinfonet.ecu.edu.au	School of Nursing and Public Health, Edith Cowan University
HealthConnect Department of Health and Ageing's website	www.health.gov.au/	Australian Government Department of Health and Ageing
Medicine Australia (MedAu)	www.medineau.net.au	Northern Rivers Division of General Practice (NSW) Ltd
HealthNetwork	www.healthnetwork.com.au	General health information site run by Editorial Committee of medical practitioners for educational purposes

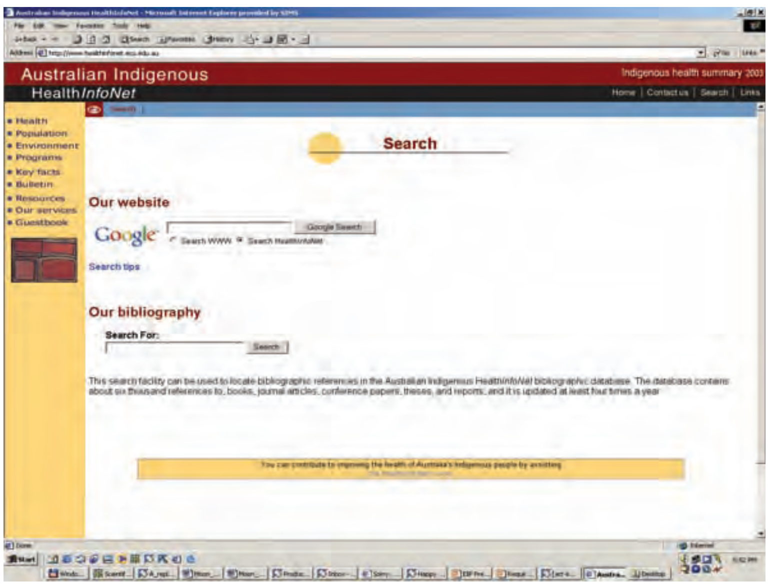
Figure 2. Rural Health gateway



Better Health Channel used an internal search engine to search the site content and used Google as an external search engine. Health*Insite* used Yahoo externally and had its own internal search engine. Health Network also offered search engines for selective links.

Figure 3 shows an example of a search facility for the Australian Indigenous Health*InfoNet* portal (http://www.healthinfonet.ecu.edu.au). It is a resource that provides medical information for Indigenous Australians. It contains literature reviews on various topics listed on the site map.

Figure 3. Search facility in Australian Indigenous HealthInfoNet



It also uses Google for external search (see Figure 3).

Spell Checking or Sounds-Like Search

None of the medical Web sites provided spell checking or sounds-like indexes. The following is an example from Better Health. When the common anti-estrogen drug for breast cancer, tamoxifen, was misspelt as *Tomoxifen*, there were no results returned, as can be seen in Figure 4a.

However, it provided links to the Google search engine. When the drug tamoxifen was misspelt as *tomoxifen*, Google provided the correct drug name, as shown in Figure 4b.

Other medical portals such as Health Network offered partial matching of spelling (sounds-like) in that you could type the first three letters of a word and still recognize the full word. For instance, if one was searching for the country Vietnam, typing *Vie* would return the country Vietnam if the first three letters were spelt **correctly**. However, it did not offer the facilities of thesaurus, spell check or sounds-like index (see Figures 5a and 5b). This feature provides some help to the users, but it could also retrieve many drugs or chemicals with the first three letters that are relevant to the query.

However, if *Vie* is mistyped as *Bie* or *betnam*, it will not recognize it.

Parsing

Parsing is the ability to separate a sentence into meaningful words. For example, if the sentence "Where can I find Tamoxifen?" is entered, a search engine should look for the key word in the sentence, "Tamoxifen," in this case, and retrieve the requested information.

Medicine Australia (MedAu), a general medical portal, was the only Web site that had the ability to show parsing. For example, the following: To the question "Where is Lithium?" the search result was the same as when the word "Lithium" was typed. The portal was able to extract the word from the sentence and retrieved information relevant to the sentence. It retrieved two articles that had information on lithium (see Figure 6).

Figure 4a. Better Health Channel sounds-like search facility

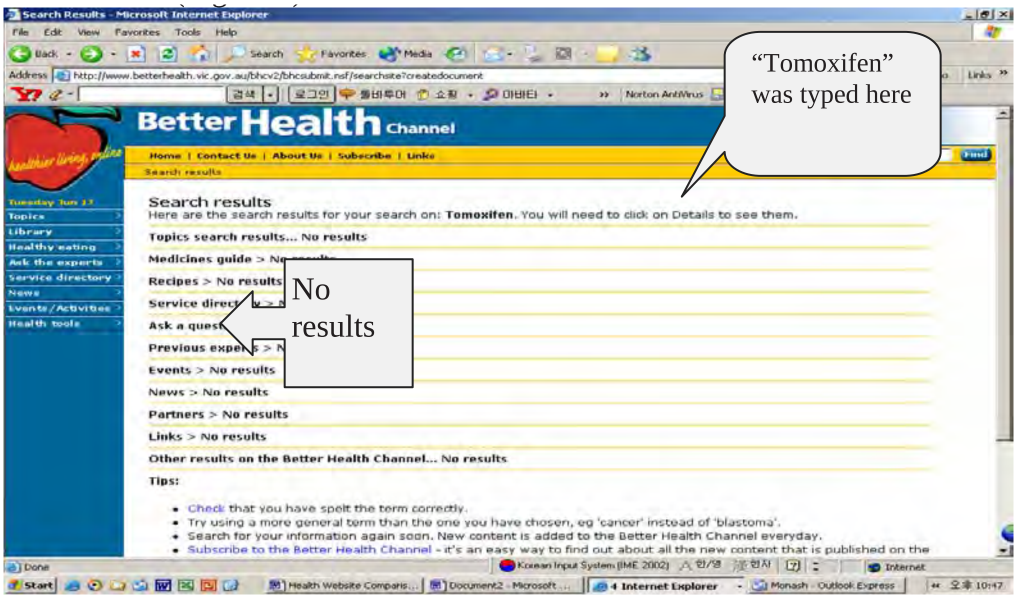


Figure 4b. Better Health Channel search facility through Google

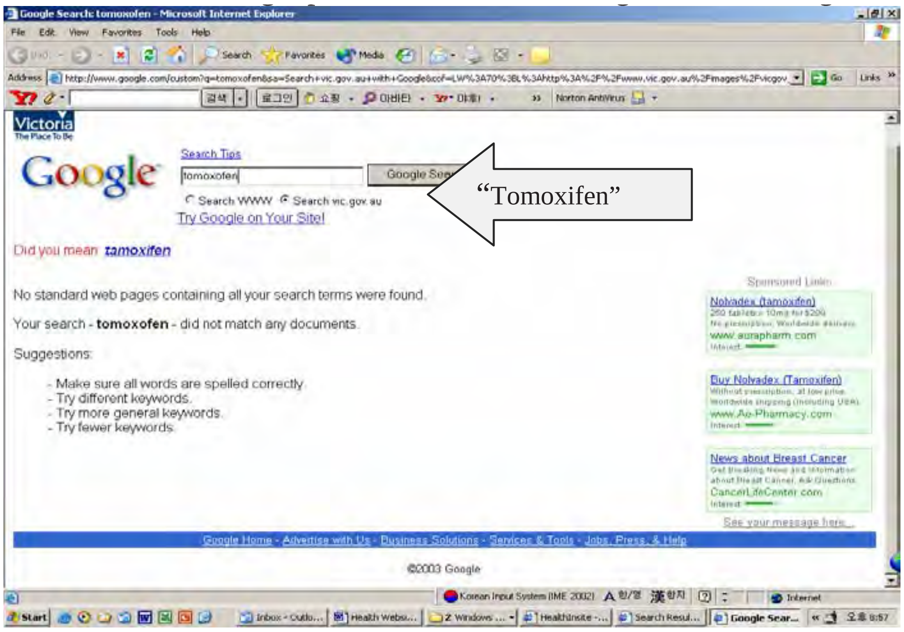
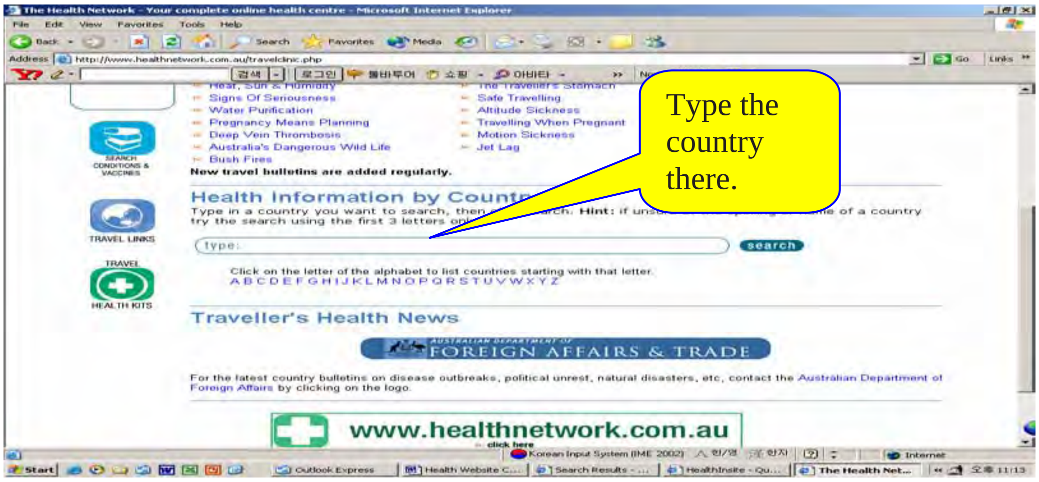


Figure 5a. Health Network sounds-like search

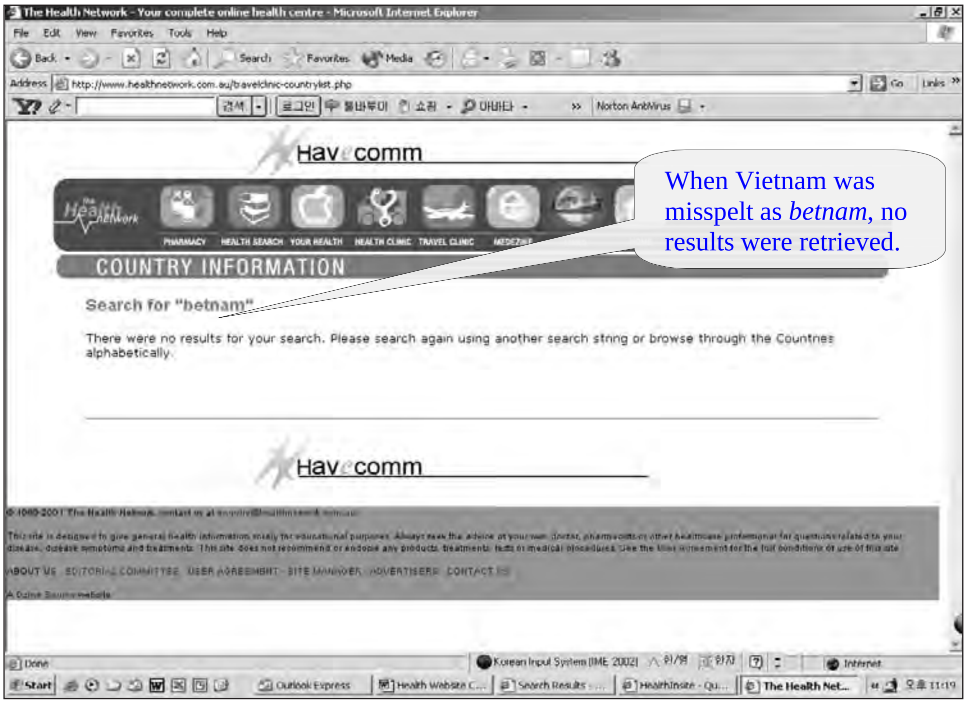


However, none of the other medical Web sites tested showed the ability to parse. In Figure 7 we present an example of the search with Better Health for the answer to the question "Where can I find tamoxifen?" As can be seen, this search found no results.

Another example was from Health*Connect* in response to "Where can I find tamoxifen?" (Figure 8).

The same question was entered into the Better Health portal through its link to Google search. When the sentence "Where can I find tamoxifen?"

Figure 5b. Health network sounds-like search



was entered, it was able to extract tamoxifen and returned the information shown in Figure 9.

Personalization

Personalization is one of the intelligence features of the portal by which users can:

- Receive an e-mail from the Web site if there is a major change to the Web site.
- Personalize the site interface—select the types of information they want to view.
- Have personal selection lists that can be placed where users can find them easily.
- Have automatic search items that match interests of users through "My interest link."

Figure 10 shows an example of how Health-*Insite* provides personalization features.

Health Network provides personalized contacts where the users can write to medical experts, who will promptly answer their questions (see Figure 11).

This provides the possibility for an added intelligence that offers the users personalized service, in addition to searching facilities. However, it is entirely manual, the questions being answered by people who are also logged on to that Web site.

Ontology and Thesaurus

No medical portal we have analysed had an ontology linked to it. Only one medical portal, Health-*Insite*, showed some presence of a thesaurus. But it was not exactly what would be expected when the full thesaurus is implemented and accessible through the portal. For example, when the drug tamoxifen was misspelt as *tomoxifen*, the search could not find it. It offered links to the Thesaurus Navigator instead (see Figure 12).

Figure 6. Example of Medicine Australia (MedAu) parsing ability

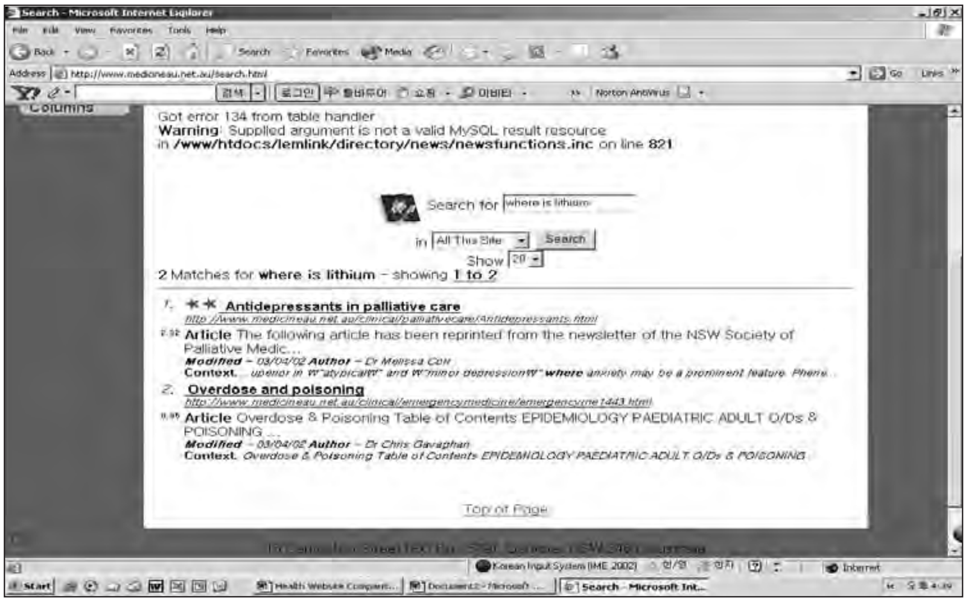


Figure 7. Example of Better Health Channel search without parsing

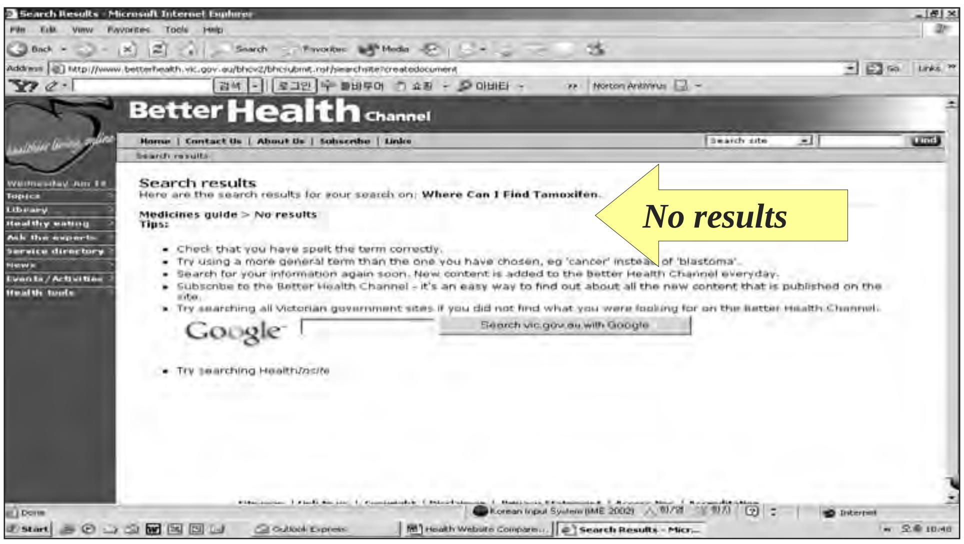


Figure 8. Example of HealthConnect search without parsing

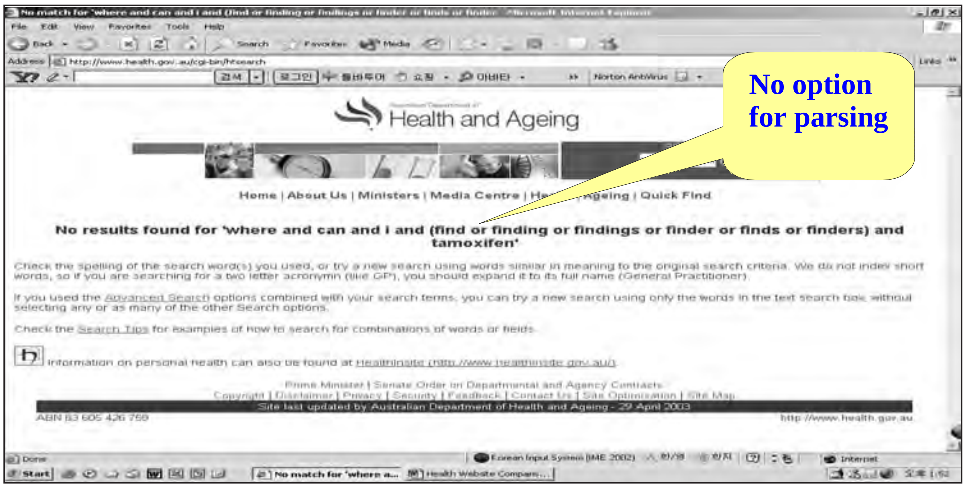


Figure 9. Improved search in Better Health Channel with Google parsing

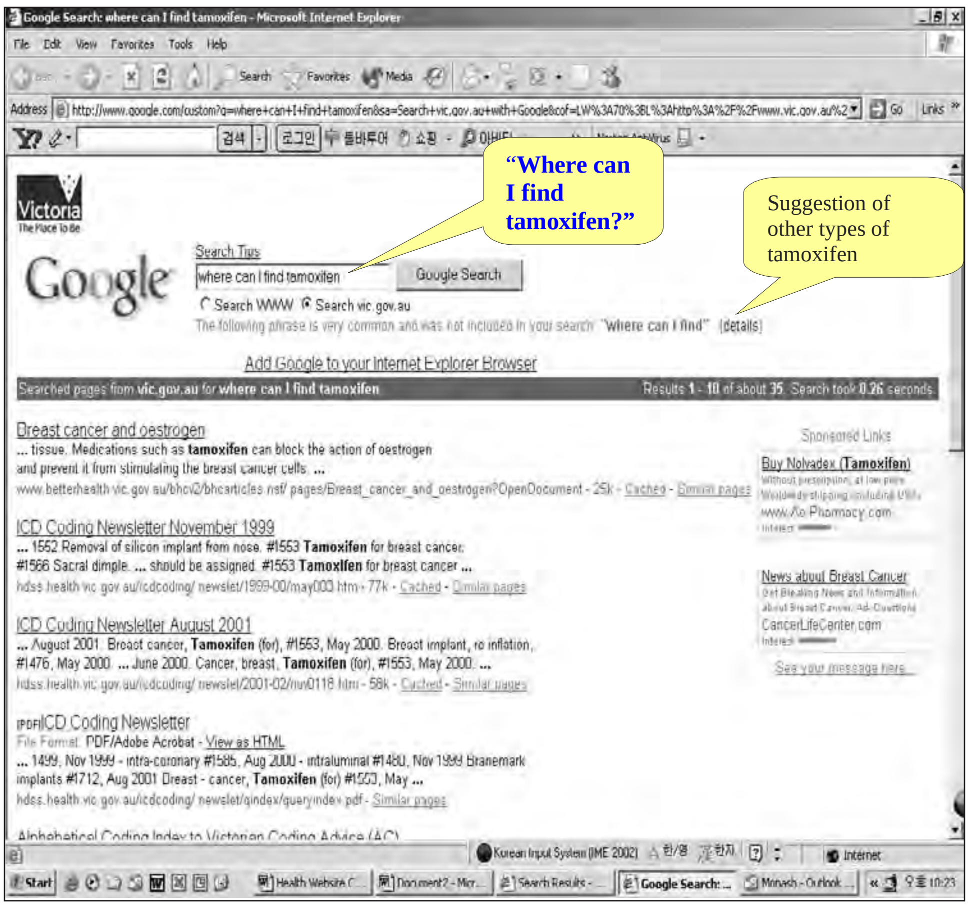


Figure 10. HealthInsite personalization facility

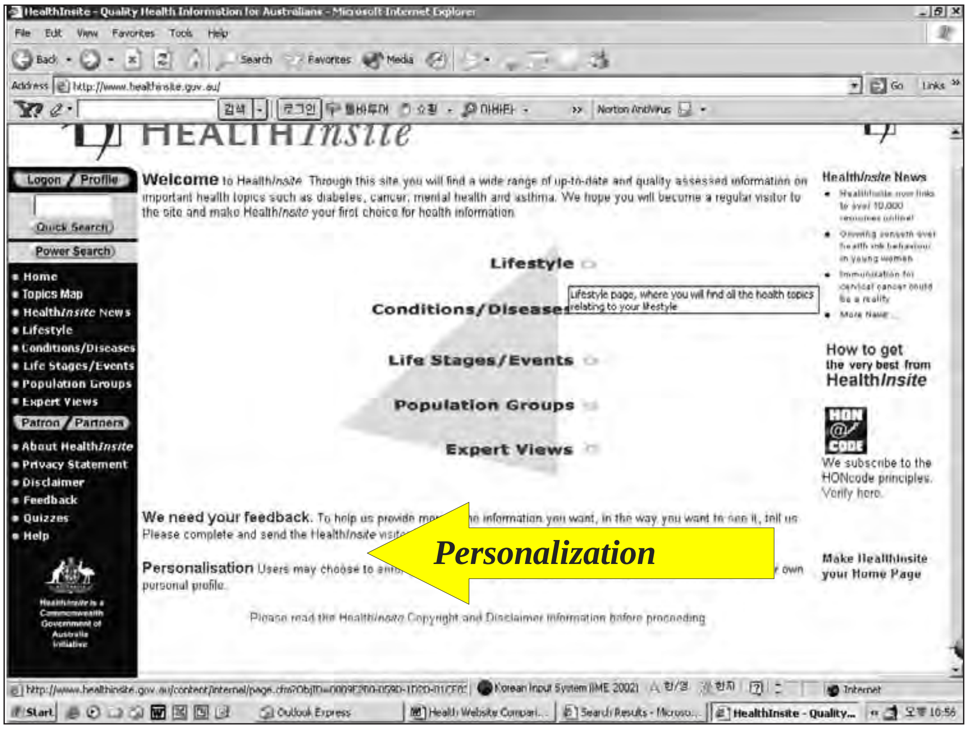


Figure 11. Health Network personalization facility

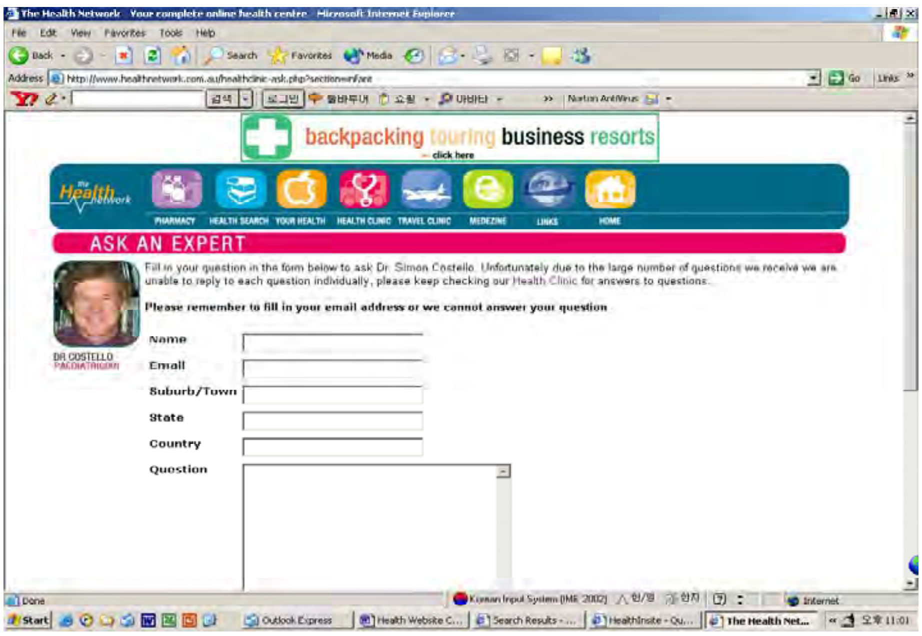
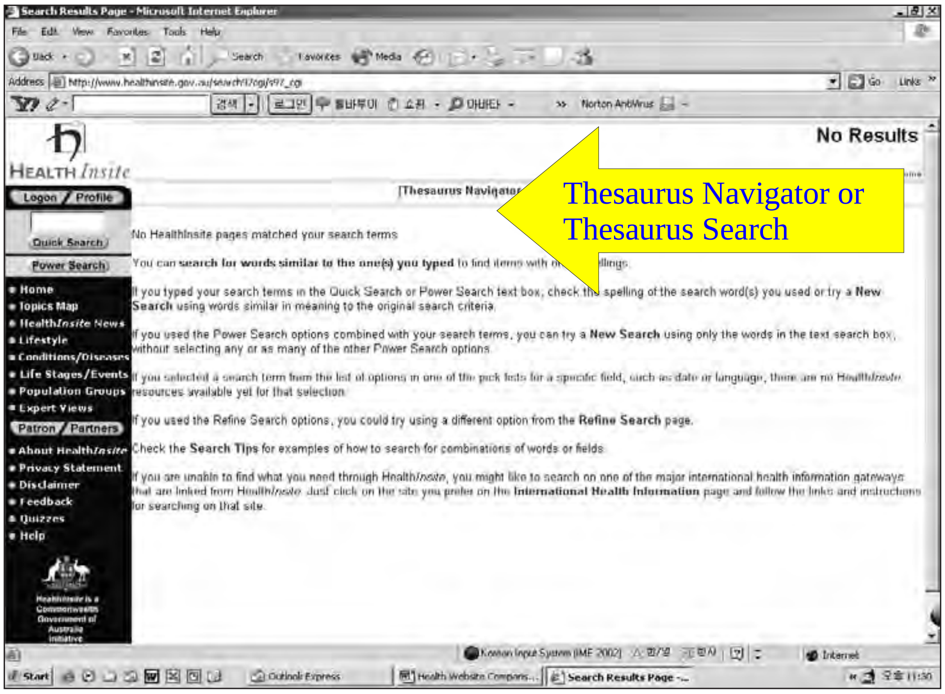


Figure 12. Thesaurus in HealthInsite



At present a thesaurus is only implemented to store the names of diseases but is not useful in searching for types of drugs. And, the information on diseases is only available if the correct spelling is typed in. Also, Medical Subject Headings – MeSH is used in medical portals which makes it extremely difficult for users to navigate, i.e., they have to understand the tree hierarchy, which is expressed all in generic/medical terminology.

Analysis of Results

Table 2 provides a comparative summary of medical portals. It demonstrates that very few of the intelligence features identified in the literature and listed in the section on portal technology are currently implemented in medical portals, and none of the portals analysed have more than three of them present.

Most of the portals had a search engine, the exception being Rural Health. However, the position of the search engine was not always obvious and in some cases was hard to find For instance, it took some effort to find the search engine in the Health*InfoNet* as it is located on a separate page as opposed to a rectangular box on the front screen, as in most of the portals.

None of the portals provided spell checking, sounds-like indexing, parsing, ontology, or decision facilities. Health*InfoNet* had features where if you typed the first three letters correctly, it could retrieve the information.

Health*Insite* offered personalization via self-registration. It also offered some level of thesaurus, in a form of a list of all topics covered in the site. However, on the whole, the Health*Insite* had more intelligence features than the other sites analyzed.

CONCLUSION

The advantage of Web-based medical portals lies in their ability to make medical information

Table 2. Medical portal comparison

X = No, absence of the feature *Yes = Presence*

Australian Medical Portal	Search Engine	Spell check	Sounds like index	Parsing	Ontology	Personalization	Thesaurus	Dec. Facilities
BetterHealth Channel	Google	X	X	X	X	X	X	X
Health*Insite*	Yahoo	X	X	X	X	Yes	Yes	X
Rural Health	X	X	X	X	X	X	X	X
Health*InfoNet*	Yes	X	X	X	X	X	X	X
HealthConnect	Yes	X	X	X	X	X	X	X
Medicine Australia (MedAu)	Yes	X	X	X	X	X	X	x
HealthNetwork	Yes/ selective	X	X	X	X	X	X	X

available anywhere in the world at any time. The latest trend in health Internet usage shows that the Internet is here to stay and will be a new dimension in our lives (Shepherd et al., 2000).

However, at present not many intelligence features have been implemented for medical Web sites despite the high level of development and availability of the theoretical approaches (such as general portal architectures, artificial neural networks, and fuzzy logic) and growing evidence that there is widespread interest in the Internet medical development area. This absence can lead to high levels of dissatisfaction by users.

The frustration felt by many users is understandable (Finkelstein & Aiken, 2000), as demonstrated by the results of our medical portals comparison. From this evaluation of general medical portals, it is clear why tailored extraction of information is difficult.

The analysis of seven Australian medical portals suggests the following categorization:

1. **Portals with search engines:** Better Health, Health*Insite*, Health*Infonet*, Health*Connect*, and MedAu.
2. **Portals with a knowledge repository:** Rural Health.
3. **Portals with a combination of 1 and 2:** Health Network.
4. **Portal as a gateway:** Rural Health.

From our analysis it can be concluded that from the users' perspective, the most important features to ensure that the information available from the portal is relevant and useful are the **quality of the search engine** and the **level of intelligence** that can be incorporated into the search.

The portal's advanced search facility should ensure that users can retrieve relevant information with minimum clicks and format of input. The search facility covers the "pull" function of the portal; whereas personalization provides a means for pushing the right content depending on some kind of user profile (Probst et al., 2000).

None of the seven Australian community medical portals had sufficient intelligence features to satisfy users with diverse backgrounds and needs. The challenge lies in ensuring that only relevant medical information is retrieved. Portals are complex Web-based systems that can act as gateways to achieve this goal. The use of intelligence built into portal systems, this can help to achieve this goal and facilitate efficient information retrieval.

Portal technology is rapidly improving with the aim to better serve the community. Modern portals are built with intelligent agents, providing artificial intelligence as a part of portal functionality to improve information retrieval (Jafari & Sheehan, 2003). Intelligent portals can improve their performance by learning from experience. Those portals will develop a better understanding of human nature and thus become better tools for meeting users' expectations and information needs.

REFERENCES

Australian Bureau of Statistics. (2002). *Year book Australia* (No. 84). Commonwealth of Australia, Canberra.

Brodie, M., Flournay, R. E., Altman, D. E., Blendon, R. J., Benson, J. M., & Rosenbaum, M. D. (2000). Health information, the Internet and the digital divide. *Health Affairs*, *19*(6), 255-265.

Clarke, I., & Flaherty, B. T. (2002). Web-based B2B portals. *Industrial Marketing Management*, *32*, 15-23.

Collins. H. (2001). *Corporate portals: Revolutionizing information access to increase productivity and drive the bottom line*. AMACON.

Ehrenberger, H. E. (2001). Cancer clinical trial patients in the information age: A pilot study. *Cancer Practice*, *9*(4), 191-197.

Finkelstein, C., & Aiken, P. (2000). *Building corporate portal with XML*. McGraw-Hill.

Harvard Computing Group. (2002). *Portals and business functions: New tools to improve operations*. Retrieved August 2003 from www.Portals Community.com/library/white_papers.cfm

Ho, J., & Tang, R. (2001). Towards an optimal resolution to information overload: An informediary approach. *ACM*.

Jafari, A., & Sheehan, M. (2003). *Designing portals: Opportunities and challenges*. Information Science.

Manaszewicz, R., Fisher, J., Williamson, K., & McKemmish, S. (2002). Breast cancer knowledge online: Supporting the information needs of rural women. Paper presented at *Using IT: Making it happen: Information technology in rural areas conference*, Rockhampton, Australia.

McKemmish, S., Burstein, F., Manaszewicz, R., & Fisher, J. (2002, August). Towards meeting the decision support needs of a community via an "intelligent portal": Breast cancer knowledge online. Paper presented at *Improving quality by lowering barriers: 10th National Health Informatics Conference*, Melbourne, Australia.

Millen, D. R. (2000). Community portals and collective goods: Conversation archives as an information resource. *Proceedings of the 33rd Hawaii International Conference on Systems Sciences*. 1-9.

Murray, M. (2002). An investigation of specifications for migrating to a Web portal framework for the dissemination of health information within a public health network. *Proceedings of the 35th Hawaii International Conference on System Sciences*.

National Archives of Australia. (2000). *AGLS*. Retrieved August 2003 from http://www.naa.gov.au/recordkeeping/gov_online/agls/summary.html

National Health Information Management Advisory Council. (2001). *HealthOnline action plan*. Retrieved September 19, 2003, from http://www.health.gov.au/mediarel/yr2001/dept/mr01018.htm

Negnevitsky, M. (2002). *Artificial intelligence. A guide to intelligent systems*. Addison-Wesley.

Notes, R. (2002, July/August). Free full text; Find articles and MagPortal. *Online Wilton*, *26*, 54-56.

O'Leary, M. (1999, January/February). Portal wars. *Online Wilton, 23,* 77-79.

Probst, G., Raub, S., & Romhardt, K. (2000). *Managing knowledge: Building blocks for success.* New York: Wiley.

Quirk, M. J. (2002). *A guide to defining an enterprise portal project*. Retrieved October 19, 2002, from http://www.OpenDoorTechnologies. com

Rao, S. S. (2001). Portal proliferation: An Indian scenario. *New Library World*, *102*, 325-331.

Shepherd, M., Zitner, D., & Watters, C. (2000). Medical portals: Web-based access to medical information. *Proceedings of the 33rd Hawaii International Conference on System Sciences,* 1-10.

Taylor, H. (2002). Cyberchondriacs update. *Harris Interactive*. Retrieved April 20, 2003, from http://www.harrisinteractive.com/harris_poll/index.asp?PID=229

Tushkar, K. H. (2000). Building enterprise portals: Principles to practice. *ACM*.

Wilson, P. (2002). How to find the good and avoid the bad or ugly: A short guide to tools for rating quality of health information on the Internet. *BMJ*, *324*, 598-602.

This work was previously published inWeb Portals: The New Gateways to Internet Information and Services, edited by A. Tatnall, pp. 270-296, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 5.16
Health Portals and Menu-Driven Identities

Lynette Kvasny
The Pennsylvania State University, USA

Jennifer Warren
The Pennsylvania State University, USA

INTRODUCTION

In this article, we make a case for research which examines the cultural inclusiveness and salience of health portals. We make our case from the standpoint of African-American women. While healthcare should be a ubiquitous social good, health disparities exist among various demographic groups. In fact, health disparities have been placed on the U.S. disease prevention and health promotion agenda. *Healthy People 2010* is an initiative sponsored by policy makers, researchers, medical centers, managed care organizations, and advocacy groups across the country. Although there is no consensus regarding what a health disparity is, sponsors agree that "racial and ethnic minorities experience multiple barriers to accessing healthcare, including not having health insurance, not having a usual source of care, location of providers, lack of transportation, lack of child care, and other factors. A growing body of evidence shows that racial and ethnic disparities in health outcomes, healthcare access, and quality of care exist even when insurance, income, and other access-related factors are controlled."[1]

In addition to healthcare, African American women have less access to the internet. Even at equivalent income levels, African Americans are less likely than either whites or English speaking Hispanics to go online. Demographically, the composition of populations not online has not changed dramatically since 2000. Overall, 60% of the total U.S. population is online with African Americans making up 11% of the total U.S. population, 8% of the online population, and 14% of the offline population. However, when looking at those who are offline, African Americans are more likely than offline whites or Hispanics to believe that they will eventually go online (Lenhart, 2003).

Although online health information is available from multiple sources, we focus solely on those health portals sponsored by the U.S. government. We made this choice based upon some early interviews with physicians and managers at a healthcare facility which serves predominantly African American clients. We learned that most clients exhibited a low degree of trust in information provided by pharmaceutical companies and other sources which seemed too commercial. Instead, clients searched for information from recognizable sources, and tended to use portals and search pages like Yahoo and Google. We found that portals sponsored by U.S. government agencies were received positively by clients. Also, portals like healthfinder.gov and cdc.gov are highly regarded by the Medical Library Association[2]. Moreover, the government is entrusted to uphold values of democracy and social justice therefore the health information that they provide should be accessible to a demographically diverse audience.

To gain insights into the cultural inclusiveness and salience of health portals, we use Nakumura's notion of menu-driven identities. For Nakumara (2002), the internet is a discursive place in which identity is enacted. She uses the term "menu-driven identities" to signify the ways in which content providers represent identities through the design of the interface and the personalization of content, and users perform their identity as they engage with the content. In what follows, we discuss health disparities and the promise of the internet in redressing inequities. Next, we further explain the ways in which users perform identity and health portals represent identities. We do this by theorizing about the health portals as mediating two-way communication between users and information providers. We conclude with directions for future research.

BACKGROUND

Health portals hold promise as an informational source for improving the health of historically underserved populations. This promise is extremely exciting given the state of health disparities in the U.S. We know from prior studies that health provider bias, stereotyping, prejudice, and clinical uncertainty may contribute to disparities along the lines of gender, class, race, and ethnicity (Balsa & McGuire, 2003). For instance, in a study by Bird and Bogart (2001), 63% of survey participants indicated that they had experienced discrimination in their interactions with their healthcare provider because of their race or ethnicity. Similarly, African Americans interviewees reported perceived discriminatory experiences such as inferior treatment, negative attitudes, being treated as if they were unintelligent, being ignored, inappropriate allegations, and racist remarks (Hobson, 2001). These negative experiences may profoundly impact attitudes towards receiving care, and willingness to comply with physician recommendations. For example, Hobson (2001) found that nearly 27% of African American survey respondents reported that, as a result of a discriminatory event, they were more hesitant to seek health services. Others avoided the healthcare facility (25.6%), avoided the provider (23.1%), avoided the personnel involved (10.3%), stopped using specific services (15.4%), or used service less frequently (7.7%).

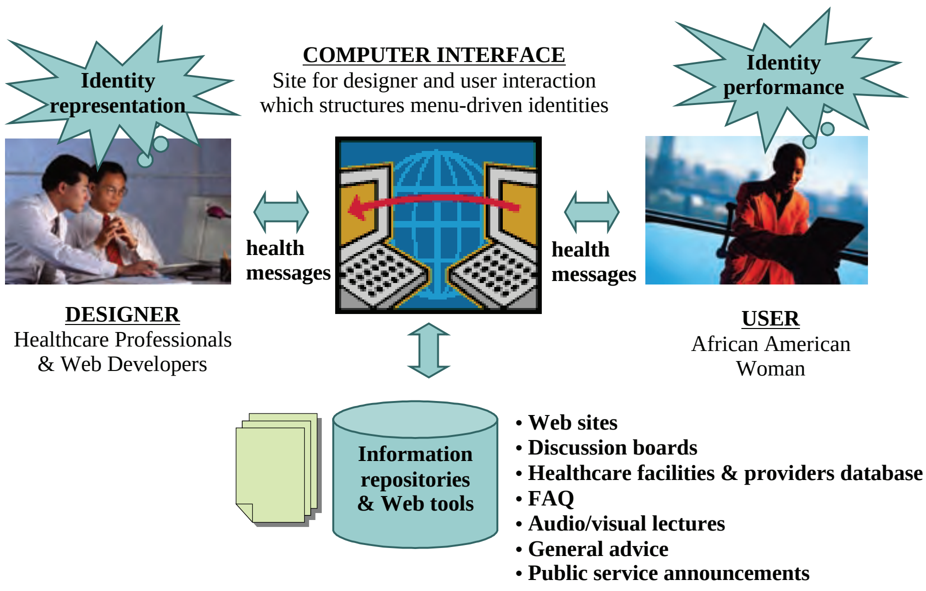
Computer mediated communication may help minorities, women and other underserved groups to receive healthcare information in a more hospitable climate. The popularity of the internet as a medium for health communication is evidenced in two ways. First, the number of health-related Web sites has dramatically increased from a mere 15,000 sites in 1999 (Rice, 2001) to 100,000 as of 2003 (Cates, 2003). Secondly, although these sources vary in quality and relevance, the number of people seeking online health information rose to 97 million in 2001 from 60 million in 1999 (Rimal & Adkins, 2003). In a 2002 national survey (see Figure 2), researchers found that 73 million people in the U.S. or 62% of internet users have gone online to search for health information. On a typical day, about 6 million Americans go online for medical advice. This exceeds the number of Americans who actually visit health professionals according to figures provided by the American Medical Association (Fox & Rainie, 2002). And while 42% of Americans say they don't use the internet, many of them either have been internet users at one time or have a once-removed relationship with the internet through family or household members. In fact, some exploit workarounds that allow them to use the internet by having email sent and received by online family members and by having others in their home do online searches for information they want (Lenhart, 2003). Women are more likely than men to say their latest search was at least in part for someone else—62% compared to 50% of men. Women are also more likely than men to seek healthcare and health information both online and offline (Fox & Fallows, 2003).

Identity

As an increasing number of Americans obtain health-related information online, it is important to consider that the internet is not race and gender neutral. Rather, it is a discursive space in which identities can be represented, performed, swapped, bought, sold, and stolen. Users can create profiles to personalize their experience, and create avatars which serve as visual representations of the body in cyberspace. But while spaces for fluid subjectivity abound, the internet often fails to accommodate minority cultural identities (Kolko, Nakamura, & Rodman, 2000; Kvasny, forthcoming).

Identities are inextricable from communication and are enacted in messages (Hecht, 1993). These enactments transmit and exchange values, beliefs, and norms, which may or may not affirm individuals' or groups' understandings of their own identities (Jackson, Warren, Pitts, & Wilson, under review). Identities also act as interpretative frames in the communication process (Hecht, 1993). Messages are filtered through and made sense of in relation to how individuals perceive themselves. If health messages communicate an identity, which is in conflict with how African American women perceive themselves, then the information may be viewed as unusable and we have done little to combat health disparities. As

Figure 1. Menu-driven identities



with all communication, messages that diverge from the identities of minority populations are unlikely to prove effective. Hence, health information must be situated within the target audiences' sphere of experiences and understandings, or they may go unheeded.

When users search and consume information online, they perform identity. In fact, identity is the first thing that you do (create a profile or user account) before you can perform any activity. Identity performance is often a practical necessity constructed by designers who create the interface. Nakumara (2002) uses the term "architecture of belief" to signify how designers, through their choice of keywords, images, and use of language, create interfaces which represent the identities of some idealized user population(s). The interface reflects the cultural imagination of the designer, and performs familiar versions of race, gender, sexual orientation, and class. The relation between the user and the interface has been termed "menu-driven identity" performance and representation (see Figure 1) (Nakumara, 2002).

Thus, contrary to the popular notion that physical characteristics are erased online, the fact that Web users must reveal aspects of their identity suggests that bodies are often "outed" in cyberspace. For instance, we often have to define our race, gender, age, weight, marital status, and other identity factors when using health portals. In a cultural sphere such as health, these aspects of identity are crucially important factors to be considered, not superficial characteristics to be erased. Consequently, identity is not entirely fluid and physical bodies remain important even though they are largely hidden.

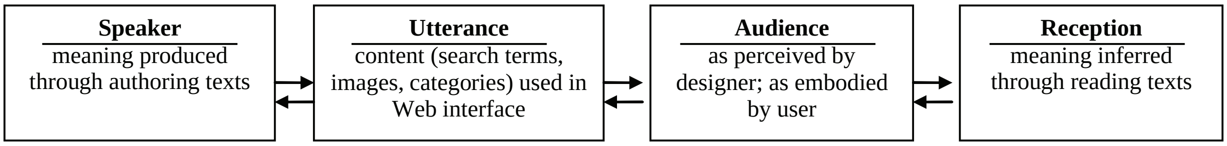
Health Portals

Health portals are made spaces in which identities are enacted through computer mediating conversations between users and designers. The interface serves as a site for the production, representation, distribution, and reception of texts (health messages) which convey meanings that affects relations of power. The distinctive rhetorical conditions of the speaker, utterance, audience, and reception are created as the designer and user co-create a communicative situation in cyberspace (see Figure 2).

In what follows, we demonstrate how menu-driven identities are produced on two government Web sites. The first example is taken from the U.S. Department of Health and Human Services healthfinder portal—http://www.healthfinder.gov/justforyou/ (see Figure 3). The navigation scheme tends to reflect the unspoken biases and categories imposed by the more privileged actor (the government sponsor) in the communicative situation. Users must select an identity which is limited to two gender categories, four age categories, four ethnic/racial categories, and five roles.

For African American women, predefined categories force the performance of race and gender in ways that marginalize, and in some cases deny, their existence. This occurs when there is limited space for identity expression because the categories reproduce the limited number of choices based on historical labels and ideologies around race, gender, age, and role. Health portals, therefore, become another discursive field in which African American women are rendered invisible because they are assumed to be only African American or female—multiple selections are not allowed because there are no categories which capture both race and gender. Other underserved groups, such as gays and lesbians, are rendered completely invisible because they have no category. White is also omitted from the choices, but this is because whiteness is assumed as the default category and simply goes without saying. These familiar versions of identity are scripted and ascribed by designers when they create interfaces based upon these types of simplified categorizations. Notice how the author "selected very specific information from our library so that it is easy to zero in on health topics of special interest to you. Just choose one of the special groups below."

Figure 2. Communicative situations



This example demonstrates how Web portals may serve as platforms for reproducing simplified discourse around difference. Category schemes are not simply passive tools through which labeling takes place, but rather are the outcomes of practices of meaning making. Designers of classification schemes constantly have to decide what categories are important, and in doing so, they develop an economy of knowledge that articulates omissions and inclusions, and ensures that only relevant features are classified (Bowker & Star, 1999). In doing so, health portals limit choices and the full participation of people who exist at the margins because users can only take the paths prescribed by the interface.

A second Web site taken from the Center for Disease Control, Office of Minority Health—http://www.cdc.gov/omh/Populations/populations.htm (see Figure 4)—demonstrates the instability of categories. For instance, race is an unstable signifier, which is socially constructed in dissimilar ways in various cultures. Blacks in the U.S. include all (including mixed race) people who trace their ancestry to Africa, but Blacks in South Africa don't include people of mixed African and European ancestry. In Britain, Black includes people with ancestry to non-African parts of the former empire such as Pakistan and China (Kolko, Nakamura, & Rodman, 2000). Racial signifiers are also unstable because they

Figure 3. Healthfinder3

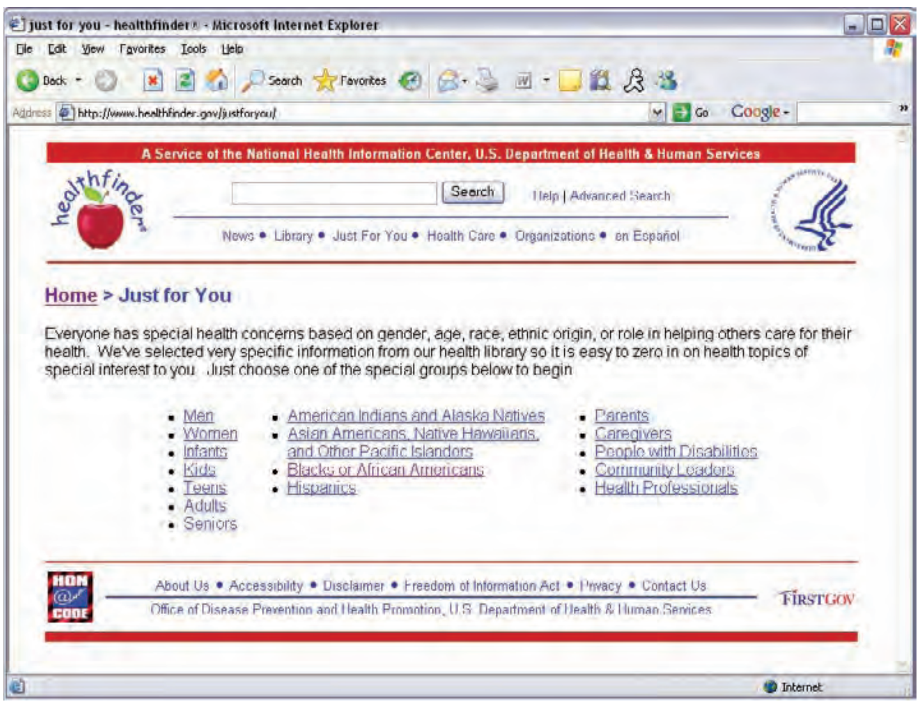
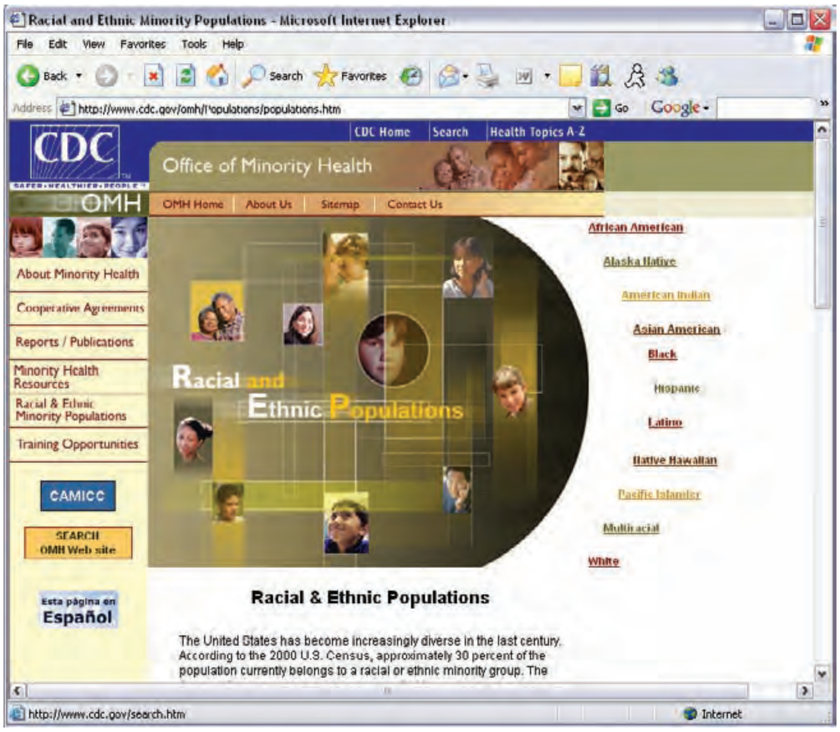


Figure 4. The center for disease control



change over time. In the U.S., for example, people of African ancestry have been labeled Negro, Black, and African American.

In this health portal, the categories have been expanded to include white. Multicultural Americans are included even though it is difficult to make generalizations about health conditions because little research exists about this group. On the surface, it looks as though African American is treated as an ethnic marker distinct from Black, a racial category which would include people throughout the African Diaspora. However, when selected, each category is linked to the same Web page. Hispanic and Latino are also listed as two distinct categories which, when clicked, send the user to a single Web page.

FUTURE TRENDS

Given the increasingly diverse populations of internet users and the growing usage of eHealth resources, we pose several important questions for future research:

1. How are underserved groups socially constructed on health portals?
2. What do these social constructions suggest about inclusiveness and health disparities?
3. What types of information do members of underserved groups seek, and why?
4. How well do existing government health portals serve the needs of underserved population?
5. How might we design more inclusive information resources?

We also suggest that existing health portals could better serve the needs of underserved groups. While we cannot account for the accuracy of the information provided through these portals, we did observe a wealth of information that is pertinent to various demographic groups.

The challenge that we raise in this article is how best to tailor user interfaces to improve the online experiences of culturally diverse users. A tailored user interface would include:

- A mixture of media formats including texts, audio, video, and slide presentations. Images should be representative of diverse people. Single page checklists, fact sheets, and brochures may be especially useful for women who are obtaining information for other family members and friends.
- Texts that are comprehensible by low-literacy users and availability in languages other than English.
- The ability to select multiple identities. This could be done with checklists or a menu structure that enables users to drill down through several demographic categories. For instance, a middle aged African American women could select from the *gender, race/ethnicity, age* categories to refine her information.
- Normative categories such as white and heterosexual should be listed explicitly.
- The ability to declare identities in ways that go beyond demographics. For instance status such as smoker, diabetic, HIV positive, and cancer survivor are important components of identity that influence health care needs and outcomes
- Spaces such as chat rooms and forums for users to act as speakers and authors in the communicative situation.

CONCLUSION

The internet offers a space where African American women and other underserved groups can become empowered health consumers who access health information on their own terms (Ferguson, 1997). However, it is important to understand the extent to which online health information is inclusive of diverse users. Providing computers and internet access, and showing underserved groups the value proposition of online health resources are simply not enough. We must also consider the cultural salience of content and inclusiveness of the interface design.

REFERENCES

Balsa A. I., & McGuire T. G. (2003). Prejudice, clinical uncertainty, and stereotyping as sources of health disparities. *Journal of Health Economics*, *22*, 89-116.

Bird S. T., & Bogart L. M. (2001). Perceived race-based and socioeconomic status (SES)-based discrimination in interactions with health care providers. *Ethnicity and Disease*, *11*, 554-563.

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.

Cates, L. K. (2003). Meet online ethical expectations: New internet guidelines assist in monitoring information dissemination. *Advance News Magazine*. Retrieved November 19, 2003, from http://www.advanceforaud.com/common/editorial/editorial.aspx?CC=10095&CP=1

Ferguson, T. (1997). Health online and the empowered medical consumer. *Journal on Quality Improvement*, *23*(5), 251-257.

Fox, S., & Fallow, D. (2003, July 16). *Internet health resources: Health searches and email have become more commonplace, but there is room for improvement in searches and overall internet access.* Washington, DC: Pew Internet and American Life. Retrieved January 2005, from http://www.pewinternet.org/PPF/r/95/report_display.asp

Fox, S., & Rainie, L. (2002, May 5). *Vital decisions: How internet users decide what information to trust when they or their loved ones are sick*. Report. Washington, DC: Pew Internet and American

Life. Retrieved January 2005, from http://www.pewinternet.org/PPF/r/59/report_display.asp

Hecht, M. L. (1993). 2002—A research odyssey: Toward the development of a communication theory of identity. *Communication Monographs*, *60*, 76-82.

Hecht, M. L., Jackson II, R. L., & Ribeau, S. A. (2003). *African American communication: Exploring identity and culture*. Mahwah, NJ: Lawrence Earlbaum Associates.

Hecht, M. L., Warren, J. R., Jung, E., & Krieger, J. (2004). The communication theory of identity: Development, theoretical perspective, and future directions. In W. Gudykunst (Ed.), *Theorizing about intercultural communication* (pp. 257-278). Thousand Oaks, CA: Sage Publications.

Hobson, W. D. (2001) *Racial discrimination in a health care interview project*. Special Report. Seattle, WA: Public Health Seattle and King County.

Jackson, R. L., Warren, J. R., Pitts, M. J., & Wilson, K. B. (under review). It is not my responsibility to teach culture! White graduate teaching assistants negotiating identity and pedagogy. In L. Cooks (Ed.), *Whiteness, pedagogy, and performance*. Landham, MD: Lexington Books.

Kolko, B., Nakamura, L., & Rodman, G. (2000). *Race in Cyberspace*. New York: Routledge.

Kvasny, L. (forthcoming). *Understanding the digital divide from the standpoint of the "other"*. The Data Base Advances in Information Systems.

Lenhart, A. (2003). *The ever-shifting internet population: A new look at Internet access and the digital divide*. Washington, DC: Pew Internet and American Life Project. Retrieved January 2005, from http://www.pewinternet.org/pdfs/PIP_Shifting_Net_Pop_Report.pdf

Nakumara, L. (2002). *Cybertypes: Race, ethnicity, and identity on the Internet*. New York: Routledge.

Rice, R. E. (2001). The Internet and health communication. In R. E. Rice &. J. E. Katz (Eds.), *The Internet and health communication* (pp. 5-46). Thousand Oaks, CA: Sage Publishers.

Rimal, R. N., & Adkins, D. A. (2003). Using computers to narrowcast health messages: The role of audience segmentation, targeting, and tailoring in health promotion. In T. L. Thompson, A. M. Dorsey, K. I. Miller, & R. Parrott (Eds.), *Handbook of health communication* (pp. 497-514). Mahwah, NJ: Lawrence Erlbaum and Associates.

KEY TERMS

Digital Divide: The term "digital divide" describes the fact that there exist people who do and people who don't have access to—and the capability to use—modern information technology, such as the telephone, television, or the internet. Access to these resources follows along demographic lines such as gender, income, race, ethnicity, geography, and age.

Health Disparity: Some demographic groups, such as racial and ethnic minorities, experience multiple barriers to accessing healthcare, including not having health insurance, not having a usual source of care, lack of transportation, lack of childcare. A growing body of evidence shows that health disparities persist even when insurance, income, and other access-related factors are controlled.

Health Portal: A Web site often sponsored by a large institution, which provides extensive information and contains links to a wide range of health and medical information on the internet.

Identity: Hecht (1993) identifies four frames through which identity is communicated. *Personal identity* involves self-concept, and develops through socially ascribed meanings and behaviors which are learned as one is socialized in a society (e.g., Black women stereotyped as mammy—the nurturer; jezebel—the seductress; sapphire—the wisecracking emasculating women; the welfare queen—the lazy, economically unstable mother of many bad kids). *Enacted identity* focuses on how messages express identity (e.g., I'm a black woman). *Relational identities* are those formed through one's relationships (e.g., I'm a mother). *Communal identities* are those shared by groups of people in some particular community (e.g., I belong to the Penn State community).

Menu-Driven Identity: This term signifies the ways in which content providers represent identities through the design of selection-oriented interfaces that are used to personalize content. Conversely, the term signifies the ways in which users perform their identity as they navigate within the choices that are both enabled and constrained by the interface.

ENDNOTES

1. http://www.healthypeople.gov/
2. http://www.mlanet.org/resources/medspeak/topten.html
3. This Web page in this screenshot served as the healthfinder start page when we conducted our analysis. This healthfinder Web site has been vastly improved, and the Web page used in our analysis is still accessible by clicking "Just for You" on the current start page.

Chapter 5.17
Assisting Users Seeking Medical Information through Government Portals

Jane Moon
Monash University, Australia

INTRODUCTION

There has been an explosion in the number of different types of portals in the last decade, and at the same time there has been a lot of confusion with them, especially in relation to the enormous number of portals and their differences from Web sites or Web-pages. This coincides with increased use by consumers seeking medical information on the Internet, and with the important role played by medical portals for evidence based medicine.

This article explores current portal technology available from an evaluation of market leaders in the industry and identifies important functional components that are necessary in building an intelligent portal to assist users seeking information on the Internet.

The emphasis will be on government to consumer portals (G2C) and uses two reputable government portals Betterhealth and Healthinsite as examples to discuss issues involved with those.

BACKGROUND: INCREASED USE OF PORTALS FOR SEEKING MEDICAL INFORMATION

Reliance on portals for medical information is high, and recent statistics show that this trend is growing. At the same time medical information is widely dispersed and information retrieval is inadequate (Shepard, Zitner, & Watters, 2000). There is an urgent need to develop portals that help users to retrieve quality information.

The National Health Medicine Advisory Council reported that better health outcomes are predicted when users are better informed, suggesting the need for digitization of health services (Bodenheimer, Lorig, Holman, & Grumbach, 2002; Kennedy, 2002; National Health Information Management Advisory Council, 2001). However this is yet to be confirmed, as Internet content has not yet proven to be satisfactory. There is a lot of concern about the materials found on the Internet (Ciolek, 1997, Moon, 2005).

Modern portals are built with the aim of better catering for the different needs of users. Portal technology is improving and features such as knowledge management, content management, and search engines, along with effective Web site design, help users to find information more effectively, and yet portals remain ineffective in information retrieval (Clarke & Flaherty, 2003; Elias & Ghaziri, 2004; Quirk, 2001; Rao, 2001). An effective evaluation tool to assess the validity and effectiveness of the portal is quintessential in assisting both users and portal builders.

PORTALS

Portal terminology has been loosely used in the industry and the term "portal" means different things to different people. Some argue that the word "portal" should be used as an adjective—"portal framework," "portal structure," and "portal architecture"—rather than as a noun as it is commonly used (Roth, 2003). Others are confused by the difference between a Web-page or Web site and a portal. The same confusion is applied to health portals vs. medical portals. The following sections seek to clarify these terms.

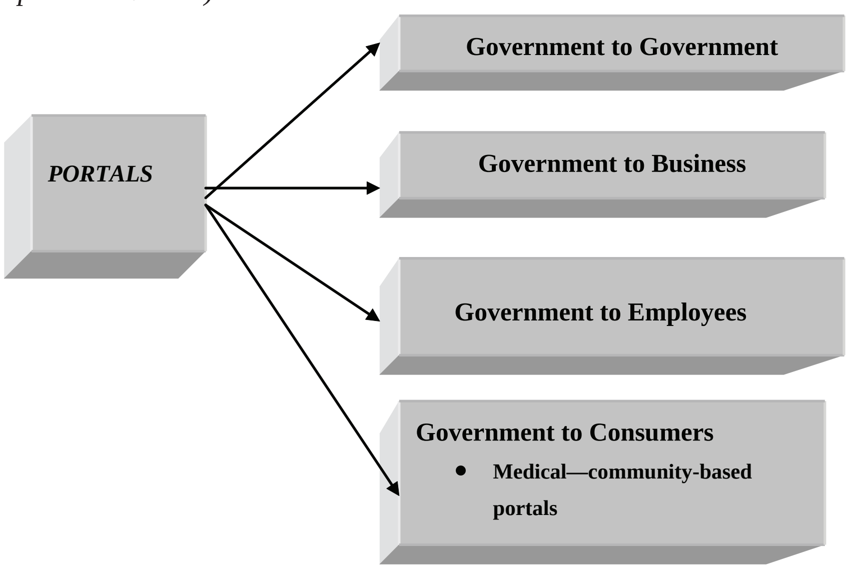
Definition of Portals

For the purpose of this research, the working definition of a portal is an Internet-based information system providing uniform access to different sources of information in an enterprise and maintaining dynamic links to information resources (Moon & Burstein, 2004). A portal supports communication within the enterprise, and connects people with information and applications they need for performing tasks. Unlike a conventional Web site, portals should support both push (subscription) and pull (search) functions in assisting users to gain access to essential contextual information (Probst, Raub, & Romhardt, 2002; Shepard et al., 2000).

Portals vs. Web sites

The word "portal" is often confused with Web sites. The difference between Web sites and portals is that the former are static and the latter are

Figure 1. Government portal market segment diagram (Adapted from Collins, 2003; Clarke and Flaherty, 2003; Shepard et al. 2000)



dynamic. Portals are Web-based, have dynamic links to the information resources, effective information and document managers and connect people with information. A portal is dynamic, can have personalisation features built into it and has the ability to bring together "business professionals involved in their area of interests" (Hazra, 2002, p. 623). A portal is a mechanism to deliver content, using a Web browser. It runs on an application server, and can integrate with knowledge management and workflow. It must unify a large range of applications, services and information flows.

A Web site is static, and is a collection of Web pages relating to a given subject, including and beginning with the home page; for example, a company Web site. A portal is often confused with a Web server where many different Web sites can be stored on a single machine. Companies generally have many Web sites that are stored by commercial Web servers in different locations, sometimes worldwide.

For the purposes of this chapter, any Web sites which have at least one of these listed features is referred to as a "portal."

CLASSIFICATION OF PORTALS AND FUNCTIONALITIES

There was an explosion in the number of different portals in the late 1990s. A great multitude of terms are used to describe different types of portals. Some examples are: EIP (enterprise information portal), corporate portal, community portal, employee portal, business intelligence portal, horizontal portal, vertical portal, public portal and private portal, to name just a few. These names are dependent on who the stakeholders are (Collins, 2003; Portals Community, 2004).

A government portals can be categorized broadly into four types depending on the stakeholders (government to government, government to business, government to employees, and government to consumers).

- G2G (Government to Government): Supports work flow between government sectors and assisting and collaborating work, not accessible to consumers
- G2B (Government to Business): Supports business flow, activities, suppliers, and partners for distribution and supply chain management
- G2E (Government to Employee): Supports the access and availability of personalised information to employees
- G2C (Government to Consumer): Supports ordering, service, support, and billing for customers

All of the portals in Figure 1 fall in to the function of the following three types:

- Knowledge Portal: Has the ability to combine information, but provides no commercial transaction
- Decision-Based Portal: Collection of information from a wide variety of sources, which can be structured or unstructured
- Process-Based Portal: Changing focus to delivery, user-centric, collecting information to be distributed to future processes such as business to consumers (B2C) or business to business (B2B), where transactions are involved

Recent development has been on process-based portals, those that promise to deliver information based on user needs. All of the previously mentioned portals are rapidly evolving to support changes in the information needs of customers, businesses, suppliers, and individuals. The names of portals vary according to services and the users they serve (Wege, 2002).

Portal Server Technology

Portals integrate diverse interactions: access to information, applications, and people. The

Table 1. Medical portals comparison

Australian Medical Portal	Search Engine	Spell Check	Sounds Like Index	Parsing	Ontology	*Personalization*	Thesaurus	Dec. Facilitis
BetterHealth Channel	Google	X	X	X	X	X	X	X
Health*Insite*	Yahoo	X	X	X	X	Yes	Yes	X
Rural Health	X	X	X	X	X	X	X	X
Health*InfoNet*	Yes	X	X	X	X	X	X	X
HealthConnect	Yes	X	X	X	X	X	X	X
Medicine Australia (MedAu)	Yes	X	X	X	X	X	X	x
HealthNetwork	Yes/ selective	X	X	X	X	X	X	X

infrastructure, which is a connection between hardware and software, is critical for the deployment of a portal. The portal server technology items needed are:

- Hosting Service Provider: The hosting service provider can be internal (the IT department within the organization), or external (an application service provider (ASP)) or an off-site hosting services vendor. Hosting service providers are responsible for maintaining the portal services. The tasks involve systems management and site administration
- Platforms: Several types of platforms have evolved. The most pervasive platform is the Operating System, especially Microsoft Windows OS, Windows 2000, and Windows NT. Others include Macintosh (for client layer), Unix, and its variants (Sun Solaris, Linux, IBM AIX, IBM OS/390) for the servers also. Some traditional main frames are OS/390, OS/400, etc
- Interoperability: Interoperability is crucial to the success of the portal. A typical portal would integrate several Web based applications (Microsoft and Sun architecture), XML, and database applications as well as desktop applications (word processors and spread sheets). A good portal would consolidate all these applications into a single organized desktop
- Personalization: Portal personalization can be made at different levels. An individual can have personal settings; a group of people sharing the same function can have group settings. An organization can choose its own settings. These features allow the portal to be customized according to needs
- Single Sign-On (SSO): Signing on only once to get access to all the facilities portals offer is crucial. This is a significant time saving in logging on to different sites, and a significant reduction in training and also time saving for administrators in re-issuing passwords for those who have forgotten them (Collins, 2003, p. 39; Portals Community, 2002, p. 12; Wege, 2002).

Functional Components of Medical Portals

Portals should provide wide functionality to allow users to find information, and to manage, categorize, and use applications. They should ensure that the features needed by employees in the organization are met. The implementation of functionalities will vary depending on the nature of the business. Some of the functionalities that are available and can be implemented in medical portals are as follows:

- Browse/Navigate Documents: Allowing users to search for and locate information manually by navigating a directory structure
- Collaboration: An important aspect of a portal and a powerful tool. It includes instant messages such as "chat," document sharing, video conferencing, virtual conferencing, and discussion forums
- Content Management: A process of approving, authoring, delivering, maintaining, and publishing content integrated with or accessed from a portal or other Web site. This can be implemented as a separate component with a common database accessible through the portal
- Directory: A directory within the portal's enterprise taxonomy. It is a collection of data structured into a hierarchy of categories
- Document Management: Similar to content management but dealing mainly with control and management of an enterprise's documents. It involves managing electronic files including scanned images of pictures and documents
- Personalization: Portal personalization can be made at different levels. An individual can have a personal setting; a group of people sharing the same function can have group settings. An organization can set up its own setting. This feature allows a portal to be customized according to needs
- Search Engine: The ability to search is the fundamental implementation of a portal. It allows users to browse the content, retrieve information on a content basis, and link to other repositories for information
- Subscribe/What's New?: Allows users to register an interest in a particular aspect of the portal. Portals then notify the user when any changes in the content occur
- Taxonomy: A classification scheme to organize information. Additional functionalities, such as metadata added to taxonomy, could help to organize documents into different categories that could make the information easy to browse, search, or navigate ideally.
- Online Community: Opportunity for users with similar interests to discuss their medical conditions and experiences. This could range from sharing experiences of treatment to discussing medical conditions and alternative medicine (Collins, 2003; Millen, 2000; Moon & Burstein, 2004; Tushkar, 2000).

Table 2. Evaluation of portal vendors on intelligence features

Vendors	CM	KM	Navigation and Search	Personalization	Community Definition	AI
ArtTechnology Group	√	√	√	√	√	√
Autonomy	√	√	√	√	√	√
BEA systems	√	√	√	√	√	
BroadVision	√	√	√	√	√	
Citrix	√	√	√	√	√	
Computer Associates	√	√	√	√	√	
Corechange	√	√	√	√	√	
Epicentric	√		√	√	√	
IBM	√	√	√	√	√	
Oracle	√	√	√	√	√	
PeopleSoft	√	√	√	√	√	
Plumtree	√	√	√	√	√	
SAP	√	√	√	√	√	
Sun Microsystems	√	√	√	√	√	
Sybase	√	√	√	√	√	
TIBCO	√	√	√	√	√	
Vignette	√	√	√	√	√	
WebMethods	√	√	√	√	√	

THE IMPORTANCE OF INTELLIGENT MEDICAL PORTALS

The term "intelligent medical portal" refers to a Web gateway environment that allows users with varied medical interests and diversified backgrounds to access medical resources and information to support their decision-making processes, which often involve critical medical problems.

Current medical portals do not provide for people with different educational backgrounds, age, ethnicity, location (urban or rural), and stages of disease.

For example, there is no specific information for a 63-year-old female with early breast cancer or for a 35-year-old mother experiencing advanced breast cancer living in a rural area. The hyperlinks regarding breast cancer seem to employ a "one-size-fits-all" approach (McKemmish, Burstein, Manaszewicz, & Fisher, 2002).

In the case of breast cancer, a third of sufferers in Australia live rurally. For them the quality of information is absolutely critical, often life-threatening (Manaszewicz, Fisher, Williamson, & McKemmish, 2002). The development of an intelligent portal is to some extent a response to this problem. A major concern is that there are no real safeguards and standards concerning the "quality, precision, trustworthiness, currency, and authorship of this information" (Ciolek, 1997, p. 5).

How can a portal be made to behave intelligently so that users can easily gain access to vastly distributed medical information? McKemmish et al. (2002) suggests an intelligent portal "will improve the structure and manage the quantity of information presented" (p. 5). Indeed an intelligent portal should provide information that is high-quality, relevant, and adequate. It should provide for the needs of users by helping them to solve issues that are often critical to life. With the use of intelligent portal technology, combining knowledge management issues of personalization and decision support, the needs of Internet users of medical portals could be improved.

Intelligence Features

Previous research has established intelligence features that are available technically as follows (Moon & Burstein, 2004):

- Search engine
- Spell checking
- Sounds-like index
- Parsing
- Ontology
- Personalization
- Thesaurus

As medical jargon, treatment and drug names are often very difficult to remember and spell, it seems that it is essential to implement the available technology previously mentioned as much as possible to assist users in their search for medical information.

Australian Medical Portals

Despite the surge of interest and importance of the Internet for providing users with medical information, users were dissatisfied with the results. The following table illustrates the results of intelligence features against seven Australian medical portals.

All the portals had search engines. In general, most of them lacked in providing any form of intelligence. HealthInsite is the only portal that provided "personalization" and a "Thesaurus."

Evaluation of Portal Vendors on Intelligence Features

For the purpose of this research, it was important to look for what was currently available on the market in terms of functionalities for medical portals that make them more intelligent.

To do this it was necessary to explore various vendors of portal software and find the functionality available as well as to compare vendor's views with users' views to check the validity of vendor's claims. Though the technology and the concept of the portal have been around for more than 10 years, the idea of implementing portal solutions as a whole package is still fairly new. Most of the portal solutions that are available are developed to solve commercial problems, and are expensive.

After the evaluation of leading providers of portal software, eighteen high-profile providers were chosen for the analysis of intelligence features.

All the leading portal software packages provided most of the intelligence features. Only two vendors, Autonomy and ATG, provided AI, in particular artificial neural networks, for their search. ATG is in partnership with Autonomy for their search engine.

FUTURE RESEARCH

The results of this study confirm wide use of Government portals in searching for medical information. However, the study indicates that Australian medical portals have a limited intelligence features and are not meeting the needs of their users despite of the available technology. Further research should be undertaken in:

- Intelligent Medical Portals: How intelligence features can be effective in implementing medical portals.
- Implementing Intelligent Agents: Implementing intelligent agents such as "digital secretary," which acts as personal secretary, or "digital sister-in-law," which helps users with movie suggestions, make portals smarter (Jafari & Sheehan, 2003).
- User Profiles: Better identification of common profiles of information usage and matching these with correct content by employing the best available technology to improve user acceptance of portals.

CONCLUSION

The research reports that current Australian medical portals are not meeting their users' needs. The research sought to discover the ways to improve the portals technology to assist the users in their search for medical information. From the literature review and analysts reports the definition of portals and of the many different types of portals are identified.

The analysis of portal vendors amongst market leaders, the technical functionality available for portals was identified. All the portal vendors had most of the listed intelligent features with the exception of artificial intelligence (AI). Autonomy was the only vendor that provided with AI that of Artificial neural network.

An intelligent medical portal is one way of meeting consumers' needs. Understanding human search behaviour can help to improve the technology if this understanding is integrated into interface design. Available portal technology such as ANN and fuzzy logic can help to improve information retrieval. Further incorporating intelligence features and portal functionalities such as CM, DM and KM can add value to the service and thereby increase the effectiveness of portals.

REFERENCES

Bodenheimer, T., Lorig, K., Holman, H., & Grumbach, K. (2002). Patient self-management of chronic disease in primary care. JAMA, 288(9), 2469-2475.

Ciolek, M. T. (1997). Today's WWW, tomorrow's MMM: The specter of multi-media mediocrity. Educom Review, 32.

Clarke, I., & Flaherty, B. T. (2003). Web-based B2B portals. Industrial Marketing Management, 32, 15-23.

Collins, H. (2001). Corporate portals: Revolutionising information access to increase productivity and drive the bottom line. New York: AMACON.

Collins, H. (2003). Enterprise knowledge portals. New York: AMACON.

Dewan, R., Freimer, M., & Seidman, A. (1999, January 5-8). Portal kombat: The battle between Web pages to become the point of entry to the world wide Web. Proceedings of the 32nd Hawaii International Conference on Systems Sciences, Maui.

Elias, M. A., & Ghaziri, H. M. (2004). Knowledge management tools and knowledge portals. Upper Saddle River, NJ: Prentice Hall.

Firestone, J. M. (2003). Organizational intelligence—the ability of an organization to adapt to its environment. Amsterdam, Sydney: Butterworth-Heinemann.

Golier Incorporated. (1980). The new book of knowledge, Dictionary (Vol. II), p. 1021. Danbury, CT: Golier.

Hazra, T. K. (2002). Building enterprise portals: Principles to practice. ACM, 19-25.

Jafari, A., & Sheehan, M. (2003). Designing portals: Opportunities and challenges. Hershey, PA: Idea Group Publishing.

Kennedy, I. (2002). Patients are experts in their own field. British Medical Journal, 326, 1276.

McKemmish, S., Burstein, F., Manaszewicz, R., & Fisher, J. (2002). Towards meeting the decision support needs of a community via an "Intelligent portal": Breast cancer knowledge online. Retrieved May 21, 2003, from http://sims.monash.edu.au/research/eirg/

Manaszewicz, R., Fisher, J., Williamson, K., & McKemmish, S. (2002). Breast cancer knowledge online: Supporting the information needs of rural women. Paper presented at Using IT: Making it Happen: Information Technology in Rural Areas Conference, Rockhampton, Australia.

Millen, D. R. (2000). Community portals and collective goods: Conversation archives as an information resource. Proceedings of the 33rd Hawaii International Conference on Systems, IEEE (pp. 1-9).

Moon, J. (2005). Discussing health issues on the Internet. In S. Dasgupta (Ed.), Encyclopedia of virtual communities. Hershey, PA: Idea Group Reference.

Moon, J., & Burstein, F. (2004). Intelligent portals for supporting medical information needs. In A. Tatnall (Ed.), Web portals: The new gateways to internet information and services. Hershey, PA: Idea Group Publishing.

National Health Information Management Advisory Council. (2001). Health online action plan. Retrieved September 19, 2003, from http://www.health.gov.au/mediarel/yr2001/dept/mr01018.htm.

Notess, R. (2002). Free full text; Find articles and MagPortal. Online Wilton, 26, 54-56.

O'Leary, M. (1999). Portal wars. Online Wilton, 23, 77-79.

Portals Community. (2002). Portals and Business Functions: New tools to improve operations. By Harvard computing group. Retrieved October 19, 2002, from http://www.PortalsCommunity.com/library/white_papers.cfm

Portals Community. (2004). The definitive enterprise portals resource. Retrieved January 14, 2004, from http://www.PortalsCommunity.com/about_categories.cfm

Probst, G., Raub, S., & Romhardt, K. (2000). Managing knowledge: Building blocks for success. New York: Wiley and Sons.

Quirk, M. J. (2001). A guide to defining an enterprise portal project. Retrieved October 19, 2002, from http://www.OpenDoorTechnologies.ccom

Rao, S. S. (2001). Portal proliferation: And Indian scenario. New Library World, 102, 325-331.

Roth, C. (2003). Mr. Grammar Says … "Portal" is an adjective, not a noun. Retrieved January 13, 2004, from http://www.metagroup.com/us/displayArticle.do?oid= 43637

Shepard, M., Zitner, D., & Watters, C. (2000, January 4-7). Medical portals: Web-based access to medical infomation. Proceedings of the 33rd Hawaii International Conference on System Sciences, IEEE, Maui (pp. 1-10).

Tushkar, K. H. (2000, May). Building enterprise portals: Principles to practice. ACM, 19-25.

Wege, C. (2002). Portal server technology. IEEE Internet Computing, 6(3), 73-79.

KEY TERMS

Content Management: A process of approving, authoring, delivering, maintaining, and publishing content integrated within or accessed from a portal or other Web site. It can be implemented as a separate component with a common database accessible through the portal.

Health Portals: Health portals on the other hand, are broad; they cover a very wide range of health and medical topics, anything relating to health (e.g., shampoo products, cosmetics, diets, medicine, and alternative therapy. This research is concerned with the narrower cencept of medical portals.

Intelligent Medical Portal: Refers to a Web gateway environment that allows users with varied medical interests and diversified backgrounds to access medical resources and information to support their decision-making processes, which often involve critical medical problems.

Knowledge Management: This can be seen as a mechanism for capturing "know-how" and "know-why" in a knowledge repository.

Medical Portals: Medical portals provide information such as causes of diseases, medications, treatments, alternative therapies and lists of consultants. Medical portals can be either general or specific. General medical portals provide general medical information, for example BetterHealth or HealthInsite. Disease-specific portals provide information specific to a particular disease (e.g., cancer portal, breast cancer portal).

Portal: In Latin, the word porta means a gate or entrance. A portal can be defined as a gateway to information (Golier Inc., 1980). It is Web-based and connects a set of commonly used information or services via a link from a single Web page (Clarke and Flaherty, 2003; Elias and Ghaziri, 2004; Notess, 2002; O'Leary, 1999; Rao, 2001).

Chapter 5.18
Empowerment and Health Portals

Mats Edenius
Stockholm School of Economics, Sweden

INTRODUCTION

We know that interest in employing Web portals for communication between the health care sector and the public is constantly increasing (Kapsalis, Charatsis, Georgoudakis, Nikoloutsos, & Papadopoulos, 2004). We can also find an increasing demand for various kinds of such communication (Sciamanna, Clark, Diaz, & Newton, 2003). It promises to become an important and valuable tool for e-health (i.e., computer-based health care and health care management). Patients can log into a Web portal in order to find an appropriate medical treatment, communicate personal matters, and/or find the right way and place to find adequate health care.

In the last 10 years, we have also seen "empowerment" flourish, not least in an Internet context. In conventional research about Internet and Web portals, empowerment refers to a quite general process of gaining influence over events and outcomes. For example, it is argued that people are becoming more informed and managing their situations in better ways thanks to portal solutions and thereby becoming more empowered (Fergusson, 2004). Empowerment is central in the discourse of health and an important analytical concept to understand how portals work and can improve health care (Cathain et al., 2005).

In this article, I will argue that empowerment is indeed a fruitful concept to capture the potential of Web portals (in the health care sector). However, what is largely missing in the contemporary analyses is a more dynamic approach to analysing empowerment than found in conventional research, and how from such an approach we may justify the way Web portals are used to reach better results. The argumentation will be supported and illustrated by empirical material based on how different health portals are used.

UNDERSTANDING EMPOWERMENT: A TERM BASED ON POWER

As a researcher coming mainly from the social science and not medical science, it is natural to base an analytical discussion about empowerment on central thoughts on the concept of power. A modern definition of power tells us that power has to do with circumstances where one actor is able to make another actor perform against his or her will and interests (Lukes, 1974). Hardy and Leiba-O`Sullivan (1998) use Luke´s definition of power and explore it in terms of empowerment, in a way that could be very relevant to health portals too. They discuss how power is exercised by using various resources to influence the outcome of the decision-making processes. Several assumptions underlie this view of power. It says that all individuals are aware of their grievances and act upon them by participating in the decision-making process and using their influence to determine key decisions. We can find a whole range of possibilities related to health care where the physician brings different resources into the discussion (authority and information, etc) or hides information from the patient. Power could also be maintained by patients. They can abuse their power by giving the physicians spurious information for their own sake (to get drugs for example) or by misusing the health resources. The physician can also offer some information to the patient who becomes the sole decision maker. It is a situation that often is related to differences in cultures and deep values (e.g., Hofstede & Hofstede 2005) between paternalism and autonomy and between fidelity and humanity.

Power, rather than simply being exercised within decision-making processes, could also be used to exclude certain issues and patients from that process. Physicians could have the possibility to squeeze patients out and not let them come to the places where the decisions are made.

What both these two dimensions of power are based on is a situation where all parties more or less know their status and their will. But what about if the patients do not know what is best for them because the communication has been distorted? What this third dimension of power tells us is the importance of investigating what the fundamental base is that takes place in the decision arena. The patients could be said to be duped, coerced, or manipulated into political inactivity (or the opposite way around) via a Web portal (or by not accessing one).

Inspired by these thoughts we can draw the scheme presented in Table 1.

The scheme can help us to analyse empowerment in the context of Web portals; for example, can patients gain access to the decision arena thanks to a portal, etc? By using such a scheme, we also say that technology performs in a determined way and is therefore, in principle, determinable. However, to understand Web portals and empowerment I suggest it could also be fruitful—in a quite pragmatic way—to loosen up such a view by linking the discussion to the insight that we also apprehend and constitute the world through a technological frame, which is not innocent in this context. What I mean is that Web portals do not strictly answer this or that question, satisfy this or that demand, or extend this or that capacity. Rather, as Arnold (2003, p. 236) calls a substantive approach:

Technology works at a more fundamental level, it enframes the world such that the question is changed in a certain social context along with the answer, the need is changed along with its gratification, and direction is changed along with the mechanism.

So, a Web portal also enframes a particular brand of reality and functions therefore as a kind of knowledge making. This world is continually working to structure our thoughts and our thinking and acting processes.[1] The problem with the conventional way of regarding empowerment and Web portals is that the logic does not allow for opposite effects to be placed within the same effect frame. My argument is that an analytical discussion about Web portals and empowerment has much to gain from being complemented by a more—what could be called—dynamic approach. I will illustrate this by examples taken from two case studies—one focuses on Vårdguiden (Edenius & Westelius, 2004) and the other one on different patient communities[2] (Edenius & Åberg, 2005). Both of these case studies are based on interviews of users and owners of different health portals.

EMPOWERMENT AND HEALTH PORTALS: TWO MINI CASES

Vårdguiden

Vårdguiden is the Stockholm County Council's Web portal for health care and telephone consultation. The portal is a neutral arena where citizens in Stockholm can get information about the health care sector and get health care advice. What could be said to be unique with Vårdguiden, compared with most other Web health portals, is that it includes some interactive services. At the time of our study (2004), the portal included three such different communication services. Patients who had registered as users of the system could book, alter, and cancel appointments with physicians. They could also renew a prescription and renew registration on the sick list.

The patients really loved the portal. They said that they can reach health care in a good way thanks to the portal. The patients stress how thanks to the written language they have been able to communicate certain matters in different ways. They have been able to use the written discussion with the physicians as qualitative check. They can always go back to what they have discussed and the physicians´ answers. They also say that thanks to the portal they have gotten important information about their illnesses, information, and knowledge they can bring with them to their meeting with the health care. It is not a big leap to say—even if the analysis is quite schematic for illustrative purpose—that using the portal could help the patient in many different ways to become more empowered (cf. ability to mobilize resources, gaining access, and consciousness-raising). But, something else is going on at the same time. A more dynamic approach tells us something in addition.

Thinking in Terms of a Booking System

Several patients expressed a desire for more control over finding an appropriate date and time to see a physician. In the system, patients could request an appointment either morning or afternoon, adding a few sentences in free format. The reason for this (limited) selection of options was that the design group, while giving the patient some possibility to specify preferences, had decided to allow the health care provider considerable discretion in the booking procedure. However, the ability to request an AM or PM appointment did not decrease the patients' desire for an even more sophisticated time-booking system. One user said for example:

You could do it like the motor vehicle inspection site. There, you can decide on the date and the time. You can tick the time you want, and then it's yours.

In this case, the users only seem to be concerned with the functionality of the portal as a booking system. They do not view it as a means of making contact with their health care provider.

	First Dimension	Second Dimension	Third Dimension
Power of A over B.	Control of resources.	Control of decision-making processes.	Control of meaning.
Empowerment of B requires.	Acquisition of resources and ability to mobilize them.	Ability to gain access to the decision arena.	Consciousness-raising.

Compared to their previous experiences with booking systems, this appointment booking system seemed rather constrained and afforded the user less control than other booking systems. However, the user demand for quicker service and greater control when utilizing the portal does have a basis. The user's knowledge of the application is created through making generalisations based on other applications they are familiar with.

Thinking About the Organisation of Health Care

As previously seen, the system generated a demand for further control. However, the way the patient uses the system also imparts knowledge about how the health care is administrated. We can say that the portal opened a gate through which you could see more clearly what was going on in the health care administration.

One comment concerns the use of communication channels, questioning the rationality of the health care providers:

But why can't they reply to an e-mail if they can talk on the phone? Then, they would be rid of the call and could attend to the e-mail instead.

A number of comments also concerned how the Web service put them in contact with the least booked doctor (and maybe the least experienced), and how the e-messaging channel of communication was believed not to match the existing routines at the hospital. The e-messaging service can also convey the impression that the administration does not work properly. As one patient said:

Getting a reply…I had a letter with an appointment time. I do not know how long it took, but I am sure that it was a number of days before I received the letter. I also received an e-mail, but it said that I would receive a letter by post. Then I had the feeling that one hand did not know what the other was doing.

This quotation conveys a sense that the Web portal is more of a black box than traditional ways of communicating with the health care provider in a potentially empowering way. This black box is disturbing to the users, and they want more control, or at least to be able to look into the box, to achieve increased transparency.

Patient Communities

There are many statements that could be interpreted as the patients having become empowered by participating in a patient community too. Actually, it lies at the heart of patient communities. As a vice president of a community said:

Take these pills and get back in 3 months', or 'we write a letter of introduction from a doctor to the doctor specialised in rheumatology,' this is hard to influence with IT but how you get there and are able to make those demands, that is what we are trying to support our members with ...

The information found on the portal is frequently very practical. *The portal* supports their members by giving them advice and writing out lists of questions the patients can ask when seeing a doctor. The patient community is also based to a high extent on different people actively sharing their experiences by communicating online. This is done by different Web-based tools such as writing messages in the *guest book*, discussing in a forum, or traditional chatting. Another example of Web-based communication is the function *question of the day*, which is one of the strongest signs that the patient community is expecting to have some kind of dialogue with their members. By sharing this kind of information, it could be said that the members become aware that they are not alone and understand that it is possible to demand things since they learn that many people have already asked, wondered, or requested the same thing. Relating these statements—tentative as they might appear—to the analytical scheme of empowerment tells us that the patient communities could indeed empower the patient, by giving them resources and the ability to use them, ability to gain access to the decision arena, and not at least by raising consciousness.

Now a more dynamic approach; the fact that the knowledge is shared, extended, and widely spread is one of the strongest tools for a patient to use in the role as an independent patient. The identity issue appears to be very important for the member since they get empowered by speaking the same *language* based on a mutual understanding, as one patient said:

It has been useful to me in situations when everything feels meaningless and you feel disappointed with the doctors…people who understand!

In this non-medical related information service, the patient community is providing an essential type of information, not clinically related to the disease, but to the human being behind the disease who is dependent on practical solutions for daily life. This information seems even more important for the member and contains practical tips, advice, and support on how to handle day-to-day things. The uniqueness about this kind of information is that, since it is not scientific, the health care does not give any information about it, even if these kinds of questions appear to be the most important ones to the patient. The members *talking* means that they teach each other and learn from each other when discussing different issues around the disease.

Another function of the patient community is to follow up on the information given to the patient in a medical/health situation. The instructions/information is corrected, adjusted, developed, and diversified by the patient community to suit the patient/member better. In a way, this means that they give the patient distance to what the health care informs them about. Which in turn means that the members' *new* knowledge leads them to act and think more freely than before, leading to more independence.

I have an example of a girl at a conference, we were suppose to have supper at 6 PM, but she came and said to me, 'I have to eat my supper at 5.30,' ok, 'why' I asked her, 'I have to eat at that time because that is what my doctor told me,' ok but we are going to eat at 6, 'no I have to eat at 5.30!' I told her to take a sandwich or something in the mean time to be able to wait for the dinner at 6…. This person was so focused on what the doctor told her, how she should live her life, instead of looking at 'what can I do to have a normal life in my situation?'

According to the previous quotations, we could say that patient communities are taking the lead and science is following. Since patient communities have the *empirical* experience, in that they hear, see, and understand the patients' problems in another way than health care does, you could say they are the ones *making the doctor's rounds* based on empirical research. Such as the patient community is responsible for a certain way of creating knowledge based on experience from members who have done this in reality and practised it over many years. The patient community is the *practical researcher*, like a pioneer in patients' knowledge of health care, however not under the flag of scientific knowledge. As formulated by a patient:

There are no medical doctors who will tell a patient who suffers from pollen allergy wanting to eat an apple, because it is common to suffer from cross-allergy against apples, if you run it in the microwave for one minute the proteins disappears. That is one thing, which if you are talking to the doctor they would never say something like that, we have different roles, but I think that it is our strength in this …

Hence, the point is that the patient community and its Web portal application can offer the patient something the public health care cannot. They give the patient different knowledge compared to health care in general. The patient community considers the members as individuals who are in need of practical advice and not only general information. That is why communicating, using the members' language and not that of medical professionals, is a winning concept. In more theoretical terms we can say that the patient communities can maintain (at least partly via the portal) an embedded and embodied kind of knowledge, which allows for the recognition of individual diversity and the relevance of particularity. It recognises a role for practical reasons as opposed to a health care that could more be described in disembodied and disembedded terms, supported by proliferation of different techniques that have a universal relevance and may be applied to any patient (cf. evidence based medicine).

FUTURE RESEARCH OPPORTUNITIES

From what is said previously, I don't argue that everything is on the move. Several authors have stressed, and believe in, the possibility for actors to manage dynamic meanings too (cf. Giddens 1979). Therefore, it is maybe time not only to discuss how people can reach empowerment through Web portals, but also how we can manage the new kind of knowledge, which using a portal will generate. In the case of Vårdguiden, it might be done with information to meet the patients' new knowledge. The illustration of patient communities shows a more complicated case where we can ask if the patients overall could be empowered, while health care is stuck in a rather clear modern scientific discourse. The question is difficult to deal with, but nonetheless important to focus on and to research further.

CONCLUSION

This article is an analytical attempt not only to capture the concept of empowerment in a health portal context, but also to expand such a discussion. Using a Web portal is an active framing process animated by forces, tensions, and interests. I have argued that the tricky concept of empowerment could be fruitfully empirically discussed, starting from three different dimensions of power. However, I have also argued that empowerment is not only a process of problem-solving or emancipating processes, but at the same time a process of problem-constructing and thereby making the concept of empowerment more extensive; compared to something that might be more or less reached thanks to a portal in a deterministic way. This is a far cry from the realist view of empowerment, but nevertheless also an important way for analysing and discussing empowerment in an empirical context. Hopefully, my illustrations could be one starting point to discuss how we can improve the potential of portals in line with human behaviour.

REFERENCES

Arnold, M. (2003). On the phenomenology of technology: The "Janus-faces" of mobile phones. *Information and Organization*, *13*, 231-256.

Cathain, A., Goode, J., Luff, D., Strangleman, T., Hanlon, G., & Greatbatch, D. (2005). Does NHS direct empower patients? *Social Science & Medicine*, *61*(8), 1761-1771.

Edenius, M., & Åberg-Wennerholm, M. (2005). Patient communities, ICT, and the future: An explorative study about patient communities and their status in the ICT-era. ITS, Porto, September.

Edenius, M., & Westelius, A. (2004). Patients' knowledge formations through a health care e-messaging system. *Journal of Human Resource Costing and Accounting*, *8*(1), 21-34.

Feenberg, A. (1999). *Questioning technology*. New York: Routledge.

Fergusson. (2004). *God bless my CML support group: An interview with Gabe Lewis.* Retrieved from www.furguson.report.com

Giddens, A. (1979). *Central problems in social theory.* Oxford University Press.

Hardy, C., & Leiba-O'Sullivan, C. (1998). The power behind empowerment: Implications for research and practice. *Human Relations*, *51*(4), 451-483.

Hofstede, G. H., & Hofstede, G. J. (2005). Cultures and organizations: Software of the mind. New York: McGraw-Hill.

Kapsalis, V., Charatsis, K., Georgoudakis, M., Nikoloutsos, E., & Papadopoulos, G. (2004). A SOAP-based system for provision of e-services. *Computer Standard & Interfaces, 26*(6), 527-541.

Lukes, S. (1974). Power: A radical view. London: Macmillan.

Sciamanna, C., Clark, M., Diaz, J., & Newton, S. (2003). Filling the gaps in physician communication. The role of the Internet among primary care patients. *International Journal of Medical Informatics, 7*2(1-3), 1-8.

KEY TERMS

Deterministic Approach: Begins with science and technology and ends with social conditions.

Disembodied Knowledge: Knowledge that has a universal relevance.

E-Health: Computer-based health care and health care management.

Embodied Knowledge: Knowledge, which allows for the recognition of individual diversity and the relevance of particularity.

Empowerment: A process of gaining influence over events and outcomes.

Power: Circumstances where one actor is able to get another actor to perform against his or her will and interests.

Substantive Dynamic Approach: Technology enframes the world such that the question is changed in a certain social context along with the answer.

ENDNOTES

1. What I really do is, in a quite pragmatic way, to use a blend of a theoretical orientation that is referred to as the "Social Construction of Technology" and a more "Substantive" approach (e.g., Feenberg, 1999)
2. The Swedish Rheumatism Association, the Swedish Asthma and Allergy Association, and the Swedish Diabetes Association, including their Youth associations.

This work was previously published in Encyclopedia of Portal Technologies and Applications, edited by A. Tatnall, pp. 332-336, copyright 2007 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 5.19
Digital Divide and E-Health Implications for E-Collaboration Research

Michele Masucci
Temple University, USA

INTRODUCTION

E-health has rapidly gained attention as a framework for understanding the relationship between using information and communication technologies (ICTs) to promote individual and community health, and using ICTs for improving the management of health care delivery systems. The use of e-collaborative tools is implicit to the delivery and access of e-health. Development of the capacity to transmit and receive digital diagnostic images, use video telecommunications for supporting the remote delivery of specialized care and surgical procedures, and the use of e-communication technologies to support logistical elements of medical care (such as scheduling appointments, filling prescriptions, and responding to patient questions) are just a few ways in which e-communications are transforming how medical care is embedded within institutional, organizational, family, and community settings.

The emerging field of e-collaboration focuses attention on the need for society to critically examine how electronic communication technologies facilitate, shape, and transform the ways in which organizations, groups, and communities interact. There are many works that explain how to (a) develop e-health systems, (b) assess the use of such systems, and (c) analyze the health outcomes that can be achieved with effective e-health applications (Brodie et al., 2000; Eder, 2000; Spil & Schuring, 2006). Less attention has been paid to how advances in e-collaboration research might inform e-health applications development and scholarly discourse. Because of this gap in the literature, few discussions pertain to understanding patient perspectives about the advantages and disadvantages that may result from rapidly emerging interconnections among access to health care, health information, health support systems, and ICTs (Berland et al., 2001; Hesse et al., 2005; Gibbons, 2005; Gilbert & Masucci, 2006).

E-HEALTH AND THE DIGITAL DIVIDE

Facilitating equitable e-health remains a difficult challenge because of persistent disparities in using and accessing ICTs among vulnerable and marginalized population groups (Atkinson & Gold, 2002; Brodie et al., 2000; Gibbons, 2005; Skinner, Biscope, & Poland, 2003; West & Miller, 2006). More research is needed to examine how differing experiences, self-efficacies, and adaptive styles among users of e-communication tools relate to the collaborative aspects of accessing and delivering health care (Atkinson & Gold, 2002; Hsu et al., 2005; Katz, Nissan, & Moyer, 2004). In particular, the collaborative aspects of implementing effective e-health policies could focus on such issues as (a) the role of educational training in using e-systems for accessing health information, health care provider consultations, and health management protocols, (b) the effects of alternative information delivery systems for enhancing patient care and community wellness, (c) the ways in which patient knowledge acquisition processes are related to the use of e-communication systems, (d) the privacy concerns related to e-collaboration strategies for accessing patient health care records, and (e) the tradeoffs associated with a movement to integrate e-communication approaches across the continuum of health care access by patients and health care providers.

In addition, e-collaboration research can lead to an understanding of the ethical implications of advances in e-health. Such methodological approaches as social action research applications in e-collaboration can result in creating tools for implementing e-health systems (such as using e-mail exchanges to foster system compliance) while also investigating the means by these approaches work to improve e-health outcomes (Kock, 2004, 2005). What society stands to gain from inquiry into these issues is a greater understanding of how e-collaborative approaches can enhance the rapid move toward using e-technologies in achieving patient health outcomes and managing the delivery of health care systems (Gibbons, 2005).

Gilbert and Masucci (2005, 2006) have examined the ICT use frameworks among such population groups as a basis for determining the most effective means of understanding and supporting empowerment goals for those groups. A focus on e-health suggests that a consideration of values and experiences with ICTs could connect an understanding of how individuals relate educational training, ICT access, health knowledge acquisition, and health care access to examine the ultimate value placed on the adoption of e-health approaches for one's personal as well as family health (i.e., Cline & Haynes, 2001; Cotten & Gupta, 2004; Houston & Allison, 2002; Kickbush, 2001; Kivits, 2006; Reddick, 2006). And, as e-collaboration tools are examined for their potential to support equitable access to e-health systems, it is important to understand that the context within which they are used relates directly to the potential outcomes that can be achieved.

For instance, an e-health system that is designed to use e-mail reminders for checking blood pressure at home among patients with diabetes may not be effective if the health care provider examines the e-mails once per week due to workplace constraints. E-mail messages sent from a privacy-secured e-mail system within a hospital may not be accessible from remote locations by health care providers, further delaying responses to patients. Patients may not have frequent access to e-mail systems as a basis for reporting blood pressure or other health characteristics. An understanding of the use of the tool for enhancing e-health delivery should examine context as well as how different ICT use patterns shapes perspectives about (a) the benefits of e-health systems, (b) the challenges associated with learning how to use such systems, and (c) the different ways in which patients and providers approach e-communications and other e-collaboration tools for implementing such systems.

IMPLICATIONS FOR USING INTERNET TELEMEDICINE TO MANAGE HEALTH CONDITIONS

Internet telemedicine refers to the use of health communication tools delivered through the use of Web interfaces for managing specific health conditions. Implicit in the use of Internet telemedicine is the goal of using such systems to foster collaborations among patients and health care providers to manage specific health conditions and procedures. Such collaborations by definition recreate the geographies of health care access and delivery (Cutchin, 2002). This reconstitution of patient-provider communications can involve such e-communication enabled tasks as (a) transmitting self-monitored information about specific health conditions to physician accessible data bases, (b) patient and provider tracking of variables related to specific health conditions, and (c) improved patient-provider communications about the implications of trends related to specific conditions.

The collaborative roles of patients and health care providers ultimately take shape around the use of specific e-communication tools for specific health conditions. Patient empowerment has the potential to increase as they are drawn into more proactive involvement in the gathering and examination of data pertaining to their health conditions (Prokosch, Ganslandt, Dumitru, & Ückert, 2006).

A recent study of the use of an Internet telemedicine system to manage risk factors for cardiovascular disease illustrates the complexities involved in examining the relationships among geographic, social, and networked access to the Internet among low income patients (Masucci et al., 2006). The activities that were undertaken to improve the likelihood that individuals impacted by digital divide barriers would use the Internet telemedicine system included (a) developing and implementing an internet training protocol that addresses infrastructure and educational barriers to accessing ICTs among participants enrolled in the study, (b) assessing self-efficacy issues related to acquiring skills needed to use the internet communication tool developed for the study, and (c) analyzing social, demographic, and spatial patterns associated with health outcomes among patients who use the system (Masucci et al., 2006).

The participant group using the Internet telemedicine system was generally representative of people who have mitigated access to ICTs due to a combination of economic status, educational background and age (NTIA 1995, 1998, 1999a, 1999b, 2002, 2004). The 44 participants in the study were from inner-city Philadelphia and rural (non-suburban) northeastern Pennsylvania, with an average age of 60. Fifty-two percent of the study participants were African American; 73% were women; nearly 65% earned less than $25,000. The study found that training participants to use the telemedicine system for self-monitoring such information as blood pressure, steps walked, weight, and number of cigarettes smoked, resulted in strong compliance using the system, with 84% sending data after being trained (Masucci et al., 2006).

Discussions held during the training sessions suggest that inexperience with using e-communications is closely associated with living at the economic and social margins of society for many individuals. Most participants (66%) had no prior Internet experience; those who did were often relying on settings with old computers and low bandwidths (Masucci et al., 2006). However, nearly all participants placed a high value on using telemedicine and e-communication tools for managing their health. This may account for their willingness to allocate time needed to negotiate the problem of accessing ICTs outside of the home and involving others to assist them with using the system. Among those who had prior experience, use of the telemedicine system relied on accessing ICTs through the support of family members, community groups, shelters, and local libraries in order to use the telemedicine system. This involved needing to learn about the locations and terms of use that would apply to each setting. It also required that their understanding of the Internet telemedicine system was transportable among the different settings they used.

Perhaps due to the patchwork of access described by participants, significant gaps in their conceptual understanding of how e-communications and the Internet function needed to be addressed in the training discussions. Nonetheless, each had specific interests in gaining information through using the Internet, indicated through the wide variety of Internet uses demonstrated through training sessions. Included were searches related to specific health conditions, applying for jobs, shopping, games, recipes, and e-mail. Few seemed to understand the implications of designing the system to be Health Insurance Portability and Accountability Act (HIPAA) compliant. However, the issue of information privacy was a central theme of training discussions. Despite the limited experiences using ICTs among many participants in the study, discussions were highly nuanced. Participants expressed concerns about protecting personal identifying information (such as study IDs, usernames and passwords, and social security numbers), maintaining privacy around Internet search procedures, and the potential for e-communications with health care providers to be read by unauthorized individuals.

CONCLUSION

The perspectives gained through this study about the use of ICTs for managing health are important ones to consider in the debates about equitable uses of e-collaboration among patients and health care providers. Particularly as e-health approaches are increasingly relied upon to deliver health care, the embedded nature of e-communications in Internet telemedicine systems has the potential to empower patients in their interactions with health care providers and improve their ability to effectively use information related to their health disseminated through the Internet. However, such factors as (a) the availability of computers and the Internet at home, (b) prior ICT use experiences of individuals, (c) awareness of health information resources available on the Internet, and (d) skill levels with using e-communication tools and overall technological literacy, can affect the overall ability of an individual to benefit from using such systems. Moreover, it is likely that individuals who come from the most disadvantaged groups will continue to manage a tightly knit set of circumstances including low education and literacy levels, low income, and marginal housing stability, to benefit from e-health approaches such as using Internet telemedicine systems. Individuals and families who are challenged in all of these ways may have specific ideas about how they can benefit from e-health yet struggle to achieve the technology self-efficacy required to gain optimal results.

These concerns form the basis for a significant societal ethics and policy debate. How can individuals from marginalized groups maximize opportunities to benefit equally from e-health advances? A dialogue about this issue must attend to the underlying health disparities that are associated with socio-economic vulnerability. Internet health systems need to be developed with the needs of underserved populations at the forefront of design considerations. Perhaps one approach is to increase efforts to develop systems and information resources that address the health concerns faced disproportionately by marginalized groups, such as increased risks for diabetes and cardiovascular disease, as compared with mainstream populations. Another consideration is to develop systems that can be accessed and used by those with limited experience and skill using ICTs. In the case of the telemedicine system developed for this study, the patient interfaces were purposefully designed to be easy to read, free of distracting hyperlinks, and intuitive for individuals with basic literacy levels.

These changes were implemented based on learning about the skill levels of likely users and their perspectives about the value of the system. Prior to the design changes being implemented, the system was information rich, to the extent that patients with little prior experience using ICTs did not know where or how to get started. It is critical for the development of equitable e-health systems that the system development strategies reflect a concern for the highly contextual ICT experiences of potential users. An understanding of the value of health information, e-health delivery systems, and telemedicine by marginalized groups can be accomplished through interacting around the use of ICTs. This requires breaking the social isolation that can exist among groups through finding opportunities for collaborating around the shared goals for developing e-health systems. The use of e-communications for such collaborative processes may seem implicit, but given the digital divide barriers that may constrain some, the form and meaning of e-collaboration may have to be adjusted to reflect existing disparities. This work shows that such adjustments can provide insights into the complex array of decision-making that forms the context for the use of e-communication tools for improving health.

ACKNOWLEDGMENT

This study was supported, in part, by grants from the Pennsylvania Department of Health and the National Science Foundation (ESI-0423242). Any opinions, findings, and conclusions or recommendations are those of the author and do not necessarily reflect the views of the granting agencies.

REFERENCES

Atkinson, N. L., & Gold, R. S. (2002). The promise and challenge of e-health interventions. *Proceedings of the 2nd Scientific Meeting of the American Academy of Health Behavior, 26*(6), 494-503.

Berland, G. C., Elliott, M. N., Morales, L., Algazy, J. I., Kravitz, R. L., Broder, M. S., et al. (2001). Health information on the Internet: Accessibility, quality, and readability in English and Spanish. *Journal of the American Medical Association, 285,* 2612-2621.

Brodie, M., Flournoy, R. E., Altman, D. E., Blendon, R. J., Benson, J. M., & Rosenbaum, M. D. (2000). Health information, the Internet, and the digital divide. *Health Affairs, 19*(6), 255-265.

Cline, R. J. W., & Haynes, K. M. (2001). Consumer health information seeking on the Internet: The state of the art. *Health Education Research, 16*(6), 671-692.

Cotten S. R., & Gupta, S. S. (2004). Characteristics of online and offline health information seekers and factors that discriminate between them. *Social Science and Medicine, 59*(9), 1795-1806.

Cutchin, M. P. (2002). Virtual medical geographies: Conceptualizing telemedicine and regionalization. *Progress in Human Geography, 26*(1), 19-39.

Eder, L. (2000). *Managing healthcare information systems with Web-enabled technologies.* Hershey, PA: Idea Group Publishing.

Gibbons, M. C. (2005). A historical overview of health disparities and the potential of e-health solutions. *Journal of Medical Internet Research, 7*(5), e50.

Gilbert, M., & Masucci, M. (2005). Research directions for information and communication technology and society in geography. *Geoforum, 36*(2), 277-279.

Gilbert, M., & Masucci, M. (2006). Geographic perspectives on e-collaboration research. *International Journal of E-Collaboration, 2*(1), i-v.

Hesse, B. W., Nelson, D. E., Kreps, G. L., Croyle, R. T., Arora, N. K., Rimer, B. K., et al. (2005). Trust and sources of health information: The impact of the Internet and its implications for health care providers: Findings from the first health information national trends survey. *Archives of Internal Medicine, 165,* 2618-2624.

Houston, T. K., & Allison, J. J. (2002). Users of Internet health information: Differences by health status. *Journal of Medical Internet Research, 2002-4*(2), e7.

Hsu, J., Huang, J., Kinsman, J., Fireman, B., Miller, R., Selby, J., et al. (2005). Use of e-health services between 1999 and 2002: A growing digital divide. *Journal of the American Medical Informatics Association, 12,* 164-171.

Katz, S. J., Nissan, N., & Moyer, C. A. (2004). Crossing the digital divide: Evaluating online communication between patients and their providers. *American Journal of Managed Care, 10*(9), 593-598.

Kickbusch, I. S. (2001). Health literacy: Addressing the health and education divide. *Health Promotion International, 16*(3), 289-297.

Kivits, J. (2006). Informed patients and the Internet: A mediated context for consultations with health. *Professionals Journal of Health Psychology, 11*(2), 269-282.

Kock, N. (2004). The three threats of action research: A discussion of methodological antidotes in the context of an information systems study. *Decision Support Systems, 37*(2), 265-286.

Kock, N. (2005). Using action research to study e-collaboration. *International Journal of e-Collaboration, 1*(4), i-vii.

Masucci, M., Homko, C., Santamore, W., Berger, P., McConnell, T., Shirk, G., et al. (2006). Cardiovascular disease prevention for underserved patients using the Internet: Bridging the digital divide. *Telemedicine and E-health, 12*(1), 58-65.

National Telecommunications and Information Administration (NTIA). (1995). *Falling through the net: A survey of the "have nots" in rural and urban America*. Washington, DC: National Telecommunications and Information Administration.

National Telecommunications and Information Administration (NTIA). (1998). *Falling through the net ii: New data on the digital divide*. Washington, DC: National Telecommunications and Information Administration.

National Telecommunications and Information Administration (NTIA). (1999a). *Falling through the net ii: New data on the digital divide*. Washington, DC: National Telecommunications and Information Administration.

National Telecommunications and Information Administration (NTIA). (1999b). *Falling through the net: Defining the digital divide.* Washington, DC: National Telecommunications and Information Administration.

National Telecommunications and Information Administration (NTIA). (2000). *Falling through the net: Toward digital inclusion*. Washington, DC: National Telecommunications and Information Administration.

National Telecommunications and Information Administration (NTIA). (2002). *A nation online: how americans are expanding their use of the Internet*. Washington, DC: National Telecommunications and Information Administration.

National Telecommunications and Information Administration (NTIA). (2004). *A nation online: entering the broadband age.* Washington, DC: National Telecommunications and Information Administration.

Prokosch, H., Ganslandt, T., Dumitru, R. C., & Ückert, F. (2006). Telemedicine and collaborative health information systems. *Information Technology, 48*(1), 12-23.

Reddick, C. G. (2006). The Internet, health information, and managing health: An examination of boomers and seniors. *International Journal of Healthcare Information Systems and Informatics, 1*(2), 20-38.

Skinner, H., Biscope, S., & Poland, B. (2003). Quality of Internet access: Barrier behind Internet use statistics. *Social Science and Medicine, 57*(5), 875-80.

Spil, T., & Schuring, R. W. (2006). *E-health systems diffusion and use: The innovation, the user and the use IT model.* Hershey, PA: Idea Group Publishing.

West, D. M., & Miller, E. A. (2006). The digital divide in public e-health: Barriers to accessibility and privacy in state health department Web sites. *Journal of Health Care for the Poor and Underserved,* 17(3), 652-67.

KEY TERMS

Digital Divide: The National Telecommunications and Information Administration developed a report series (NTIA, 1995, 1998, 1999a and b, 2000, 2002, and 2004) that conducted an analysis of the ownership and use of computers by households in the U.S. The reports indicated that disparities between the households with the highest and lowest degrees of computer use and Internet access were related to income, race, and gender. The reports called the disparities a "digital divide."

E-Health: The use of information and communication technologies (ICTs) to support individual and community health through creating health management information systems, electronic scheduling systems, electronic prescription services, transmittal of health records and diagnostic imagery among providers, and the remote delivery of care and consultations.

E-Health Ethics: Also referred to as Health Internet Ethics (Hi-Ethics), considerations to ensure the equitable distribution of health benefits for individuals and families through using e-health to disseminate health services. Concerns include protecting information privacy, ensuring that current, reliable health information, and advice is shared using e-health approaches, and that access to services is available to all.

HIPAA: The Health Insurance Portability and Accountability Act created standards for regulating the transactional records related to the delivery of electronic health care. This involved creating national identifiers for health care providers, insurance plans, and places of employment. HIPAA requirements also exist to protect the privacy and security of health information.

Internet Telemedicine: The use of Web interfaces for accessing and delivering telemedicine services.

Technological Literacy: The knowledge, skills, and self-efficacy required to critically engage advancements in technologies and technological influences on daily life.

Telemedicine: The delivery of medical care between settings that are geographically separate through the use of telecommunication systems, including traditional telephone systems and electronic communication systems.

This work was previously published in the Encyclopedia of E-Collaboration, edited by N. Kock, pp. 153-158, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 5.20
Benefits Derived from ICT Adoption in Regional Medical Practices:
Perceptual Differences Between Male and Female General Practitioners

R. C. MacGregor
University of Wollongong, Australia

P. N. Hyland
University of Wollongong, Australia

C. Harvie
University of Wollongong, Australia

B. C. Lee
University of Wollongong, Australia

ABSTRACT

Information and communications technologies (ICTs) are being used more and more by general practitioners (GPs) in their day-to-day activities. While a number of studies have shown that ICT adoption and use can provide real benefits to medical practices, there have been few studies to determine whether the perception of those benefits is uniform across the sector. This study examines whether differences in the perception of benefits exist between male and female GPs. The results suggest that the groupings and priorities of benefits arising from ICT use differ substantially between male and female GPs. Results also show, amongst other things, that male GPs are focussed on business efficiency, while female GPs are focussed on communication and practice expansion.

INTRODUCTION

The use of ICTs within medical practices is by no means a new phenomenon. An examination of the literature provides numerous studies detailing the design of clinical ICT systems (Baldwin, Clarke, & Jones, 2002; Hsu et al., 2005; Pelletier-Fleury et al., 1999) as well as the uses of such systems within medical practice or health care facility (Ammenwerth, Mansmann, Iller, & Eichstadter, 2003; Catalan, 2004; Shohet & Lavy, 2004; Waring & Wainwright, 2002). Along with the increased adoption and use of ICTs in the small-to-medium enterprise (SME) sector (MacGregor, 2004a; MacGregor & Vrazalic, in press; Stockdale & Standing 2004; Woerndl, Powell, & Vidgen, 2005), there has been an increased interest both by academics as well as practitioners in the use of ICTs in the overall management of medical practices (Anderssen, Vimarlund, & Timpka, 2002; Ash, Gorman, Seshadri, & Hersh, 2004; Leung, Yu, Wong, Johnston, & Tin, 2003). These studies suggest that health care managers and professionals are increasingly examining ICTs, not only to improve medical services but to reduce costs and to operate more effectively and efficiently.

Not only has there been an increased use of ICTs both within the SME sector as well as the medical sector, a number of studies (Boerma & van den Brink-Muinen, 2000; MacGregor & Vrazalic, in press; Teltscher, 2002,) have shown that there is a substantial increase of females in the wider SME sector as well as in general practice. The increase in female participation and ICT adoption has, not unexpectedly, given rise to studies concerning the role of gender in ICT adoption and use. A number of studies (Brush & Hisrich, 1999; Buttner & Moore, 1997; Carter, 2000; Sandberg, 2003; MacGregor & Vrazalic, in press) have found gender differences in the perception of benefits derived from ICT adoption in the small business sector. Yet despite the increasing numbers of females in the workforce, there have been no studies examining whether differences exist in the perception of benefits derived from the use of ICTs between male and female GPs.

The purpose of this paper, then, is to examine whether the groupings and priorities of benefits derived from ICT adoption, differs between male and female GPs in regional Australia. As almost all of these practices are, by nature, specialized small businesses, the study examines the benefits (both the intrinsically medical benefits as well as the more general small business benefits) derived from ICT adoption and use.

The paper begins by providing a brief review of the literature concerned with gender differences from the wider small business sector. The paper then examines gender differences within the general practice sector as well as those differences found in the adoption and use of ICTs. It further investigates the nature of benefits derivable from ICT adoption and use, both from a general SME perspective as well as from a medical perspective. A description of the study and results are presented. This includes a series of Factor Analyses of the benefits derived from ICT adoption and use by male and female GPs. Finally the implications of the study results are discussed along with conclusions.

Gender Differences in the SME Sector

As most general practices in regional Australia are specialized small businesses, it is appropriate to examine the findings from this "wider" perspective, before examining medical practices or the use of ICTs. It is also appropriate to note the observations of Baker, Aldrich, and Liou (1997) and Carter (2000), who found that research into gender differences in the ownership/management of SMEs is scarce by comparison to research that has examined SMEs in general. Nonetheless, there are a number of interesting findings in the literature that compare various facets of gender differences in the ownership/management of SMEs. Studies by Carter (2000) and Brooksbank

(2000) suggest that that the primary motivation for moving into the SME sector is the desire to become self-employed. An examination of the United Kingdom's labor force figures for the 1990s shows that while the growth in self-employment for males was 4.73%, the growth in self-employment for females was 19.06% (Labour Force Survey 1990-1999). Studies by Nillson (1997), Brush & Hisrich (1999), and Sandberg (2003) have provided similar figures in Europe, the United States and Scandinavia.

A number of studies (Carter, 2000; Rosa, Hamilton, Carter, & Burns, 1994; Sandberg 2003) have examined gender differences both in the acquisition of financing and the use of that financing within the small business. A study of 600 UK SMEs (Carter & Rosa, 1998) found that males were more likely to make use of bank loans and overdrafts than females. Indeed, females were less likely to use or rely on financial institutional arrangements including cheaper sources of financing, such as extended supplier credit, than were their male counterparts. The UK study also showed that, on average, the capital expenditure on actual start-up by female owner/managers was 33% of that spent by males.

Not only has the use of financing been shown to differ between males and females, Carter (2000) and Sandberg (2003) have shown that the ability to access financing often differs between male and female owner/managers. Many of these studies have concluded that while financial institutions may have a nondiscriminatory policy, through stereotyping, the application of those policies are often prejudiced against women.

The increased difficulty of obtaining financing by female owner/managers has impinged on four areas:

- the ability to raise start-up financing (Carter, 2000);
- differences in guarantees required to attract financing (Carter, 2000; Sandberg, 2003);
- Attraction of ongoing financing through females' failing to penetrate the informal financial networks (Sandberg, 2003); and
- sexual stereotyping (Carter 2000)

A number of recent studies (McGregor & Tweed, 2001; Verheul, Risseeuw, & Bartelse, 2002) have examined differences in management style between male and female small business owner/managers. These studies found that female managers of SMEs were more relaxed with giving instructions to staff through informal conversation than were their male counterparts. Indeed, while the male managers stressed the role and use of power, female managers stressed the importance of interpersonal communication. Both studies also found that female owner/managers were more likely to hire external expertise and were more inclined to develop business strategies that were specific to their particular business than were their male counterparts.

Studies in Sweden (Sandberg, 2003) and New Zealand (McGregor & Tweed, 2001) showed that:

- female owner/managers paid more attention to business-to-business links;
- female owner/managers paid more attention to strategic alliances than males;
- female owner/managers managers were more mindful of both their customers and their staff than were male managers;
- female owner/managers were significantly better at dealing with the details of the day-to-day business; and
- female owner/managers were far more aware and capable of managing budgets than their male counterparts.

The role of networking, particularly social networking, has become an important area of research and government strategy within the small business sector. Networks provide: legitimacy (Dennis, 2000), business and technical support (MacGregor, 2004b), business structure (O'Donnell, Gilmore, Cummins, & Carson, 2002), distribution of power (Johanisson, Ramirez-Pasillas, & Karlsson, 2002), and business process, including supply chain management (Johanisson et al., 2002). It is interesting to note that while female managers tend to be less formal than male managers within their own businesses, when it comes to dealing with other managers in a business networking situation, male networks are often far more informal than female networks (DeWine & Casbolt, 1983; Smeltzer & Fann, 1989).

More recent studies (Brush, 1997; Carter, 2000; Carter & Rosa, 1998; Sandberg, 2003) support these earlier findings, adding that females appear to be less welcome in social networks often resulting in a reduced ability to use network partners to gain financing or attract technical or marketing assistance.

Gender Differences Among GPs

There have been a number of studies examining gender differences among GPs. Chambers & Campbell (1996) and Boerma and van den Brink-Muinen (2000) found that female GPs apply fewer technical procedures than males. A study by van den Brink-Muinen, Bensing, and Kerssens (1998) found that female GPs were more attentive to problems from staff than were their male counterparts.

In a large study of 8,183 GPs from across Europe, Boerma and van den Brink-Muinen (2000) found a number of gender differences between male and female GPs. These included data that showed that:

- female GPs more inclined to work fewer hours,
- female GPs tended to allocate more time to patients,
- female GPs had less technical equipment at their disposal,
- female GPs made fewer house calls,
- female GPs were less involved in services outside the practice (with the exception of providing training), and
- female GPs preferred group practices to solo practices.

The study by Boerma and van den Brink-Muinen (2000) provided a number of interesting implications for the future provision of care in general practice. These included the likelihood that with the increase in numbers of female GPs, there would be a similar increase in the number of group practices that would allow female GPs work flexibility and the ability to limit the need for a full-time commitment. The finding that female GPs had less technical equipment at their disposal suggested that modifications may need to be made in the education of medical practitioners.

In light of the findings by Boerma and van den Brink-Muinen (2000), it is appropriate that gender differences in the use of the Internet is now considered.

Gender Differences in Internet Use

A number of early studies (Gilroy & Desai, 1986; Meier & Lambert ,1991) found that males were less anxious about using computer technology than females. Yet, according to Gebler (2000), female Internet users exceeded male users in the year 2000. The implications of this event are significant considering the previous research into the use of the Internet by females. Shade (1998) and Sheehan (1999) both found that females were more concerned with privacy and security issues and, subsequently, more cautious about using the Internet for online shopping and trading. Presumably, the same concerns may apply to female business owners considering electronic commerce (e-commerce) adoption.

With the proliferation of ICTs, particularly into the small business sector, a number of studies have explored potential determinants of Internet use and found gender to be an influential variable in predicting this use (Sexton, Johnson, & Hignite, 2002). Other studies have attempted to develop frameworks to explain gender differences in Internet use and online behaviour (see, for example, Rodgers & Harris, 2003).

Although the gap between male and female Internet adoption rates has disappeared, resulting in a more gender-balanced use of the Internet, differences remain in how the Internet is actually used. For example, Akhter (2003) found that men were still more likely to use the Internet for shopping than women. This would suggest that males may be more open to e-commerce adoption as business owners because they are more willing to adopt the technology as consumers. However, empirical evidence of such a trend is still unavailable. Although our knowledge of gender differences in relation to *Internet* adoption as users and consumers is broad, our understanding of gender differences in relation to *e-commerce* adoption as business owners is scant and inadequate. One exception is a study of e-commerce and teleworking in 112 Spanish small businesses by Perez, Carnicer, and Sanchez (2002). The authors found that small businesses with female managers were significantly more concerned with the difficulty of using the technology, than were their male counterparts. The study also cited cost of the technology and changes to work procedures as being of more concern to female managers.

As already indicated, all regional general practices are really specialized small businesses. It is appropriate, then, that we examine the findings from the wider SME literature before considering the literature specific to general practice.

Benefits of ICT Adoption (SMEs)

There have been many studies concerned with the benefits derived from ICT adoption. Some of these studies (MacGregor, 2004a) have shown that benefits often fall into the category termed intangible benefits and are often not realized by owner/managers at the time of adoption. The following section will examine some of the findings.

Studies by Poon and Swatman (1997) and Abell and Lim (1996) found that SMEs benefited in their ability to reach new customers and new markets through the use of ICT. This finding also has been supported in more recent studies (Quayle, 2002; Raymond 2001;, Ritchie & Brindley, 2000; Sparkes & Thomas, 2001;, Vescovi, 2000).

Earlier studies found that other benefits reported by SME operators included lowering of administrative costs (Abell & Lim, 1996; Poon & Swatman, 1997;), reduced lead time (Abell & Lim, 1996), increased sales (Abell & Lim, 1996), improved relations with business partners (Poon & Swatman, 1997), and improved quality of information (Abell & Lim, 1996; Poon & Swatman, 1997). A recent study by Quayle (2002) found that benefits derived from ICT use, as reported by SME owner/managers, included reduced administrative costs, reduced production costs, reduced lead time, reduced stock, improved marketing, and improved quality of information, while Tetteh & Burn (2001) found that the adoption of ICTs substantially increased internal efficiency.

As mentioned, many of the benefits enjoyed by small businesses are intangible. Woerndl et al. (2005) found that many small businesses involved in ICT use had expanded their activities beyond those performed prior to adoption. Stockdale and Standing (2004) found that benefits included changes in business methods and costs, enhanced levels of communications, stimulated competition, and reduced transaction costs. Hurwitz (2000) concluded that ICT adoption led to greater potential for partnerships, while Brunn, Jensen, and Skovgaard (2002) showed that ICT use led to flexibility in administration and communication.

Benefits of ICT Adoption (Medical)

The adoption and use of ICTs within the health care sector is not entirely a new phenomenon. Studies of the National Health Service (NHS) in the United Kingdom by Gallagher (1998) noted that among other benefits were the ability to undertake strategic planning and better manage the health care environment. Rees (1998), in a similar study, noted that the use of ICTs appeared to enhance efficiency. Pullen, Atkinson, and Tucker (2000) showed that costs were better managed using certain forms of ICTs. These studies are supported by the findings of Nelson and Alexander (2002). It is interesting to note, however, that despite the apparent benefits, the approach to the design of ICT inclusions has been questioned by several authors (Gelnay, 2002; Shohet & Lavy, 2004; Waring & Wainwright, 2002), who noted that many fail to fully account for all the stakeholders of the enhanced system.

Other benefits noted in the literature include contact with other clinicians regarding patient care (Baldwin et al., 2002); elimination of redundancy in patient care (Pelletier-Fleury et al., 1999); enhancements to the effectiveness of the practice (Andersson et al., 2002); control of economic demands (Prop, 2000); improved patient care (Leung et al., 2003); and order entry of consumables by the practice (Ash et al., 2004).

METHODOLOGY

A series of interviews with GPs was undertaken to develop a set of benefits that are derived from ICT adoption (see Table 1). A questionnaire was then developed. Respondents were asked, amongst other things, to rate the benefits of ICT adoption across a 5-point Likert scale (1 very unimportant, 5 very important). Respondents also were asked their gender. The 690 surveys were distributed across five locations in Australia (Illawarra, Hunter, Ballarat, S.E.NSW, and Rockhampton). The scales of measurement were tested for reliability using a Cronbach's alpha reliability test. Cronbach's alpha was 0.942, indicating a high level of reliability.

RESULTS

Responses were obtained from 122 GPs, giving a response rate of 17.9%, which is within normal limits (10-30%) for questionnaire studies using mail-out as a vehicle (Sohal & Ng, 1998). Of the respondents 88 were male, and 34 were female.

An examination of the correlation matrices (available from the authors) suggested the use of factor analysis as a statistical method to investigate any separate underlying factors and to reduce any redundancy in the correlation matrix. The results of Kaiser-Meyer-Olkin MSA (0.880 for males; 0.718 for females) and Bartlett's test for sphericity (χ^2=1169, p=0.000 for males; χ^2=463 p=0.000 for females) indicated that the data set satisfied the assumptions for factorability. Principle components analysis was chosen as the method of extraction in order to account for maximum variance in the data, using a minimum number of factors. A three-factor solution was extracted for males with Eigenvalues 9.316, 1.546, and 1.150, while a four-factor solution with Eigenvalues 7.745, 2.312, 1.564, and 1.294 was extracted for females. Both were supported by an inspection of the Scree plot. Table 2 provides the total variance for both.

In both cases, the resulting components were rotated using a Varimax procedure to achieve simple structures. Tables 3 and 4 provide summaries of our data analysis and findings.

DISCUSSION

The first and most obvious difference between the findings in Tables 3 and 4 is that while male GPs considered that there were three factors underlying the benefits derived from ICT adoption,

Table 1. Benefits of ICT adoption and use

Expanding the patient/customer base by broadening the area of coverage
Improvement to business efficiency
Reduction of the overall workload
More time on patient care
Reduction of costs
Improvement in the way the business is operated
Allowing the business to expand
Information storage and retrieval
Communication with fellow GPs
Communication with other medical organisations
Disease management
Streamlining accounting and billing
Adding to the skills of the practice
Communication with hospitals
Ordering drugs
Communication with general practice business suppliers
Reducing the importance of remoteness in the provision of medical care

female GPs thought there were four. A comparison of the two tables, however, shows that there are several other important differences. While male GPs considered that expansion of the practice was part of the overall business effectiveness and rated this secondary to medical and business efficiency (Table 3), female GPs considered expansion of the practice to be aligned with communication and considered it to be of primary importance (Table 4). One possible explanation for this may be found in the earlier study of Boerma and van den Brink-Muinen (2000), who noted that female GPs preferred group practices, worked fewer hours, and tended to spend more time with patients. If this were the motivating force, it would explain the logical grouping of communication and practice

Table 2. Total variance explained

Males			
Component	Eigenvalue	**% Variance**	**Cum. %**
Med./bus. efficiency	9.316	54.798	54.798
Business effectiveness	1.546	9.097	63.894
Communications	1.150	6.766	70.660
Females			
Communication/expansion	7.745	45.559	45.559
Business efficiency	2.312	13.600	59.159
Efficient patient care	1.564	9.200	68.359
Communication with suppliers	1.294	7.614	75.972

Table 3. Rotated component matrix (male GPs)

Benefits of ICT adoption and use	**Medical and business efficiency**	**Business effectiveness**	**Communication**
Expanding the patient/customer base by broadening the area of coverage		0.734	
Improvement to business efficiency	**0.564**	**0.565**	
Reduction of the overall workload		0.694	
More time on patient care		0.745	
Reduction of costs		0.750	
Improvement in the way the business is operated		0.641	
Allowing the business to expand		0.535	
Information storage and retrieval	0.855		
Communication with fellow GPs			0.718
Communication with other medical organizations	0.648		0.631
Disease management	0.611		
Streamlining accounting and billing	0.766		
Adding to the skills of the practice	0.673		
Communication with hospitals			0.603
Ordering drugs			0.623
Communication with general practice business suppliers			0.741
Reducing the importance of remoteness in the provision of medical care			0.738

expansion (Table 4, column 1) as one mechanism to compensate for working fewer hours.

A comparison of Tables 3 and 4 shows that three benefits, "Reduction of the overall workload," "More time on patient care," and "Reduction of costs," were considered to be part of the improvement to business practice and effectiveness by males. By comparison, females loaded these benefits to a separate factor termed efficient patient care (Table 4, column 3). This again supports the earlier findings of Boerma and van den Brink-Muinen (2000), who found females spent longer with patients and were more attentive, both in time and procedure, with their patients than were males.

Tables 3 and 4 also show that male GPs considered ordering of drugs to be part of communication, while female GPs considered it to be part of patient care. Similarly, female GPs separated communication with suppliers from communication with other medical practitioners, hospitals, or medical associations. It is interesting to note that both groups considered business efficiency to be a greater benefit derivable from ICT adoption and use than improvement to patient care.

The results of this study are significant in several ways. The analysis has shown that e-commerce benefits can be grouped in relation to three main factors for males and four for females. This gives researchers a powerful explanatory tool because it reduces the "noise" in the data. Instead of accounting for 17 different benefits, the advantage to GPs of adopting and using ICTs can be

Table 4. Rotated component matrix (female GPs)

Benefits of ICT adoption and use	Communication and expansion	Business efficiency	Efficient patient care	Communication with suppliers
Expanding the patient/customer base by broadening the area of coverage	**0.643**			
Improvement to business efficiency		**0.747**		
Reduction of the overall workload			**0.743**	
More time on patient care			**0.799**	
Reduction of costs			**0.809**	
Improvement in the way the business is operated		**0.744**		
Allowing the business to expand	**0.460**			
Information storage and retrieval		**0.780**		
Communication with fellow GPs	**0.828**			
Communication with other medical organizations	**0.847**			
Disease management	**0.812**			
Streamlining accounting and billing		**0.881**		
Adding to the skills of the practice		**0.670**		
Communication with hospitals	**0.691**			
Ordering drugs			**0.580**	
Communication with general practice business suppliers				**0.837**
Reducing the importance of remoteness in the provision of medical care				**0.711**

explained in relation to one of three or four factors. The rotated component matrices shown in Tables 3 and 4 also enable the prediction of the scores of each individual benefit, based on the score of the other factors. Whereas before researchers have identified various benefits (such as the ones listed in Table 1), this is the first time a study has shown that certain benefits can be logically grouped according to three or four factors.

CONCLUSION

The results of this study are significant in several ways. The analysis has shown that e-commerce benefits can be grouped in relation to three main factors for males and four for females. This gives researchers a powerful explanatory tool because it reduces the "noise" in the data. Instead of accounting for 17 different benefits, the advantage to GPs of adopting and using ICTs can be explained in relation to one of three or four factors. The rotated component matrices shown in Tables 3 and 4 also enable the prediction of the scores of each individual benefit, based on the score of the other factors. Whereas before researchers have identified various benefits (such as the ones listed in Table 1), this is the first time a study has shown that certain benefits can be logically grouped according to three or four factors.

The results of this study also showed that the perception of benefits of ICT adoption and use in medical practices appear to differ depending on the gender of the GP. The results show that both the grouping as well as the priority of the groups differs substantially between male and female GPs. The current study appears to support the earlier findings of Boerma and van den Brink-Muinen (2000), who found, amongst other things, that female GPs have less technical equipment at their disposal. While the level of availability of technical equipment was not examined, the results of the current study suggest that female GPs see ICTs primarily as business oriented, while male GPs rated, as their highest priority, a medical benefit for their adoption and use. Clearly further research needs to be undertaken to determine whether the day-to-day use of ICTs by male and female GPs differs and the implications of any differences to patients.

The results of this study also show that ITC adoption benefits can, indeed, be grouped, simplifying the discussion of benefits in general. The apparent relationship of gender to perception of benefits is worthy of further examination because by formally explicating it, researchers and governmental organizations engaged in promoting ICT adoption will have more comprehensive knowledge about the issues and factors that have an effect on the relationship and will be able to provide better advice to GPs on ICT adoption.

REFERENCES

Abell, W., & Lim, L. (1996). *Business use of the Internet in New Zealand: An exploratory study.* Retrieved from http//www.scu.edu.au/sponsored/ausweb96

Akhter, S. H. (2003) Digital divide and purchase intention: Why demographic psychology matters? *Journal of Economic Psychology, 24*, 321-327.

Ammenwerth, E., Mansmann, U., Iller, C., & Eichstadter, R. (2003). Factors affecting and affected by user acceptance of computer-based nursing documentation: Results of a two-year study. *Journal of the American Medical Informatics Association, 10*(1), 69-84.

Andersson, A., Vimarlund, V., & Timpka T. (2002). Management demands on information and communication technology in process-oriented health-care organisations. *Journal of Management in Medicine, 16*(2/3) 159-169.

Ash, S., Gorman, P. N., Seshadri, V., & Hersh, W. R. (2004). Perspectives on CPOE and patient care. *Journal of the American Medical Informatics Association, 11*(2), 95-99.

Baker, T., Aldrich, H. E., & Liou, N. (1997). Invisible entrepreneurs: The neglect of women business owners by mass media and scholarly journals in the USA. *Entrepreneurship and Regional Development, 9*(3), 221–238.

Baldwin, L. P., Clarke, M., & Jones, R. (2002). Clinical ICT systems: Augmenting case management. *Journal of Management in Medicine, 16*(2/3), 188-198.

Boerma, W. G. W., & van den Brink-Muinen A., (2000). Gender-related differences in the organisation and provision of services among general practitioners in Europe: A signal to health care planners. *Medical Care*, 38(10), 993-1002.

Brooksbank, D. (2000). Self employment and small firms. In S. Carter & D. Jones-Evans(Eds.), *Enterprise and small business: Principles, policy and practice.* London: FT Prentice Hall.

Brunn, P., Jensen, M., & Skovgaard, (2002). E-marketplaces: Crafting a winning strategy. *European Management Journal, 20*(3), 286-298.

Brush, C. G. (1997). Women's entrepreneurship. *Proceedings of the OECD Conference on Women Entrepreneurs in Small and Medium Enterprises OECD, Paris.*

Brush, C. G., & Hisrich, R. (1999). Women owned businesses: Why do they matter? In Z. Acs (Ed.), *Are small firms important? Their role and impact* (pp.111-127). Boston: Kluwer, Academinc Publishers.

Buttner, E. H., & Moore, D. P. (1997). Women's organisational exodus to entrepreneurship: Self reported motivations and correlates with success. *Journal of Small Business Management,* (January), 34-47.

Carter, S. (2000). Improving the numbers and performance of women-owned businesses: Some implications for training and advisory services. *Education & Training 42*(4/5), 326-333.

Carter, S., & Rosa, P. (1998). The financing of male and female owned businesses. *Entrepreneurship and Regional Development, 10*(3), 225-241.

Catalan, J. (2004). Internet medicine sales and the need for homogeneous regulation. *International Journal of Medical Marketing, 4*(4), 342-349.

Chambers, R., & Campbell, I. (1996). Gender differences in general practitioners at work. *British Journal of General Practice, 46,* 291-293.

Dennis, C. (2000). Networking for marketing advantage. *Management Decision*, *38*(4), 287-292.

DeWine, S., & Casbolt, D. (1983). Networking: External communication systems for female organisational members. *Journal of Business Communication*, *20,* 57-67.

Gallagher, M. (1998). *Evolution of facilities management in the healthcare sector* (Construction Paper No. 86, pp. 1-8), Ascot, UK: The Chartered Institute of Builders, Ascot–.

Gebler, D. (2000). Rethinking e-commerce gender demographics. *E-Commerce Times,* October 6, Retrieved from http://www.ecommercetimes.com

Gelnay, B. (2002). Facility management and the design of Victoria Public Hospitals. *Proceedings of the CIB Working Commission 70: Facilities Management and Maintenance Global Symposium 2002 Glasgow,* 525-545.

Gilroy F.D., & Desai H.B., (1986) Computer anxiety: sex, race and age. *International Journal of Man-Machine Studies 25(6),* pp. 711-719.

Hsu, J., Huang, J., Kinsman, J., Fireman, B., Miller, R., Selby, J., & Ortiz, E. (2005). Use of e-health services between 1999 and 2002: A growing digital divide. *Journal of the American Medical Informatics Association, 12*(2), 164-171.

Hurwitz, (2000). *E-marketplaces: Issues, risks and requirements for success* (White Paper). Framingham, MA: Hurwitz group Inc.

Johannisson, B., Ramirez-Pasillas, M., & Karlsson, G. (2002). Theoretical and methodological challenges bridging firm strategies and contextual networking. *Entrepreneurship and Innovation* (August) 165-174.

Kolsaker, A., & Payne, C. (2002). Engendering trust in e-commerce: A study of gender-based concerns'. *Marketing Intelligence & Planning, 20*(4/5), 206-214.

Leung, G. M., Yu, P. L. H., Wong, I., O. L., Johnston, J.M., & Tin K. Y. K. (2003). Incentives and barriers that influence clinical computerisation in Hong Kong: A population based physician survey. *Journal of the American Medical Informatics Association*, *10*(2), 201-212.

MacGregor, R. C. (2004a). The role of formal networks in the ongoing use of electronic commerce technology in regional small business. *Journal of Electronic Commerce in Organisations, 2*(1), 1-14.

MacGregor, R. C. (2004b). Factors associated with formal networking in regional small business: Some findings from a study of Swedish SMEs. *Journal of Small Business and Enterprise Development, 11*(1), 60-74.

MacGregor, R. C., & Vrazalic, L. (in press). E-commerce adoption in regional SMEs. Hershey, PA: Idea Group.

McGregor, J., & Tweed, D. (2001). Gender and managerial competence: Support for theories of androgyny. *Women in Management Review, 16*(6), 279-286.

Meier, S. T., & Lambert, M. E. (1991). Psychometric properties and correlates of three computer aversion scales. *Behaviour Research Methods, Instruments and Computers, 23*(1), 9-15.

Nelson, M. L., & Alexander, K. (2002). The emergence of supply chain management as a strategic facilities management tool. In K. Alexander (Ed.), *Proceedings of the Euro FM Research Symposium in Facilities Management.* The University of Salford.

Nillson, P. (1997) Business counselling services directed towards female entrepreneurs—Some legitimacy dilemmas. *Entrepreneurship and Regional Development, 9*(3), 239-258.

O'Donnell, A., Gilmore, A., Cummins, D., & Carson, D. (2001). The network construct in entrepreneurship research: A review and critique. *Management Decision, 39*(9), 749-760.

Pelletier-Fleury, N., Lanoe, J. L., Philippe, C., Gagnadoux, F., Rakotonanahary, D., & Fleury, B. (1999). Economic studies and technical evaluation of telemedicine: The case of telemonitored polysomnography. *Health Policy 49,* 179-194.

Perez, M. P., Carnicer, M. P. L., & Sanchez, A. M. (2002). Differential effects of gender perceptions of teleworking by human resources managers. *Women in Management Review,* 17(6), 262-275.

Poon, S., & Swatman, P. (1997). The Internet for small businesses: An enabling infrastructure. *Fifth Internet Society Conference* 221-231.

Prop, (2000). Nationell handlingsplan for Utveckling av Halso – Och Sjukvarden No. 49.

Pullen S., Atkinson, D., & Tucker, S. (2000). Improvements in benchmarking the asset management of medical facilities. *Proceedings of the International Symposium on Facilities Management and Maintenance Brisbane*, 265-271.

Quayle, M. (2002). E-commerce: The challenge for UK SMEs in the twenty-first century. *International Journal of Operations and Production Management, 22*(10), 1148-1161.

Raymond, L. (2001). Determinants of Web site implementation in small business Internet research. *Electronic Network Applications and Policy 11*(5), 411-422.

Rees, D. (1998). Management structures of facility management in the National Health Service in England: A review of trends 1995-1997. *Facilities, 15*(3 /4), 254-261.

Ritchie, R., & Brindley, C. (2000). Disintermediation, disintegration and risk in the SME global supply chain. *Management Decision, 38*(8), 575-583.

Rodgers, S., & Harris, M. A. (2003). Gender and e-commerce: An exploratory study. *Journal of Advertising Research*, 43(3), 322.

Rosa, P., Hamilton, D., Carter, S., & Bums, H. (1994). The impact of gender on small business management: Preliminary findings of a British study. *International Small Business Journal, 12*(3), 25-32.

Sandberg, K. W. (2003). An exploratory study of women in micro enterprises: Gender related difficulties. *Journal of Small Business and Enterprise Development, 10*(4), 408-417.

Sexton, R.S., Johnson, R.A., & Hignite, M.A. (2002). Predicting Internet/e-commerce use. *Internet Research,* 12(5), 402-410.

Shade, L.R. (1998). A gendered perspective on access to the information infrastructure. *The Information Society, 14,* 33-44.

Sheehan, K. (1999). An investigation of gender differences in on-line privacy concerns and resultant behaviour. *Internet Marketing,* 159-173.

Shohet, I.M., & Lavy, S. (2004). Healthcare facilities management. State of the art review. *Facilities, 22*(7/8), 210-220.

Smeltzer, L.R., & Fann, G. L. (1989). Gender differences in external networks of small business owner/managers. *Journal of Small Business Management,* 25-32.

Sohal, A. S., & Ng, L. (1998). The role and impact of information technology in Australian business. *Journal of Information Technology, 13*(3), 201-217.

Sparkes, A., & Thomas, B. (2001). The use of the Internet as a critical success factor for the marketing of Welsh agri-food SMEs in the twenty first century. *British Food Journal, 103*(4), 331-347.

Stockdale, R., & Standing, C. (2004). Benefits and barriers of electronic marketplace participation: An SME perspective. *Information Management, 17*(4), 301-309.

Teltscher, S. (2002). E-commerce and development report 2002. United Nations Conference on Trade and Development.

Tetteh, E., & Burn, J., (2001). Global strategies for SME-business: Applying the SMALL framework logistics. *Information Management, 14*(1-2), 171-180.

van den Brink-Muinen, A., Bensing, J. M., & Kerssens, J. J. (1998). Gender and communication style in general practice: Differences between women's health care and regular health care. *Medical Care, 36,* 100 - 106

Verheul, I., Risseeuw, P., & Bartelse, G. (2002). Gender differences in strategy and human resource management. *International Small Business Journal, 20*(4), 443-476.

Vescovi ,T. (2000). Internet communication: The Italian SME case. *Corporate Communications: An International Journal, 5*(2), 107-112.

Waring, T., & Wainwright, D. (2002). Enhancing clinical and management discourse in ICT implementation. *Journal of management in Medicine, 16*(2/3), 133-149.

Woerndl, M., Powell. P., & Vidgen, R. (2005). Netwourcing in SMEs: E-ticketing in art venues. *Electronic Markets, 15*(2), 119.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 2, Issue 1, edited by J. Tan, pp. 1-13, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 5.21
Factors Motivating the Acceptance of New Information and Communication Technologies in UK Healthcare:
A Test of Three Models

Janice A. Osbourne
Brunel University, UK

Malcolm Clarke
Brunel University, UK

ABSTRACT

This paper discusses the use of three published models, the Technology acceptance model (TAM), Rogers diffusion of Innovation theory (IDT), and the Triandis theory of interpersonal behaviour (TIB), and attempts to bring them together in an integrated model to better predict the adoption of new information and communication technologies by a cohort of health professionals within UK primary care in an attempt to aid implementers in bringing technology in at an organizational level.

INTRODUCTION

Over the last 25 years, public healthcare delivery has been undergoing continuing changes. This has included the use of new information and communication technologies in a bid to improve services to patients, speed up waiting times, and addressing structural problems in the National Health Service (NHS). These changes have been largely driven by technical competence on the medical side but not matched sufficiently in technical organizational improvements. This article discusses the use of three published models, the Technology Acceptance Model (TAM), Rogers Diffusion of Innovation theory (IDT), and the Triandis Theory of Interpersonal behaviour (TIB), and attempts to bring them together to assist in the political decision to bring technology in at the organization level too.

PUBLIC HEALTHCARE IN THE UK: AN OVERVIEW

Within the United Kingdom, there exists a plethora of organizations and bodies providing the majority of healthcare in the UK including general practitioners to accidents and emergency departments, and dentistry. These organizations all fall under the National Health Service (NHS), the publicly funded healthcare system of each part of the UK, which in theory is managed by the Department of Health. Services provided under this organization are characterised by free service to all citizens and is divided into two levels of care, primary and secondary.

In the United Kingdom, a patient must first see their own doctor (referred to as the GP) located in close proximity to the patient's home. GP's are the first point of contact for users in the UK. This level of service provided is known as primary care. At present, 90% of all health and social care contacts with the NHS are through primary healthcare (NHS, 2001). Primary healthcare is provided through a combination of general practitioners and community medical workers. Services such as district nursing and child health monitoring are provided by community medical workers. If specialist help is required by a patient, he or she will be referred to a hospital or a consultant by their GPs. This is referred to as secondary care, as self-referral is not allowed and the clinical condition presenting normally cannot be dealt with by a primary care specialist and so is dealt with at this level.

One major problem in the NHS is that of communication between hospital specialists and general practitioners particularly in inner city areas. The written communication between GPs and consultants have been highlighted as being of poor quality (Rowland, 1992) and often having poor educational value (AGHTA, 1996). This problem has led to problems occurring in the outpatient referral system in terms of delays for hospital appointments, leading to frustration by patients (DOH, 1991).

In addition to this, there have been problems with out of hours GP services (Hallam, 1994), which has led to recommendations being made for more access to the healthcare system through entry points such as NHS Direct (Rogers, Chapple, & Sergison, 1999).

The government, in an effort to modernize the National Health Service and to deal with the numerous structural problems, have emphasized in policy initiatives the vision for connecting health policies with the capabilities of new information and communication technologies, which are able to provide new kinds of service that are more responsive to public needs and speed up access to healthcare. *The Information Strategy for the Modern NHS* (1998-2005) was seen as both visionary and relevant to the needs of the NHS. In the *Information for health* policy document for example, removing distance from healthcare was seen as a goal of the innovative technology Telemedicine.

Opportunities in the field of telemedicine will be seized to remove distance from healthcare, to improve the quality of that care, and to help deliver new and integrated services. GP's will be able to send test readings or images electronically to hospital specialists many miles away and in the same way receive results and advice more quickly (National Health Service Executive, 1998)

One of the major programmes, which the government has initiated, is the National Programme for IT for which an overview is provided.

National Programme for IT (Formerly NPfit)

The National Programme for Information Technology (Npfit) came about in 2002 as a result of the UK government's decision to make an unprecedented investment in information technology as essential to its plans for the National Health Service. This programme has been seen as one of the most expensive information technology programmes in the world with costs estimated to run over £18billion over ten years. (Brennan, 2005).

The programme aims to connect 30,000 general practitioners and 270 acute, community, and mental health trusts to a single, secure, national system to make information available when and where it is needed, including to patients themselves (Hutton, 2004).

In tracing its history, the Npfit programme came about as a result of suggestions made in the Wanless Treasury's report (Wanless, 2002). The report gave reasons as to why previous targets set by the NHS Executive, of all NHS trust having electronic medical reports implemented by 2005, had not been achieved. These reasons included budgets for information technology being used to relieve financial pressures elsewhere instead of locally and an inadequate setting of central information technology (IT) standards.

The Report recommended that the NHS double the proportion of its budget invested in information technology to 4%, to bring it closer into lines with the healthcare in the United States, which has a budget of 6%. They concluded that without a major advance in the effective use of information and communication technology, the health service would find it difficult to deliver the efficient high quality service, which the public will demand. The government's response to this was to allocate £2.3bn for a new national programme for IT (Doh, 2002). The Primary aim of the programme is **for** electronic patient records to be implemented in all acute trusts by the end of 2007. In 2005, the government established an agency, *Connecting for health*, as the single national provider of IT, which would deliver the National programme ensuring the maintenance, development, and effective delivery of the IT products and services delivered by the former NHS information authority.

There are several products that the National Programme plans to deliver, which includes:

- **A National Care Record System (NCRS):** *Connecting for health* plans to produce holistic records for patients from birth to death gradually, holding a review of the person's health and all his or her health contacts within the National Health Service. This summary patient record called the "spine" will be accessible 24 hours a day, seven days a week by health professionals, whether they work in hospitals, primary care, or community services.
- **Electronic Booking:** This programme, *Choose and book*, will allow GPs and other primary care staff to make initial hospital or clinic outpatient appointments at a convenient time, date, and place for the patient. If the patient prefers, he or she can make their appointment later online or through a telephone booking service.
- **Electronic Transfer of Prescriptions:** This will allow for prescribers to create and transfer prescriptions directly to a patient's community pharmacist and the Prescription Pricing Authority (PPA). It's aim is to reduce reliance on article prescriptions and it is expected to reduce prescription errors and provide better information at the point of prescribing.
- **New National Networking Service (N3):** This will provide the NHS with world-class networking services, including secure broadband connectivity, from every site where NHS services are delivered or managed

Based on the expenditure of this and similar programmes, we deduce that money is available so the choice is between spending the time to do it well, seeing that the government intends to have information technology playing a central part in healthcare within the very near future, or to do it badly on time. This then leads us to look at external models on which to identify factors, which play a part in technology adoption.

DEVELOPMENT OF RESEARCH MODEL

There have been several theoretical perspectives, which have been used to study the determinants of IT acceptance and adoption across a variety of disciplines including Communications (Rogers, 1995), Sociology (Wejnert, 2002), and Information systems research (Knol & Stroeken, 2001).

Intention-based models and behavioural decision theories have been used to explain usage of information systems (e.g., Agarwal & Prasad, 1997; Davis, 1989) and results further show that behavioural intentions are significantly and positively correlated with actual behaviour. According to these theories, user adoption and usage behaviours are determined by intention to use information technology, which in turn is influenced by beliefs and attitudes about information technology.

Although there are many theories, which can be used to explain intention we have selected three. The technology acceptance model, the theory of interpersonal behaviour, and Rogers innovation and diffusion model. These models will form the basis of our conceptual model.

Technology Acceptance Model (TAM)

The technology acceptance model (TAM, Davis, 1993) is described as the most dominant theoretical model in information technology acceptance (Misiolek, Zakaria, & Zhang) and is an adaptation of the Theory of Reasoned Action (TRA, Fishbein & Ajzen, 1980). The TAM's goal is to provide an explanation of the determinants of computer acceptance that is generally capable of explaining user behaviour across a broad range of end-user computing technologies and user populations (Davis, Bagozzi, & Warshaw, 1989).

The TAM identifies two beliefs perceived usefulness and perceived ease of use, as being of primary importance for computer acceptance behaviour. The model posts that a users adoption of an information system is determined by behavioural intention, which is determined by the user's belief about the system, similar to the TRA. The model differs from the TRA in that behavioural intention is viewed as being determined by the individual's attitude and perceived usefulness. The TAM does not however include the subjective norms of the TRA as it was not found to be significant (Davis, 1989).

The Triandis Theory of Interpersonal Behaviour

The Triandis model (Triandis, 1980, see Figure 1) explains individuals' behaviour in terms of what they have always done (habit), by what they think they should do (social norms) and by the consequences, they associate with a behaviour (perceived consequences). The model also contains aspects that are directly related to the individual, for example attitudes, genetic factors, intention, and behaviour and others that are related to the individual's environment, for example culture, facilitating conditions, and social situations.

Rogers Innovation and Diffusion Theory

The Innovation and Diffusion Theory proposes that an innovation is an idea or practice perceived as new by an individual, group, or organization

(Rogers, 1995). Diffusion is described as the process by which an innovation spreads. The individuals' decision to accept or reject an innovation occurs in the following stages, awareness of the innovation, which leads to an attitude being formed towards it based on the individuals perception of the innovation. The decision to adopt or reject is then made, implementation takes place, and the individual confirms his or her decision. The perceived characteristics of an innovation include its relative advantage, compatibility, complexity, trialability, and observability (Rogers, 1995).

REASONS BEHIND INTEGRATION

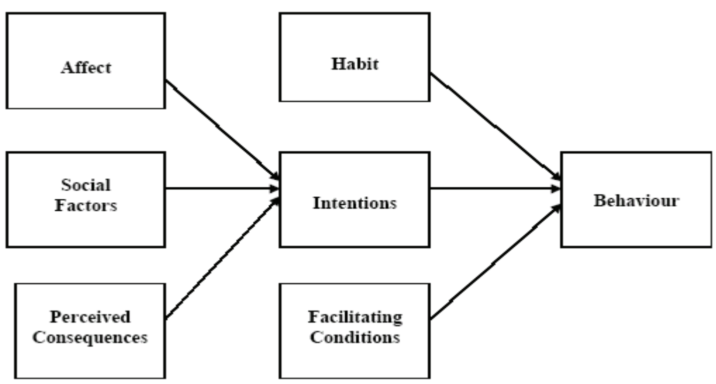
Previous models described here can contribute to our proposal of an integrated model in a variety of ways, which are:

Incompleteness

Although strong empirical support for the TAM has been established through numerous studies (Karahanna, Straub, & Chervany, 1999; Venkatesh & Davis, 2000; Venkatesh, Morris, Davis, & Davis, 2003), the TAM has had several criticisms levelled at it, including not offering sufficient understanding to provide system designers with information needed for creating and promoting user acceptance of systems (Mathieson, Peacock, & Chin, 2001), and its assumption that its use is volitional, which means that in there are no barriers to prevent an individual from using a technology or a system if he or she chose to do so (Mathieson et al., 2001). There may be situations in which an individual wants to use an IT, but is prevented by lack of time, money, or expertise (Chau & Hu, 2001; Taylor & Todd, 1995). Given the fact that the information and communication technologies, including mobile health technologies (described as computing technology, comprising software, hardware, and communications specifically associated with mobility, Zaslavsky & Tari , 1988) are fairly new to health professionals, the possibility exists that they may choose not to use the technology because they may not have sufficient skills or the ability to use the technology, hence the model would not give a complete picture of factors affecting adoption.

Because of the uncertain theoretical status of the TAM, Davis (1989) dropped the subjective norm, which has resulted in studies using the TAM seldom including variables related to the social environment. Technology adoption in healthcare is a current hot topic at the moment and we believe that social pressure plays an important role in explaining reasons behind their use.

Figure 1. Triandis Theory of Interpersonal Behaviour



Similarities Between the Models

Although these theories focus on different determinants to explain consumer behaviour in the adoption of technology, they have similarities:

1. The TAM and the Triandis model are both intention-based models, which assume that attitudes influence intention, which determines behaviour and the actual use of the technology.
2. Perceived ease of use in TAM is related to the complexity construct of the Innovation Diffusion theory in fact the complexity construct is the exact opposite of the Perceived ease of use construct.
3. Perceived consequences in the Triandis model construct is similar to the construct of perceived usefulness (PU) in the TAM and similar to relative advantage in the Innovation Diffusion theory.

We have divided the technology adoption process into three factor areas, those related to the social environment, those related to the technology, and those related to the individual. We find the Triandis model most relevant for factors related to the social environment; for the technology related section we find the attributes from the diffusion of innovation, which concentrates on both the technology and its compatibility and the TAM more suitable; and for those related to the individual, we find the IDT and the Triandis model more suitable. As these models do not give a complete overview on their own, we find it necessary to combine the three to evaluate adoption.

AN INTEGRATIVE MODEL

Given their complementary nature, a model that integrates the key research constructs from TAM, TIB, and EDT should explain more variance in IT usage intention than either model alone. Such an integrated model is depicted in Figure 3. In this section, we discuss how the assembled model attempts to do the duties assigned to it.

Social Factors

An individual's perception of social pressure to perform or not perform a behaviour affects intention, (Fishbein & Ajzen, 1975). Perception of social pressure refers to an individual's perception of whether individuals close to or important to them think that they should or should not perform a behaviour. Consequently, we view social factors as norms, values, and roles, which influence an individual's intention to adopt medical technology. These values in our context may be conveyed by interaction with patients and peers.

There are varying views on physicians being influenced socially. It has been suggested that general medical practitioners are influenced in their decision-making by medical specialists, who are seen as being innovative and creative. (Blumberg, 1999). Such opinion leaders can activate local networks to diffuse an innovation by facilitating transfer of information (Young, Hollands, Ward, & Holman, 2003).

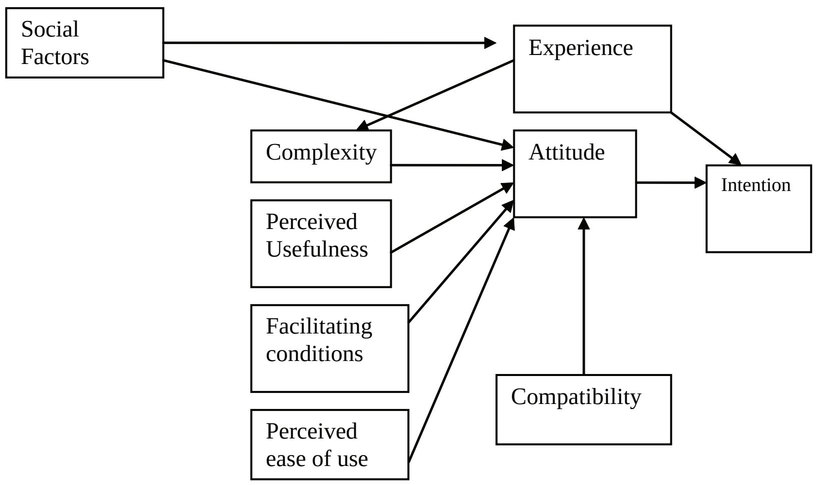
Perceived Usefulness

Among the many possible influences on technology use, Davis (1989) suggests that two determinants are of particular importance: perceived usefulness and perceived ease of use. Perceived usefulness is the tendency of people to use or not use a technology to the extent they believe it will help them perform their job better.

Perceived Ease of Use

PEOU is the degree to which a person believes using a particular system would be free of effort, and if the performance benefits of usage are outweighed by the effort of using the technology (Davis, 1989). Hence, the more complex the inno-

Figure 2. Integrated model



vation of an information technology, the lower the probability of its adoption will be (Rogers, 1995). Applying this to technology adoption in the NHS, technologies perceived to be user friendly will be more appealing to both patients and practitioners. Perceived ease of use refers to ease of using the technology, sending data electronically as in the case of mobile technology, obtaining the data necessary to proceed with consultations as in the case of General Practitioners.

Facilitating Conditions

Facilitating conditions are the objective factors that make a behaviour easy or difficult. In the Triandis model, facilitating conditions are important determinants of behaviour. Even if intentions to perform the behaviour are high, the habits are well established and the physiological arousals are optimal, there may be no action/behaviour if the situation or objective factors do not warrant the behaviour (Triandis, 1980). A simplified explanation of this is that facilitating conditions are important in that individuals with the intention of accomplishing something may be unable to do because their environment prevents the activity from being performed.

Triandis hypothesizes that facilitating conditions directly affect the actual behaviour rather than intentions because he argues that one might

Figure 3. Comparison of the three models

	TAM	TIB	IDT
Variables	Perceived ease of use	Social factors	Relative advantage
	Percieved usefulness	Facilitating conditions	Trialability/Experience
	Attitude	Perceived consequences	Compatibility
		Genetic Factors	Complexity
		Cultural Factors	Observability

Key:
TAM : Technology Acceptance Model
TIB : Triandis Theory of interpersonal behaviour
IDT : Innovation Diffusion Theory

have the intention to perform a certain act, but if the environment does not support this behaviour then the act will most likely not be executed. We define facilitating conditions as those factors in an individual's environment that facilitate the act of adopting technology. Empirical investigations have shown that facilitating conditions could also have a significant positive impact on attitude (Chang & Cheung, 2001).

We expect facilitating conditions to have a positive influence on technology adoption.

Perceived Consequences

According to Triandis, each act or behaviour is perceived as having a potential outcome that can be either positive or negative. An individual's choice of behaviour is based on the probability that an action will provoke a specific consequence.

Relative Advantage

Relative advantage is the degree to which an innovation is perceived as better than the idea it supersedes. Physician perceptions of new technologies affect the adoption of innovation in clinical practice. For example, physician perceptions related to teledermatology suggest that this new technology needs to be quick, efficient, reliable, and easily used (Weinstock, Nguyen, & Risica, 2002). If new technologies are too time consuming or complicated, they may not be widely adopted and used. In a study on the perceptions of GPs towards teledermatology, the results reported a preference for reliable, efficient, and easy-to-use technology (Collins, Nicolson, Bowns, & Walters, 2000).

Trialability/Experience

Trialability is the degree to which an innovation may be experimented with a limited basis. The more adopters experiment with a new technology and explore its ramifications, the greater the likelihood that the innovation will be used during early stages of adoption (Agarwal & Prasad, 1997; Natek & Lesjak, 2006).

Intention

Behavioural intention refers to "instructions that people give to themselves to behave in certain ways" (Triandis, 1980) or their motivation regarding the performance of a given behaviour. In our model, intention refers to health professionals' motivation to adopt information technologies.

Attitude

According to Fishbein and Ajzen (1980), "attitude is a learned predisposition to respond in a consistently favorable or unfavorable manner with a given object." Attitude is directly related to behavioural intention and adoption because people will only intend to perform behaviour for which they have positive feelings (Han, Harkke, Mustonen, Seppanen, & Kallio, 2005; Marino, 2004). We hypothesise that attitude is positively related to the users intention to use new ICTs in healthcare

Compatibility

This construct has not been included in most studies on technology adoption apart from those using the IDT (Cho,). Compatibility refers to the conformity of an innovation in our case new ICTs with the values and beliefs of users and with previously introduced ideas and needs (McCole, 2002). Therefore, the adoption of a new technology depends on its compatibility with existing practices. The importance of technology compatibility with the organization and its tasks has been shown to be a significant factor in successful technology implementations (Cooper & Zmud, 1990).We hypothesise that compatibility is positively related to the user's attitude towards new ICT's in healthcare.

DISCUSSION

The primary objective of this paper was to compare two of the more dominant technology acceptance and adoption models in IT research, namely the TAM and the IDT as well as a less popular model from the social science, the TIB, then extend these models to an integrated model of IT acceptance and adoption.

On their own these models do not fully evaluate the acceptance of technology by individual, however when integrated the weaknesses of the individual models are totally cancelled out by the strengths of the integrated model.

As described earlier, the TAM assumes that there are no barriers to preventing an individual from using a technology if he or she wished. The IDT and the TIB however feature variables, which can affect adoption by the individual, hence filling one of the gaps of the TAM. The Triandis models which attempts to cover all the major component of the individual and their environments with its 34 variables is somewhat difficult to test by virtue of its complexity hence only the variables relating to our study have selected. The IDT focuses on the process of adoption as well as the, this is complementary with both perceived characteristics of the technology complimenting both the TIB and the TAM

Based on these factors, we believe the integrated model will have better explanatory power of user adoption of technology. However dear readers, as with any cake, the proof is in the eating and as such, we will shortly be conducting an exhaustive study on a cohort of health professionals within a Primary Care Trust given new medical technology as part of a government modernization initiative to see if the model lives up to expectation. We ask that you wait with bated breath!

REFERENCES

Agarwal, R., & Prasad, J. (1997). The role of innovation characteristics and perceived voluntariness in the acceptance of information technologies. *Decision Sciences*, *28*(3), 557-583.

AGHTA. (1996). *Advisory group on health technology assessment. Assessing the effects of health technologies: Principles practice proposals.* London: Department of Health.

Blumberg, R. S. (1999). The leadership of innovation. *Gastroenterology, 116*(4), 787.

Brennan, S. (2005). *The NHS IT project.* Abingdon, UK: Radcliff Medical Press.

Chang, M. K., & Cheung, W. (2001). Determinants of the intention to use Internet/WWW at work: A confirmatory study. *Information & Management, 39*(1), 1-14.

Chau, P. Y. K., & Hu, P. J. H. (2001). Information technology acceptance by individual professionals: A model comparison approach. *Decision Sciences, 32*(4), 699-719.

Cho, V. (2006). A study of the roles of trusts and risks in information-oriented online legal services using an integrated model. *Information and Management 43*(4) 502-520.

Collins, K., Nicolson P., Bowns. I, Walters, S. (2000). *General Practitioners' Perception of Telemedicine in Dermatology: A Psychological Perspective Journal of Telemedicine and Telecare 6*(1), 50-53

Cooper, R., & Zmud, R. (1990). Information technology implementation research: A technological diffusion approach. *Management Science, 36*(2), 123-139.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340.

Davis, F. D. (1993). User acceptance of information technology: System characteristics, user perceptions, and behavioural impacts. *International Journal of Man-Machine Studies, 38*(3), 475-487.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982-1003.

Doh. (2002). *Delivering 21st century IT support for the NHS. National strategic programme*. London: Department of Health.

DOH. (1991). *Research for health: A research and development strategy for the NHS*. London: HMSO.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behaviour: An introduction to theory and research*. Reading, MA: Addison Wesley.

Fishbein, M., & Ajzen, I. (1980). Predicting and understanding consumer behaviour: Attitude behaviour correspondence. In I. Ajzen, & M. Fishbein (Eds.), *Understanding attitudes and predicting social behaviour* (pp. 149-172). Englewood Cliffs, NJ: Prentice Hall.

Hallam, L. (1994). Primary medical care outside normal working hours: Review of published work. *British Medical Journal, 308*(6923), 249-253.

Han, S., Harkke, V., Mustonen, P., Seppanen, M., & Kallio, M. (2005). Understanding physician acceptance of mobile technology: Insights from two telephone interviews in Finland. *International Journal of Electronic Healthcare, 1*(4), 380-395.

Hutton, J. (2004). *National programme for IT. House of commons official report* London: Department of Health.

Karahanna, E., Straub, D. W., & Chervany, N. L. (1999). Information technology adoption across time: A cross-sectional comparison of pre-adoption and post-adoption beliefs. *MIS Quarterly, 23*(2), 183-213.

Knol, W. H., & Stroeken, J. H. M. (2001). The diffusion and adoption of information technology in saml - and mediums - ed enterprises through IT scenarios. *Technology Analysis, & Strategic Management, 13*(2), 227- 46.

Marino, A. (2004). Innovation technology and application issues in Italian e-healthcare. *International Journal of Electronic Healthcare, 1*(2), 210-220.

Mathieson, K., Peacock, E., & Chin, W. W. (2001). Extending the technology acceptance model: The influence of perceived user resources. *DATA BASE for Advances in Information Systems, 32*(3), 86-112.

McCole, P. (2002). The role of trust for electronic commerce in services. *International Journal of Contemporary Hospitality Management, 14*(2), 81-87.

Misiolek, N., Zakaria, N., & Zhang, P. (2002). *Trust in organizational acceptance of information technology: A conceptual model and preliminary evidence*. Paper presented at the annual meeting of the Decision Sciences Institute, San Diego, CA (November 23-26)

Natek, S., & Lesjak, D. (2006). Trial work: The way to successful information system projects in healthcare. *International Journal of Electronic Healthcare, 2*(3), 223-230.

National Health Service Executive. (1998). *Information for health: An information strategy for the modern NHS 1998-2001*. London: National Health Service Executive.

NHS. (2001). *Building the information core—Implementing the NHS plan*. London: Department of Health, UK Government.

Rogers, A., Chapple, A., & Sergison, M. (1999). If a patient is too costly they tend to get rid of you: The impact of people's perceptions of rationing on the use of primary care. *Healthcare Analysis, 7*(3), 25-237.

Rogers, E. M. (1995). *Diffusion of innovations* (4th ed.). New York: The Free Press.

Rowland, M. (1992). Communications between GP's and specialists. In M. Roland, & A. Coulter (Eds.), *Hospital referrals* (pp. 108-22). Oxford: Oxford University Press.

Taylor, S., & Todd, P. A. (1995). Understanding information technology usage: A test of competing models. *Information Systems Research, 6*(2), 144-176.

Triandis, H. C. (1980).Values, attitudes, and interpersonal behaviour. In M. M. Page (Ed.), *Nebraska symposium on motivation 1979* (pp. 195-295). Lincoln: University of Nebraska Press.

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, *46*(2), 186-205.

Venkatesh, V., Morris, M., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 26*(4), 425-478.

Wanless, D. (2002). *Securing our future health: Taking a long-term view. Final report.* Retrieved September 27, 2005, from http://www.hm-treasury.gov.uk./Consultations_and_Legislation/wanless/consult_wanless_final.cfm

Weinstock, M., Nguyen, F., & Risica, P. (2002). Patient and referring provider satisfaction with teledermatology. *Journl of the American Academy of Dermatology, 47*(1), 68-72.

Wejnert, B. (2002). Integrating models of diffusion of innovations: A conceptual framework. *Annual Review Sociology, 28*(1), 297-326.

Young, J. M., Hollands, M. J., Ward, J., & Holman, C. D. J. (2003). Role for opinion leaders in promoting evidence-based surgery. *Archives of Surgery, 138*, 785-791.

Zaslavsky, A., & Tari , Z. (1988). Mobile computing: Overview and status. *The Australian Computer Journal, 30*(2), 42-52.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 1, Issue 4, edited by J. Tan, pp. 29-39, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 5.22
Gender Differences in Adoption and Use of a Healthcare IT Application

Kai Zheng
Carnegie Mellon University, USA

Rema Padman
Carnegie Mellon University, USA

Michael P. Johnson
Carnegie Mellon University, USA

Herbert S. Diamond
The Western Pennsylvania Hospital, USA

INTRODUCTION

Information technology (IT) adoption and diffusion is a central concern of information systems research and practice. The most widely-accepted method in IT adoption and diffusion research, the technology acceptance model (TAM; Davis 1989), posits that perceived ease of use and perceived usefulness are fundamental determinants of user acceptance. However, TAM and its subsequent research makes little or no reference to gender effects (Adams, Nelson, & Todd, 1992; Chin & Gopal, 1995; Venkatesh & Davis, 2000), despite the fact that researchers have shown that socio-cultural factors, such as gender and ethnic differences, influence human perceptions and behaviors (Hofstede, 1980). These socio-cultural factors can result in differences in user responses to technology innovations (Gefen & Straub, 1997).

Aiming to provide theoretical extensions to the TAM model, researchers have shown that gender differences may relate to beliefs and use of IT. For instance, males and females are found to demonstrate distinct adoption behavior in use of a wide spectrum of IT applications, such as e-mail (Gefen & Straub, 1997), mobile telephony (Ling,

2000) and Internet (Kraut, Scherlis, Mukhopadhyay, Manning, & Kiesler, 1996). Nevertheless, the exact gender effect remains a controversy. Some researchers believe that females are less technology inclined, less motivated and, therefore, less competent in masculine computer and technology culture (Wilder, Mackie, & Cooper, 1985; Qureshi & Hoppel, 1995). Others, in contrast, argue that females have the ability to be proficient in adopting new technologies (Turkle, 1995). Some research results indicate that females tend to favor some technology innovations and use them more effectively than males, such as computer-mediated communication (Kraut et al., 1996; Morahan-Martin & Schumacher, 1997).

In health care, influence of physician gender has long been noted in resident education and many practice areas. Researchers find the procedural and obstetrical care pattern of practice differs between male and female residents (Chaytors, Szafran, & Crutcher, 2001), and physician gender significantly affects treatments in adult primary care practice (Boulis & Long, 2004). An understanding of these socio-cultural issues is also of vital importance towards success of health care IT applications. This study is thereby designed to assess medical residents' acceptance and adoption of a Clinical Reminder System (CRS), by examining several key user characteristics that may relate to adoption and use of the system.

BACKGROUND

CRS is a class of computerized clinical decision support systems (CDSS) that send just-in-time alerts to clinicians when potential errors or deficiencies in patient management are detected. Beneficial outcomes of CDSS have been documented in many studies along a number of dimensions, including compliance with treatment standards, reduced costs and improved health outcomes (McDonald, Hui, Smith, Tierney, Cohen, Weinberger et al., 1995; Curtin, Hayes, Holland, & Katz, 1998). Several systematic reviews have also shown that CDSS can be an effective means of implementing medical guidelines to enhance clinical performance in wide-ranging aspects of medical care (Hunt, Haynes, Hanna, & Smith, 1998; Kaplan, 2001). However, most of these studies either focus on the accuracy and relevance of the computer-aided recommendations, or use experimental or randomized controlled clinical trials (RCCT) designs to assess system or clinical performance. Very few studies of CDSS involve a naturalistic design in routine settings with real patients (Kaplan, 2001). It is not clear whether a CDSS that has been shown to be effective in a laboratory setting will be fully utilized by end-users in clinical environments, and whether these users will adapt their practice style to efficiently accommodate computer-generated reminders.

Clinical users also differ in many ways. It has been recognized that individual users' experiences and their opinions or reactions to a technology make a difference in whether or not the technology will be adopted (Straub, Limayem, & Karahanna-Evaristo, 1995). Nevertheless, few studies have used individual-level data to measure the magnitude of user differentiation and the impact of such differentiation on technology adoption.

The present study aims to address some of these limitations. First, the system under evaluation, CRS, has been integrated into the daily operations of an ambulatory clinic. Second, longitudinal usage data for an evaluation period of 10 months were collected from computer logs, providing an objective and non-intrusive measure of the actual use over time. Third, a novel developmental trajectory method is applied to the usage data to identify groups that demonstrate distinct adoption behaviors, and to relate estimated group configurations to a variety of user characteristics.

The next section describes CRS and its basis in principles of evidence-based medicine. The study site and data collection procedures are then discussed, followed by a presentation of the trajectory analysis method for analyzing the usage data. The findings are presented next, and the final section presents some concluding remarks.

CRS AND EVIDENCE-BASED MEDICINE

Evidence-based medicine is the distillation of a large volume of medical research and standards into treatment protocols for diseases and preventive care procedures that represent the most accurate knowledge available (Sackett, Rosenberg, Muir Gray, Haynes, & Richardson, 1996). Evidence-based medicine has been widely applied to systematically review, appraise and use clinical research findings to aid the delivery and provisioning of optimum clinical care to patients.

The clinical decision support application that we developed, CRS, incorporates evidence-based medicine principles to assist in patient management decisions. It integrates the hospital's administrative, laboratory and clinical records systems into a single application, and uses patients' current medical status to provide reminders to clinicians at the point of care that reflect evidence-based medicine guidelines. Reminders generated by CRS take the form of recommendations to have tests scheduled and performed, receive vaccinations, alert clinicians to review abnormal test results or closely monitor patients with medical conditions that require unscheduled intervention. The CRS deployed at the time of this study is a distributed windows application based on client-server architecture. The clients are written in Visual Basic and communicate with an Oracle database server via the hospital's internal computer network. Evidence-based medicine guidelines are programmed in Oracle PL/SQL procedures.

STUDY SITE AND DATA COLLECTION

The study was conducted in the primary care clinic of an urban teaching hospital offering comprehensive health care services. Given the availability in every exam room of desktop computers installed with CRS, residents used the system during patient encounters.

Data collection started on February 1, 2002. In this study, we report 10 months of usage data to identify distinct user groups and maturation of their interaction with the reminder system. There were 44 active residents registered in the application. Activities generated by three of them are removed from the analysis since there were no visits recorded for these residents in 6 or more continuous months.

We have included in the study several key attributes of users that may relate to their human, contextual and cultural characteristics that in turn affect system use. These attributes are: general demographics including gender and citizenship (U.S. vs. non-U.S.); computer literacy identified by use, knowledge and optimism of computer systems; and frequency of encounters (work load). These attributes are found to influence adoption behavior in some other contexts (Berner & Maisiak, 1999). We have excluded a few other potentially influential attributes, because they are (a) invalid in the context, such as medical specialty (all users are internal medicine residents); or (b) lack of variation in the sample, such as age (mean 29.6, standard deviation 2.1); (c) sample size too small for defined subgroups, such as year of residency; or (d) inaccessible to us in recorded form, such as evaluation of residents' clinical performance (due to confidentiality concerns).

DEVELOPMENTAL TRAJECTORY ANALYSIS

We use a novel method, developmental trajectory analysis (DTA), to study user adoption behavior of this reminder system. DTA is a semi-parametric, group-based approach for identifying distinctive groups of individual trajectories within the population and for profiling the characteristics of group

Figure 1. Developmental trajectories of three distinct groups

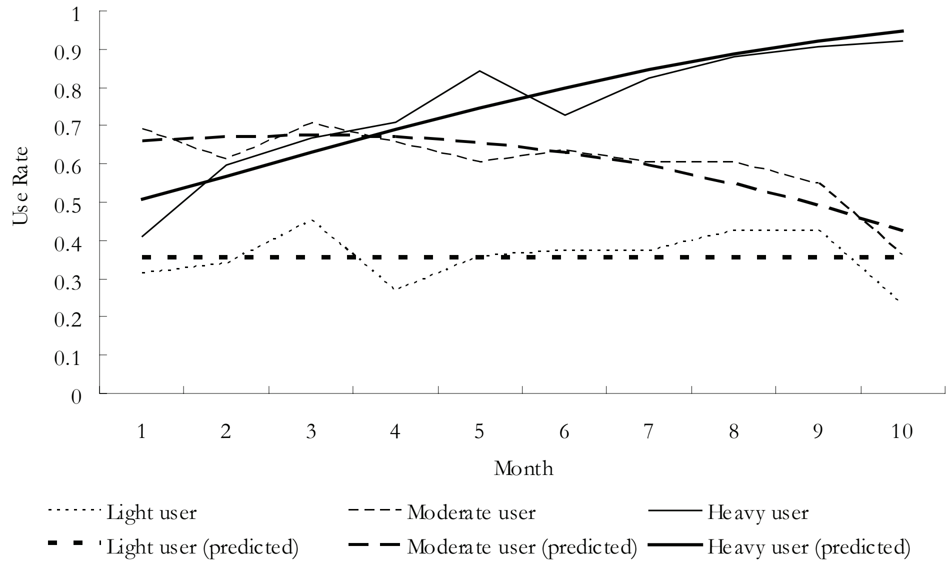


Table 1. Group profiles

User Characteristics A	ll	User Groups		
		Light M	oderate H	eavy
Female (%)	43.90	41.2	33.3	66.7
Male (%)	56.1	58.8	66.7	33.3
Non-U.S. citizen (%)	48.78	23.5	66.7	66.7
U.S. citizen (%)	51.22	76.5	33.3	33.3
Number of appointments *	14.04	14.13	14.12	13.75
Number of visits *	8.44 8	.50 8	.67 7	.95
Number of visits w/ system use *	4.53 3	.02 5	.45 5	.87
Computer use score	33.54	33.0	33.56	34.13
Computer knowledge score	32.96	36 **	31.33	31.38
Computer optimism score	53.19	48.56	52.78	58.88 ***

* Monthly average
** Significantly higher than that of the moderate user group at .05 level
*** Significantly higher than that of the light user group at .001 level

members (Nagin, 1999). It has provided valuable insights into studying physical aggression among youth (Nagin & Tremblay, 1999, 2001) and Web utilization and saturation patterns (Christ, Krishnan, Nagin, Kraut, & Guenther, 2001).

A "developmental trajectory" describes the course of a developmental behavior over age or time. DTA assumes that the population is composed of a mixture of distinct groups defined by their developmental trajectories. This method is useful for modeling unobserved heterogeneity in a population, where trajectories vary greatly across population subgroups both in terms of the level of behavior at the outset of the measurement period and in the rate of growth or decline over time. DTA also provides a multivariate procedure to examine impacts of individuals' characteristics on group membership probability, so the influence of the social-cultural factors, such as gender differences, can be evaluated.

Figure 2. Mean group membership probabilities for gender and citizenship

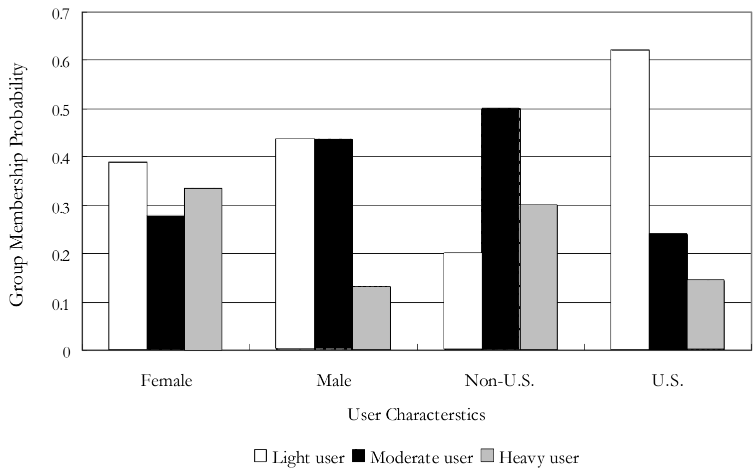


Table 2. The impact of user characteristics on the probability of assignment to usage groups

Variable condition	User Group		
	Light	Moderate	Heavy
Multinomial logit coefficients			
Gender - female	–	-.56 (-0.62)	.27 (0.24)
Non-U.S. citizen	–	1.67 (1.87)	.68 (0.6)
Computer use	–	.045 (.39)	.23 (1.46)
Computer knowledge	–	-.09 (-.985)	-.06 (-.546)
Computer optimism	–	.31 (2.18)	.98 (1.59)
Predicted membership probabilities based on multinomial logit model coefficient estimates			
Without user characteristics	.49	.36	.15
Gender - female only	.54	.27	.19
Gender - male only	.46	.41	.12
Non-U.S. citizen only	.29	.57	.14
U.S. citizen only	.63	.22	.15

RESULTS

The Trajectories

The developmental trajectories obtained from usage data are depicted in Figure 1. Bold and light lines denote observed and predicted trends, respectively. Observed data values are computed as the mean use rate of users assigned to each of these groups identified by estimation, and expected values are computed using DTA model coefficient estimates. We label these three groups as “Light” users, “Moderate” users and “Heavy” users, each comprising 17, 15 and 9 individuals (41.46%, 36.59% and 21.95%, respectively, of all users).

Figure 1 also reveals developmental trends of the adoption behavior. Users classified as “Light” initially utilized the system in about 35% of their encounters, and this rate remained steady over the 10-month study period. “Moderate” users had the highest initial use rate, about 70%, but this rate consistently decreased over the study period to a level comparable with that of the “Light” users. “Heavy” users had an initial use rate of approximately 50%, and this rate increased consistently to about 100% at the end of the study period. Changes in the use rate for members of the moderate group are of particular interest, since it indicates “Moderate” users demonstrated strong enthusiasm initially, while followed by a gradual decline in their use of the system.

Group Profiling

We now examine the association between adoption behavior and user characteristics to identify potential causal relations. Group profiles are shown in Table 1. For example, female users account for two-thirds of the heavy user group, whereas male residents account for two-thirds of the moderate user group. The lower portion of Table 1 shows real-scale measures. On average, each resident saw 8.44 patients each month during the study period (standard deviation 1.89); the maximum count of daily patient encounters is five, but this only occurred a few times. Since the frequency of patient encounters, monthly or daily, was low and relatively evenly distributed among residents, we do not deem this variable useful and will exclude it from further analysis.

The last three rows of Table 1 present the residents’ computer literacy assessments obtained from Cork’s instrument, measuring physicians’ use of, knowledge about and attitudes toward computers (Cork, Detmer, & Freidman, 1998). These three items can be roughly interpreted as how often physicians currently use computers, how much they know about computers, and their relevant beliefs and attitudes. As indicated by Table 1, “Light” users have the highest computer knowledge score and “Heavy” users have the highest computer use and computer optimism scores.

Figure 2 shows the group membership probabilities for each of the categorical variables. For example, male users tend to cluster in the light and the moderate user groups, and non-U.S. citizens are more likely to be present in the moderate user group, whereas U.S. citizens are more likely to be present in the light user group.

Group profiles (see Table 2) are a collection of univariate contrasts. A multivariate procedure is provided in the DTA model to construct a more parsimonious list of predictors to sort out redundant variables as well as to control for potential confounds. Table 2 shows the impact of the covariates of interest on group membership probabilities. The upper panel shows coefficient estimates and t-statistics. We use the light user group as a *contrast group*. That is, for the light user group, the impacts of the covariates are set to zero, and the coefficient estimates for other groups are interpreted as measuring the impact of the covariates on group membership probabilities relative to that of the contrast group.

Table 2 shows the impact of gender on group membership probabilities. If a user is female, the probability of her being in the heavy user group is increased while the probability of her being in the moderate user group is decreased, compared to the probability of membership in the light user group. The lower panel of Table 2 shows the predicted membership probabilities based on multinomial logit model coefficient estimates. For each row, the impact of a single factor is evaluated. Being a female alone increases the probability of membership in the heavy user group. Similarly, having non-U.S. citizenship alone decreases the probability of membership in the light user group and increases the probability of being in the moderate user group. Impact of computer literacy also emerge: A higher score on previous computer use or computer optimism increases the probability of being in the heavy user group, and a higher score on computer optimism increases the probability of being in the moderate user group. In contrast, computer knowledge does not seem to affect usage significantly.

CONCLUSION

In this study, we assess 41 medical residents' acceptance and adoption of a clinical reminder system for chronic disease and preventive care management in an ambulatory care environment. We use a novel developmental trajectory approach to identify distinct groups, following distinct usage trajectories, among those who recorded use of the reminder system within an evaluation period of 10 months. We find that users in this study can be clustered into three groups: "Light" users, who used the system steadily over time for about 35% of their patient encounters; "Moderate" users, whose initial use rate was the highest (70%) among all groups but declined steadily to a level comparable with that of "Light" users; and "Heavy" users, whose use rate, initially moderate (50%), increased to nearly 100% at the end of the evaluation period.

We find that several user characteristics are correlated with usage levels and group membership probabilities. Females are more likely to be frequent users, while males demonstrate the other extreme. Female residents' tendency to comply with system use may be explained by social influence processes, such as subjective norm and their inclination to follow orders. Besides the gender influence, citizenship and attitude towards computers also play a role: Non-U.S. citizens are more likely to be frequency users than U.S. citizens, and computer optimism, rather than previous use and knowledge of computers, increases a user's probability of being a frequent user.

With awareness of such potential causal relations, researchers and practitioners can utilize user characteristics as a means of predicting future levels of usage and adoption. Gender-differentiated training programs and other just-in-time intensive strategies also can be developed and implemented to accomplish higher levels of technology acceptance and use and to minimize undesirable adoption behaviors.

REFERENCES

Adams, D. A., Nelson, R. R., & Todd, P. A. (1992). Perceived usefulness, ease of use and usage of information technology: A replication. *MIS Quarterly*, *16*(2), 227-247.

Berner, E. S., & Maisiak, R. S. (1999). Influence of case and physician characteristics on perceptions of decision support systems. *Journal of American Informatics Associatio*n, *6*, 428-434.

Boulis, A. K., & Long, J. A. (2004). Gender differences in the practice of adult primary care physicians. *Journal of Women's Health*, *13*(6), 703-712.

Chaytors, R. G., Szafran O., & Crutcher R. A. (2001). Rural-urban and gender differences in procedures performed by family practice residency graduates. *Family Medicine*, *33*(10), 766-771.

Chin, W., & Todd, P. A. (1995). On the use, usefulness, and ease of use of structural equation modeling in MIS research: A note of caution. *MIS Quarterly*, *19*(2), 237-246.

Christ, M., Krishnan, R., Nagin, D. S., Kraut, R., & Guenther, O. (2001). Trajectories of individual Web usage: Implications for electronic usage. *Proceedings of the 34th Hawaii International Conference on System Science* (pp. 212-216), Island of Maui, Hawaii.

Cork, R. D., Detmer, W. M., & Freidman, C. P. (1998). Development and initial evaluation of an instrument to measure physicians' use of, knowledge about, and attitudes toward computers. *Journal of American Medical Informatics Association*, *5*,164-76.

Curtin, K., Hayes, B. D., Holland, C. L., & Katz, L. A. (1998). Computer-generated intervention for asthma population care management. *Effective Clinical Practice*, *1*, 43-46.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319-340.

Gefen, D., & Straub D. W. (1997). Gender differences in the perception and use of e-mail: An extension to the technology acceptance model. *MIS Quarterly*, *21*(4), 389-400.

Hofstede, G. (1980). *Culture's consequences: International differences in work related values*. London: Sage.

Hunt, D. L., Haynes, R. B., Hanna, S. E., & Smith, K. (1998). Effects of computer-based clinical decision support systems on physician performance and patient outcomes. *Journal of American Medical Association*, *280*,1339-1346.

Kaplan, B. (2001). Evaluating informatics applications—Clinical decision support systems literature review. *International Journal of Medical Informatics*, *64*, 15-37.

Kaplan, B. (2001). Evaluating informatics applications—Some alternative approaches: Theory, social interactionism, and call for methodological pluralism. *International Journal of Medical Informatics*, *64*, 39-56.

Kraut, R., Scherlis, W., Mukhopadhyay, T., Manning, J., & Kiesler, S. (1996). The HomeNet field trial of residential Internet services. *Communications of the ACM*, *39*, 55-65.

Ling, R. (2000). The use of mobile telephony among Norwegian youth. *Information Technology and People*, *13*(2), 102-120

McDonald, C. J., Hui, S. L., Smith, D. M., Tierney, W. M., Cohen, S. J., Weinberger, L., et al. (1995). Reminders to physicians from an introspective computer medical record. *Annals of Internal Medicine*, *100*, 130-138.

Morahan-Martin, J., & Schumacher, P. (1997). Gender differences in Internet usage, behaviors and attitudes among undergraduates. In G. Salvendy (Ed.), *HCI International '97: Proceedings of the 7^{th} International Conference on Human-Computer Interaction* (p. 122), San Francisco, CA. Amsterdam: Elsevier Science.

Nagin, D. S. (1999). Analyzing developmental trajectories: A semi-parametric, group-based approach. *Psychological Methods*, *4*, 139-157.

Nagin, D. S. & Tremblay, R. E. (1999). Trajectories of boys' physical aggression, opposition and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development*. *70*, 1181-1196.

Nagin, D. S., & Tremblay, R. E. (2001). Analyzing developmental trajectories of distinct but related behaviors: A group-based method. *Psychological Methods*, *6*, 18-34.

Qureshi, S., & Hoppel, C. (1995). Profiling computer predispositions. *Journal of Professional Services Marketing*, *12*, 73-83.

Sackett, D. L., Rosenberg, M. C., Muir Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *British Medical Journal*, *312*, 71-72.

Straub, D., Limayem, M., & Karahanna-Evaristo, E. (1995). Measuring system usage: Implications for IS theory testing. *Management Science*, *141*, 1328-1342.

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, *46*(2), 186-204.

Wilder, G., Mackie, D., & Cooper, J. (1985). Gender and computers: Two surveys of computer-related attitudes. *Sex Roles*, *13*, 215-228.

KEY TERMS

Clinical Decision Support System (CDSS): A class of computer applications typically designed to integrate a medical knowledge base, patient data and an inference engine to generate case-specific advice. They intend to support clinical diagnosis and treatment plan processes and promote use of best practices, condition-specific guidelines and population-based management.

Clinical Reminder System (CRS): A computer system developed at Carnegie Mellon University and The West Penn Allegheny Health System. This system uses patients' medical history to provide just-in-time reminders to clinicians at the point of care that reflect evidence-based medical guidelines.

Developmental Trajectory Analysis: A semi-parametric, group-based approach for identifying distinctive groups of individual trajectories within the population and for profiling the characteristics of group members.

Evidence-Based Medicine: The distillation of a large volume of medical research and standards into treatment protocols for diseases and preventive care procedures that represent the most accurate knowledge available.

Health Informatics: A field of study concerned with the broad range of issues in the management and use of biomedical information, including medical computing and the study of the nature of medical information itself.

Information Technology Adoption: A micro process that focuses on the stages through which an individual passes when deciding to accept or reject the innovation of an information technology.

Medical Guideline: A document with the aim of guiding decisions and criteria in specific areas of health care, as defined by an authoritative examination of current evidence (evidence-based medicine). In the U.S., the National Guideline Clearinghouse publishes and maintains guidelines.

Random Control Trails (RCT): In RCT, study participants are assigned at random to either a control or treatment group to avoid bias and assure more objective results. People in the treatment group receive the new drug or are treated with the new device; the control group participants receive a more standard treatment, no treatment or a placebo (a look alike that has no active drug).

Technology Acceptance Model (TAM): An information systems theory that models how users come to accept and use a technology. The model suggests that when users are presented with a new software package, perceived usefulness and perceived ease-of-use are two primary factors that influence their decision about how and when they will use it.

This work was previously published in Encyclopedia of Gender and Information Technology, edited by E. Trauth, pp. 514-521, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 5.23
Social Construction of Gender and Sexuality in Online HIV/AIDS Information

Jing Chong
The Pennsylvania State University, USA

Lynette Kvasny
The Pennsylvania State University, USA

INTRODUCTION

HIV (human immunodeficiency virus) and AIDS (acquired immunodeficiency syndrome) represent a growing and significant health threat to women worldwide. According to the United Nations (UNAIDS/WHO, 2004), women now make up nearly half of all people living with HIV worldwide. In the United States, although males still accounted for 73% of all AIDS cases diagnosed in 2003, there is a marked increase in HIV and AIDS diagnoses among females. The estimated number of AIDS cases increased 15% among females and 1% among males from 1999 through 2003 (Centers for Disease Control and Prevention, 2003). Looking closer at HIV and AIDS infections among women in the United States, Anderson and Smith (2004) report that HIV infection was the leading cause of death in 2001 for African-American women aged 25 to 34 years, and was among the four leading causes of death for African-American women aged 20 to 24 and 35 to 44 years, as well as Hispanic women aged 35 to 44 years. The rate of AIDS diagnoses for African-American women (50.2 out of 100,000 women) was approximately 25 times the rate for white women (2 out of 100,000) and 4 times the rate for Hispanic women (12.4 out of 100,000; Centers for Disease Control and Prevention). African-American and Hispanic women together represented about 25% of all U.S. women (U.S. Census Bureau, 2000), yet they account for 83% of AIDS diagnoses reported in 2003 (Centers for Disease Control and Prevention).

Women's vulnerability to HIV and AIDS may be attributed to gender inequalities in socioeconomic status, stereotypes of AIDS as a gay-male disease, and cultural ideology around sexual practices such as abstinence, monogamy, and condom use. Because of cultural mores and socioeconomic disadvantages, women may consequently have less access to prevention and care resources. Information is perhaps the most important HIV and AIDS resource for women, and the Internet provides a useful platform for disseminating information to a large cross-section of women. With the flourishing use of e-health resources and the growing number of public-access Internet sites, more and more people are using the Internet to obtain health-care information. Over two thirds of Americans (67%) are now online (Internet World Statistics, 2005). On a typical day, about 6 million Americans go online for medical advice. This exceeds the number of Americans who actually visit health professionals (Fox & Rainie, 2002). Studies also show that women are more likely to seek health information online than are men (Fox & Fallows, 2003; Fox & Rainie, 2000; Hern, Weitkamp, Hillard, Trigg, & Guard, 1998). HIV and AIDS patients are among the health-care consumers with chronic medical conditions who increasingly take the Internet as a major source of information (Kalichman, Weinhardt, Benotsch, & Cherry, 2002).

As more Americans go online for health information, the actual efficacy of the information consumption becomes salient. Recent digital divide studies call for shifting from demographic statistics around technological access to socially informed research on effective use of technology (Gurstein, 2003; Hacker & Mason, 2003; Kvasny & Truex, 2001; Payton, 2003; Warschauer, 2002). Although the Internet provides a health information dissemination platform that is continuous, free, and largely anonymous, we should not assume that broader access and use will be translated into positive benefits. We must begin to critically examine the extent to which e-health content meets the needs of an increasingly diverse population of Internet users.

To combat the AIDS pandemic, it is necessary to deliver information that is timely, credible, and multisectoral. It has to reach not just clinicians and scientists, but also behavioral specialists, policy makers, donors, activists, and industry leaders. It must also be accessible to affected individuals and communities (Garbus, Peiperl, & Chatani, 2002). Accessibility for affected individuals and communities would necessitate targeted, culturally salient, and unbiased information. This is a huge challenge. For instance, health providers' insensitivity and biases toward women have been documented in the critical investigation of TV programs (Myrick, 1999; Raheim, 1996) and printed materials (Charlesworth, 2003). There is a lack of empirical evidence to demonstrate the extent to which and the conditions by which these biases are reproduced on the Internet. In what follows, we provide a conceptual framework for uncovering implicit gender biases in HIV and AIDS information. This framework is informed by the role of power in shaping the social construction of gender and sexuality. We conclude by describing how the framework can be applied in the analysis of online HIV and AIDS information resources.

BACKGROUND

Gupta (2000) has explored the determining role of power in gender and sexuality. Gender, according to Gupta, concerns expectations and norms of appropriate male and female behaviors, characteristics, and roles shared within a society. It is a social and cultural construct that differentiates women from men and defines the ways they interact with each other. Distinct from gender yet intimately linked to it, sexuality is the social construction of a biological drive, including whom to have sex with, in what ways, why, under what circumstances, and with what outcomes. Sexuality is influenced by rules, both explicit and implicit, imposed by the social definition of gender, age, economic status, ethnicity, and so forth (Dixon Mueller, 1993; Zeidenstein & Moore, 1996).

What is fundamental to both sexuality and gender is power. The unequal power balance in gender relations that favors men translates into an unequal power balance in heterosexual interactions. Male pleasure supersedes female pleasure, and men have greater control than women over when, where, and how sex takes place (Gavey, McPhillips, & Doherty, 2001). Therefore, gender and sexuality must be understood as constructed by a complex interplay of social, cultural, and economic forces that determine the distribution of power. As far as HIV and AIDS, "the imbalance in power between women and men in gender and sexual relations curtails women's sexual autonomy and expands male sexual freedom, thereby increasing women's and men's risk and vulnerability to HIV" (Gupta, 2000, p. 2; Heise & Elias, 1995; Weiss & Gupta, 1998).

Based on this feminist approach to theorizing gender and sexuality, Gupta (2000) categorized HIV and AIDS programs in terms of the degree to which historical power dynamics in gender and sexuality were maintained. The categories summarized in Table 1 are depicted in Figure 1 ranging from the most damaging to the most beneficial ones.

In the theory of social construction, HIV and AIDS are represented as a set of social, economic, and political discourses that are transmitted by media (Cullen, 1998). In symbolic interactionism's theory of gender, mediated messages in advertising, TV, movies, and books tell quite directly how gender is enacted (Ritzer, 1996). As the latest platform for computer-mediated communication, the Internet may also adhere to these gendered representations. We theorize that online HIV and AIDS information follows a similar pattern of power reconstruction, and that these categories could be applied to empirically determine how and why online HIV and AIDS information reproduces these power relations.

FUTURE TRENDS

This theoretical framework could be employed in empirical studies that deconstruct online materials to demonstrate how HIV and AIDS gain their social meanings at the intersection of discourses about gender and sexuality. Prior studies in this area have focused on the cultural analyses of AIDS (Cheng, 2005; Sontag, 1990; Treichler, 1999; Waldby, 1996) rather than structural determinants of risk such as political policy, globalization, industrialization, and the economy. Cultural analysis is based upon the belief that this disease operates as an epidemic of signification based on largely predetermined sexual and gendered conventions. The female has now become socially constructed as a body under siege in AIDS discourse. This gendered body is not, however, a stable signifier. Previously, the body was constructed as white, gay, and male. Now the global discourses on HIV and AIDS have constructed the body as third world, heterosexual, and female. Thus, we see a feminization of HIV and AIDS.

Analysis of the social construction of AIDS using this framework could occur at different levels of analysis and with various populations. We conclude with a few examples.

- Garbus et al. (2002) provide a categorization of HIV and AIDS Web sites that could be used for a cross-category or within-category analysis of the representation of gender and sexuality.
- Cultural ideologies around condom use for AIDS prevention and reproductive health could be studied.
- Given the wide disparities in HIV and AIDS infections among women in the United States, research is needed to examine the discursive practices surrounding HIV and AIDS, socioeconomic status, geographic region, and ethnicity or race.
- The absence of lesbians in the HIV-AIDS and women discourse can be analyzed.
- The social construction of the female body in the HIV and AIDS discourse can be studied.

Table 1. Categories of HIV and AIDS programs based on gender and sexuality

Category	Description
Stereotypical	The damaging stereotypes of men are reinforced as "predator, violent, irresponsible" and women as "powerless victims" or "repositories of infection".
Neutral	The target is the general population instead of either gender or sex. Despite no harm done and "better than nothing", the different needs of women and men are ignored. Very often the basis is research that only has been tested on men, or works better for men.
Sensitive	The different needs and constraints of individuals based on their gender and sexuality are recognized and responded to. One example is to provide female condoms. Thus women's access to protection, treatment or care can be improved, but little is done to change the old paradigm of imbalanced gender power
Tranformational	The aim is to transform gender relations to make them equitable. The major focus is on the redefinition of gender roles at the personal, relationship, community and societal levels.
Empowering	The central idea is to "seek to empower women or free women and men from the impact of destructive gender and sexual norms". Women are encouraged to take necessary actions at personal as well as community levels to participate in decision-making. One misunderstanding that needs to be corrected is that empowering women isn't equal to disempowering men. The fact is more power to women would eventually lead to more power to both, since empowering women improves households, communities and entire nations.

Figure 1. Continuum of the social construction of gender and sexuality

- Discursive practices surrounding HIV and AIDS, gender, and development in developing countries are a potential research subject.
- The tension in the social construction of women as both vulnerable receivers and immoral transmitters of this deadly disease can be deconstructed.

CONCLUSION

HIV and AIDS are a complex and pressing issue. It is not just an issue of health, but has also been framed as an issue of personal responsibility, economics, development, and gender equity. It impacts every nation and individual across the globe. In this article, we argue that the increasing epidemic of HIV and AIDS among women is also an issue of information. We propose a framework for unpacking discursive practices that construct women as the new face of HIV and AIDS. We also provide examples of problem domains in which the feminist analysis informed by this framework can be conducted.

REFERENCE

Anderson, R. N., & Smith, B. L. (2003). *Deaths: Leading causes for 2001. National Vital Statistics Reports, 52*(9), 32-33, 53-54. Retrieved October, 2005, from http://www.cdc.gov/nchs/data/nvsr/nvsr52_09.pdf

Bull, S. S., McFarlane, M., & King, D. (2001). Barriers to STD/HIV prevention on the Internet health education research. *ProQuest Psychology Journals, 16*(6), 661-670.

Centers for Disease Control and Prevention. (2003). Cases of HIV infection and AIDS in the

United States. *HIV/AIDS Surveillance Report, 15*, 1-46. Retrieved February 10, 2005, from http://www.cdc.gov/hiv/stats/hasrlink.htm

Charlesworth, D. (2003). Transmitters, caregivers, and flowerpots: Rhetorical constructions of women's early identities in the AIDS pandemic. *Women's Studies in Communication, 26*(1), 60-87.

Cheng, S. (2005). Popularising purity: Gender, sexuality and nationalism in HIV/AIDS prevention for South Korean youths. *Asia Pacific Viewpoint, 46*(1), 7-20.

Cullen, J. (1998). The needle and the damage done: Research, action research, and the organizational and social construction of health in the "information society." *Human Relations, 51*(12), 1543-1564.

Dixon Mueller, R. (1993). The sexuality connection in reproductive health. *Family Planning, 24*(5), 269-282.

Fox, S., & Fallows, D. (2003). Internet health resources: Health searches and email have become more commonplace, but there is room for improvement in searches and overall Internet access. *Pew Internet & American Life Project: Online Report.* Retrieved July 18, 2003, from http://www.pewinter net.org/reports/toc.asp?Report=95

Fox, S., & Rainie, L. (2000). The online health care revolution: How the Web helps Americans take better care of themselves. *Pew Internet & American Life Project: Online Report.* Retrieved November 28, 2000, from http://www.pewinternet.org/reports/pdfs/PIPHealthReport.pdf

Fox, S., & Rainie, L. (2002). Vital decisions: How Internet users decide what information to trust when they or their loved ones are sick. *Pew Internet and American Life: Online Report.* Retrieved April 2005 from http://www.pewinternet.org/PPF/r/59/report_display.asp

Garbus, L., Peiperl, L., & Chatani, M. (2002). AIDS in the digital age. *British Medical Journal, 324*(7331), 246.

Gavey, N., McPhillips, K., & Doherty, M. (2001). "If it's not on, it's not on": Or is it? Discursive constraints on women's condom use. *Gender and Society, 15*(6), 917-934.

Gupta, G. R. (2000, July 12). Gender, sexuality, and HIV/AIDS: The what, the why, and the how (Plenary Address). *Proceedings of the 13th International AIDS Conference,* Durban, South Africa (pp. 1-8).

Gurstein, M. (2003). Effective use: A community informatics strategy beyond the digital divide. *First Monday, 8*(12). Retrieved May 2005 from http://www.firstmonday.org/issues/issue8_12/gurstein/

Hacker, K., & Mason, S. (2003). Ethical gaps in studies of the digital divide. *Ethics and Information and Technology, 5*(2), 99-115.

Heise, L., & Elias, C. (1995). Transforming AIDS prevention to meet women's needs: A focus on developing countries. *Social Science and Medicine, 40*(7), 933-943.

Hern, M. J., Weitkamp, T., Hillard, P. J. A., Trigg, J., & Guard, R. (1998). Promoting women's health via the World Wide Web. *Journal of Obstetric, Gynecologic, and Neonatal Nursing, 27*(6), 606-610.

Internet World Statistics. (2005). *Internet usage statistics for the Americas.* Retrieved May 2005 from http://www.internetworldstats.com/stats2.htm

Kalichman, S. C., Weinhardt, L., Benotsch, E., & Cherry, C. (2002). Closing the digital divide in HIV/AIDS care: Development of a theory-based intervention to increase Internet access. *AIDS Care, 14*(4), 523-537.

Kvasny, L., & Truex, D. (2001). Defining away the digital divide: A content analysis of institutional influences on popular representations of technology. In B. Fitzgerald, N. Russo, & J. DeGross (Eds.), *Realigning research and practice in information systems development: The social and organizational perspective* (pp. 399-415). New York: Kluwer Academic Publishers.

Myrick, R. (1999). Making women visible through health communication: Representations of gender in AIDS PSAs. *Women's Studies in Communication, 22,* 45-65.

Payton, F. C. (2003). Rethinking the digital divide. *Communications of the ACM, 46*(6), 89-91.

Raheim, S. (1996). The reconstruction of AIDS as a women's health issue. In R. L. Parrott & C. M. Condit (Eds.), *Evaluating women's health messages: A resource book* (pp. 402-413). Thousand Oaks, CA: Sage Publications.

Ritzer, G. (1996). *Modern sociological theory* (4th ed.). New York: The McGraw-Hill Companies, Inc.

Sontag, S. (1990). *Illness as metaphor and AIDS and its metaphors.* New York: Farrar Straus and Giroux.

Treichler, P. (1999). *How to have theory in an epidemic: Cultural chronicles and AIDS.* Durham, NC: Duke University Press.

UNAIDS/WHO. (2004). *AIDS epidemic update: 2004*. Retrieved May 2005 from http://www.unaids.org/wad2004/EPI_1204_pdf_en/EpiUpdate04_en.pdf

U.S. Census Bureau. (2000). *Census brief: Women in the United States: A profile*. Retrieved March, 2000, from http://www.census.gov/prod/2000pubs/cenbr001.pdf

Waldby, C. (1996). *AIDS and the body politic.* London: Routledge.

Warschauer, M. (2002). Reconceptualizing the digital divide. *First Monday, 7*(4). Retrieved from http://www.firstmonday.org/issues/issue7_7/warschauer/index.html

Weiss, E., & Gupta, G. R. (1998). *Bridging the gap: Addressing gender and sexuality in HIV prevention.* Washington, DC: International Center for Research on Women.

Weiss, E., Whelan, D., & Gupta, G. R. (2000). Gender, sexuality and HIV: Making a difference in the lives of young women in developing countries. *Sexual and Relationship Therapy, 15*(3), 233-245.

Zeidenstein, S., & Moore, K. (Eds.). (1996). *Learning about sexuality: A practical beginning.* New York: Population Council.

KEY TERMS

Digital Divide: Unequal access to and use of computers and the Internet resulting from such socioeconomic gaps as income, education, race, and age.

E-Health: The applications of the Internet and global networking technologies to medicine and public health.

Empowerment Theory: The study of how perceptions of power affect behaviors and how individuals can increase their power through social interaction.

Feminist Theory: Women-centered theory that treats women as the central subjects, seeks to see the world from the points of women in the social world, and seeks to produce a better world for women.

Gender: Expectations and norms of appropriate male and female behaviors, characteristics, roles, and ways of interaction that are shared within a society.

Sexuality: Social construction of a biological drive, including whom to have sex with, in what ways, why, under what circumstances, and with what outcomes.

Social Construction of Information: Information is examined not as objective missives, but rather as data inextricably intertwined with the social settings in which they are encountered.

This work was previously published in Encyclopedia of Gender and Information Technology, edited by E. Trauth, pp. 1112-1116, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 5.24
Community-Based Information Technology Interventions for Persons with Mental Illness

Rosanna Tarsiero
University of Pisa, Italy

ABSTRACT

The chapter provides the reader with an overview of the problems persons with mental illness experience in their everyday life, and guides readers through how ICT access and usage can be approached in order to empower such a marginalized population in both developed and developing countries. It argues that, since isolation is their main problem, networking those people with reliable sources of medical information, providers of distance training and learning, and online self-help communities can have a profound impact on lifting their marginalization. The author hopes that the role ICT can play for these people will no longer be overlooked or neglected, and that policymakers will be inspired to use ICT worldwide to defeat mental illness by implementing solutions tailored on these people's needs.

INTRODUCTION

In thinking about information and communication technology (ICT) helping people with disability, what generally comes to the minds of the researchers are assistive technology tools, like haptic devices providing tactile representations for visual stimuli on a display to persons with visual impairments (e.g., Colwell, Petrie, Kornbrot, Hardwick, & Furner, 1998), similar to the principle of Braille. However, physical disabilities are not the only disabilities that can benefit from ICT assistance.

When one thinks of ICT for development, the thought goes to the pioneering work of the United Nations Development Program (UNDP) in information technology for development initiated in 1993 (UNDP, 2001). Since the United Nations (UN) Secretary-General's words: "Communication and information technology have enormous potential, especially for developing countries, and in furthering sustainable development" (Annan, 1997, p. 1), and the annual World Bank report two years later, new perspectives and understandings led to the inclusion of widespread availability of ICTs into the Millennium Development Goals (UN, 2000b). In order for ICT to help generate changes, it needs infrastructures, domestic and external human resources (Mansell & Wehn, 1998). Thus, the UN launched the United Nations Information Technology Service (UNITeS) initiative (UN, 2000a; UNITeS, n.d.), in which external human resources, the United Nations Volunteers (UNV), join domestic people to solve local development problems through ICT (UNV, 2005). Nevertheless, people in developing countries are not the only ones who can benefit from ICT as a means to their daily development, as reported in many studies on ICT interventions. Populations living in rural areas of the developed countries are also in need of ICT interventions for their development (Huggins & Izushi, 2002).

Finally, the thought of ICT and networking for marginalized communities usually leads to thinking of a group of people who share one or more characteristics, values, and goals, and whose members proactively escape marginalization through some technology-mediated interaction with other people and/or through access to some valuable empowering resources (Phipps, 2000).

People with mental illness somehow belong to all of these categories and more. Many of them have a disability, in that they have special needs to be accommodated and fulfilled, mainly pertaining to emotional skills (MacDonald-Wilson, Rogers, Massaro, Lyass, & Crean, 2002). They need resources for a safe, strong, and sustainable development of their own community, in terms of information on illness and benefits, social networks, job opportunities, housing, and so forth. Finally, they are marginalized, not only and not just because mainstream people exclude them, but also due to the illness and its consequences which generate accessional relational impairment, along with chronic external and internalized stigmas.

In North America, studies on ICT services for persons with mental illness mainly focused on mental health records (Puskar, Aubrecht, Beamer, & Carozza, 2004), telemedicine (Gutierrez, 2001), and e-therapy (International Society for Mental Health Online & Psychiatric Society for Informatics, 2000), while in Europe they focused on delivering community-based mental health services (Draper & Rigby, 2000; Rigby, Lindmark, & Furlan, 1998) and enhancing communication between mental health providers and consumers (Castelnuovo, Gaggioli, Mantovani, & Riva, 2003). As far as nonprofit organizations, the most common use of the Internet for activities related to mental health is advocacy (e.g., National Alliance of Mentally Ill in the United States and Mind in the United Kingdom). In developing countries, ICT solutions have been implemented for general health only and with the specific purposes of delivering continuous medical education, telemedicine, and health e-governance (Chandrasekhar & Ghosh, 2001).

It goes beyond the scope of this chapter to examine cultural differences in causative beliefs on mental illness and in its treatment options (for a review of current theories, see Lonner, Dinnel, Hayes, & Sattler, 2004). Because of the cultural appropriateness of community-based treatments and its therapeutic successes in both developing countries (Barrio, 2000) and in developed countries (Simmonds, Coid, Joseph, Marriott, & Tyrer, 2001), and because of the potentiality of ICT to address such interventions worldwide, this chapter will focus on community-based approaches involving ICT in both these settings, highlighting differences whenever encountered. This chapter will try to show how it is possible for people with mental illness to achieve empowerment through community-based ICT interventions, and how ICT can be an appropriate answer as well as a powerful tool in both developed and developing country settings.

For this to happen, however, it needs to carefully address the issue from three unfolding perspectives: (1) psychosocial factors involved in mental illness, (2) ICT intervention design, and (3) strategic planning.

BACKGROUND

Mental health is a main problem even in countries with high rates of infectious disease (McKenzie, Patel, & Araya, 2004). In fact, people with mental and behavioral disorders are about 450 million worldwide, as many as overall Internet users in 2002 (McKenna & Green, 2002). It is calculated that one person in four will develop one or more mental or behavioral disorders during their lifetime (WHO, 2001). These rates clearly dispel the myth that mental illness is due to some sociological issues linked to higher income or education (McKenzie et al., 2004) and that it is just a problem of rich countries.

Low-income countries estimated direct mental illness costs do not reach developed country levels. That is because the lower availability of resources and coverage of mental health care services in low-income countries tend to raise productivity loss and other indirect costs instead, which account for a larger proportion of overall costs (WHO, 2001). Also, in most developing countries, available resources are misallocated, in terms of favoring low-cost-effective provisions, are not equally available to all the population, and are ineffective, while in middle-income developing countries health costs are exploding (World Bank, 1993). Using the *disability-adjusted life years* (DALY) as an indicator of burden due to the illness (World Bank, 1993) is very useful for psychiatric disorders, in that it accounts for both premature deaths due to suicide and years wasted to chronic disability. Nevertheless, even if mental illness outcome does not imply psychiatric disability in the majority of cases, neither in the developed nor in developing countries (Goldberg & Lecrubier, 1995), and people in the latter have a better prognosis for schizophrenia (Hopper & Wanderling, 2000; Sartorius et al., 1986) and mental illness in general (Jilek, 2001), 5 of the 10 actual leading causes of disability and premature death worldwide are still due to psychiatric conditions (WHO, 2004).

While rich countries neglected community-based approaches in favor of more individualistic ones (McKenzie et al., 2004), a joint paper from ILO, UNESCO, and WHO suggested community-based ones to be the best approach to use in low-income countries (1994). More individualistic Western approaches can be dangerous on other ethnicities (Barrio, 2000), while community-based approaches on Westerners are beneficial (Simmonds et al., 2001), and both the European Union (WHO European Ministerial Conference on Mental Health, 2005) and the United States (U.S. General Accounting Office, 2000) are adopting them. Developing countries, focusing on community-based interventions, obtained significant successes, most noteworthy for post-partum depression in Hong Kong and Fiji (Becker & Lee, 2002). However, those studies mainly focused on prevention, which can be successfully exerted in community-based settings without generating personal feeling of stigma, therefore leading to overall lesser drop-out rates.

Even though effective and appropriate, community-based interventions for persons with mental illness still present some difficulties in their actual implementations, mainly because of such community hallmarks: (1) mental illness as main, shared characteristic; (2) marginalization as social determinant; and (3) remitting impairment in communication among members as peculiarity to address.

One of their major problems is how to overcome difficulties with respect to communication among members, because the definition of community does not fully apply in these settings, unless interaction among members is proactively fostered. They share some characteristics (e.g., mental illness, stigmatization, marginalization, low self-esteem), some goals (e.g., stability, recovery, empowerment, self-determination), and under certain conditions (e.g., belonging to the same kind of self-help group) even values and methods to accomplish those goals. Nevertheless, since psychiatric illnesses are invisible and often concealed or denied in fear of stigma, members of the same theoretical community may be unknown to each other. This accounts for the possible lack of interaction, interaction that is essential to create and sustain a functional community of peers with which to share values and pursue common goals.

A second problem of community-based approaches is, paradoxically, the meeting settings. Again in fear of stigma, people with mental illness are tempted not to make use of available face-to-face community-based programs. Regardless of the benefits, enrollment can force disclosure, in that workers for the supporting association and any other person on the street might see them getting to the meeting place. Something similar is known to happen in both developed and developing countries when accessing mental health care (Eapen & Ghubash, 2004) and/or welfare, where just being enrolled generates fear of stigma, to the point of avoiding enrollment or withdrawing from an awarded benefit (Stuber & Kronebusch, 2004).

Addressing community-based interventions through ICT also poses other problems, in terms of ICT, Internet access, and issues arising from it. Depending on countries and actual location within that country, in both developed and developing countries ICT infrastructures, telephones, and power lines can be more (or less) available (ITU, 1997, 2004). Usually, developed countries have a lesser number of under-served areas and less people living in rural areas, while in developing countries, an average of 70% of citizens live away from urbanized areas. In both settings, cities are better served in terms of infrastructures, telephones, and power lines. Since the success of the "Asian Tigers," especially Korea, is largely due to the heavy investment government put in building infrastructures such as Internet backbones and letting the market do the rest (ITU, 2004), some governments of middle-income countries already invested into providing the country with a discrete number of backbone nodes (Mansell & Wehn, 1998).

However, access to ICT is still distributed in a very unequal way between developed and developing countries (Paua, 2003; Pentland, Fletcher, & Hasson, 2004). Within developed countries, there are many smaller digital divides (ITU, 1997), depending on income, education, and location of Internet users (Phipps, 2000). In developing countries, higher costs and lack of infrastructures and reliable continuous electric power (Mansell & Wehn, 1998), costs of hardware equipment, use of a non-English language and non-Westerner cultures sum up with lower income, education, and computer literacy. Higher rates of population living in rural areas (Chandrasekhar & Ghosh, 2001; Paua, 2003) extend the divide.

In developed countries, governments, businesses, and nongovernmental and nonprofit organizations addressed those divides by lowering Internet access fees and/or supporting tax-exempt personal computer purchases (European Council, 2000). These provisions can have an impact on actual Internet user numbers because infrastructure and electric power are already present and reliable in most countries, and even when they are not, those governments usually have or can easily get access to funds for their realization. Furthermore, there are a lesser number of people living in rural zones, average income allows users to purchase required hardware and software, education is high enough to know (or easily learn) how to use ICT, English is spoken as a first or second language, and users belong to the "dominant" Internet culture.

In developing countries, higher relative costs (especially when compared with lower wages), lack of infrastructures, and lack of reliable electric power induced many governments and non-profit organizations to opt for implementing a system of shared accesses rather than personal ones (Pentland et al., 2004). These authors also highlight the widespread use of wireless technology as a possible alternative to shared access, similarly to what happens in rural areas of developed countries, in that wireless is less expensive than copper lines, and cables in general have difficulties reaching remote areas (ITU, 2004). Pentland et al. advocate the use of an asynchronous wireless service like DakNet, implemented in India with success. Rural populations, thanks to a bus with a wireless device, get asynchronous Internet access that allows them to send and receive e-mails, for example with the purpose of interacting with governmental offices and thus saving in high transportation costs.

Aside from infrastructure, in many countries Internet access still has its own, separate costs, and universal Internet access is not a reality yet. In general, access fees are higher in Europe than in North America, because flat rates are less common in Europe. However, both Europe (European Council, 2000) and North America (Greenstein, Lizardo, & Spiller, 1997) enacted access policies for lowering access cost with the aim of reaching rural and low-income areas. A way both developed and developing countries addressed this problem is shared accesses in public places. In developed countries, hospitals, schools, and libraries generally have some provision for citizens to freely access e-government and information. In developing countries, public kiosks have this function (Badshah, Khan, & Garrido, 2004).

Even if all these considerations are necessary beyond any doubt, especially when considering ICT implementations in developing countries, Mansell and Wehn (1998) observed how too often ICT debates are geared on hardware requirements rather than software design and information content. Knowledge, its production, its circulation, and its sharing confer to a society the capability of sustaining development via the Internet (Lundvall, 1992). The appropriateness of those attributes can make all the difference in the demand for ICT applications, more so for people with mental illness for which Saraceno and Barbui (1997) theorized that the service-delivery richness, not service availability in itself, is responsible for better outcomes in developing countries.

Aside from information retrieval, the Internet can be used for online networking, since online networks can use asynchronous communication to promote a sense of community (Schuler, 1996). Community-based ICT has theoretical grounds to be employed for people with mental illness. Online settings are of help because of their remoteness and lack of physical cues. Remoteness can play a pivotal role for different, converging reasons. Most importantly, it eliminates barriers to service fruition, especially the ones consequent to symptoms and medication side effects (e.g., agoraphobia, anhedonia, low energy, dizziness) that most characteristically prevent people with mental illness from venturing outside and interacting with others. In developed countries, person with mental illness can be helped to get low-fee home Internet access, because of widespread connectivity. In developing countries and wherever resources and connectivity are limited, asynchronous wireless communication similar to DakNet is a possible way to allow participation into such online communities, because sending and receiving e-mails does not take much bandwidth. Secondarily, online settings are more disinhibiting (McKenna, Green, & Gleason, 2002; Suler, 2004) and people with mental illness are less afraid of being judged for their appearance and for what they have just disclosed (McKenna & Bargh, 1998). Since it is extremely unlikely for participants in the same online community to also live in the same physical community, advantages of disclosure might be full (for an overview on disclosure, see Ralph, 2002).

Though the findings related to virtual community behavior originated in rich countries, they have been already transposed with success to virtual communities in the developing world (Wagner, Cheung, Lee, & Ip, 2003). According to these authors, online communities in developing countries (in this study, Africa, Armenia, Bangladesh, China, India, and Peru) differ from their counterparts in developed countries because they: (1) are knowledge oriented, (2) require an e-mail hub to save bandwidth, (3) have interaction-generated content, and (4) are highly autonomous in nature. However, these findings were relative to virtual communities generated for e-governance purposes only, ruling out a virtual community as a source of emotional support a priori due to its expensive settings. Though this might be true for illnesses whose treatment is promptly available, because professionals are available and/or because the person does not have problems in reaching out to such professionals, those considerations might not apply to people with mental illness, for which self-help can cut costs in terms of favorable outcome of the illness. It is also important to highlight how the Chinese virtual community, that also served as support, became integrated at the same cost as others, possibly because of the higher value such culture poses on collaborative behavior.

Online self-help groups overcome participation barriers, because they are available in timing and locations in which face-to-face groups are not. If members have home computers and access, such as in developed countries, online groups are available every day of the week for as long as the member wishes to use them. If members access through kiosks, like in developing countries, they are available whenever the kiosk is open.

Physical cues take a heavy toll on contributing to prejudices and stigma. They determine whether a relationship of any kind will develop, just on some external characteristics. Some mental illnesses can be evident from just the way a person dresses, so appearance is a major concern for such people and rejections based on look are not that unusual. Therefore, the lack of physical cues, so characteristic of computer-mediated communication (Sproull & Kiesler, 1991), is a very important tool to foster acceptance, trust, and relationships in marginalized online communities. The richer a media is, in terms of immediate feedback, focus on the recipient, language variety, and communication cues/channels availability (Daft & Lengel, 1986), the more overwhelming it can be, especially to persons with mental illness that already tend to have a lower stimulation threshold. Therefore, an impoverished media could be a consistent advantage for such persons.

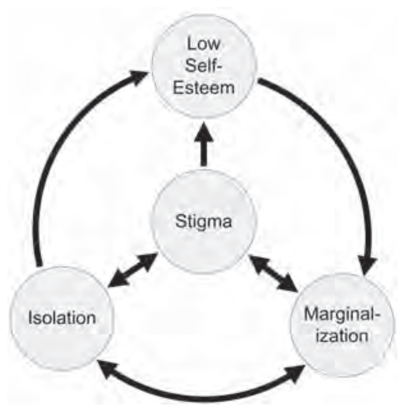
EMPOWERING ESCAPE FROM MENTAL ILLNESS

The Mental Illness Circle

Stigma, marginalization, isolation, and low self-esteem influence each other deeply (Figure 1). Interventions addressing one of these problems, therefore, have repercussions on all the others, but rarely does addressing one problem at a time lead to any resolution. For this reason, it is preferable to implement strategies that simultaneously address each of the problems, even if every one is addressed just partially.

The most common kind of stigma people with mental illness face is the interpersonal one, or

Figure 1. Vicious circle of stigma, low self-esteem, isolation, and marginalization



external stigma. A growing body of literature from developing countries (for a review see Jilek, 2001; for some country profiles see: Hugo, Boshoff, Traut, Zungu-Dirwayi, & Stein, 2003; Karim, Saeed, Rana, Mubbashar, & Jenkins, 2004; Khandelwal, Jhingan, Ramesh, Gupta, & Srivastava, 2004; Raguram, Raghu, Vounatsou, & Weiss, 2004; Regmi, Pokharel, Ojha, Pradhan, & Chapagain, 2004; Shibre et al., 2001) confirms that external stigma is also a problem for people with mental illness located outside the developed world.

However, only five Western psychiatric disorder constructs were found to be applicable transculturally: schizophrenia, brief reactive psychoses, major depression, bipolar disorder, and anxiety-related ones (Ameen, 2002). Constructs like *trauma* are culturally bound syndromes (Eyber, 2002), as well as *koro* (Jilek, 2001). Even disorders found to be transculturally applicable manifest themselves with symptoms (Bhugra & Mastrogianni, 2004; Jilek, 2001) and verbalizations (Draguns, 1990) that still vary with culture, in a process called *pathoplastic*.

Different cultures stigmatize people for different reasons and in different ways. In developed countries, often because of media, persons with mental illness are frequently stigmatized as violent (Angermeyer & Matschinger, 1996; Glasgow University Media Group, 1996) and/or incompetent (Byrne, 1997), while in Nigeria they are regarded as being possessed by an evil spirit (Kabir, Iliyasu, Abubakar, & Aliyu, 2004) and in Haiti as being voodooed (Desrosiers & St Fleurose, 2002).

Also, different mental illnesses carry different degrees of stigma. As an example, Balinese people stigmatize people with obsessive-compulsive disorder much more than people with schizophrenia or depression, while the reverse is true for Japanese (Kurihara, Kato, Sakamoto, Reverger, & Kitamura, 2000), Latin Americans, and Caribbeans (de Toledo Piza Peluso & Blay, 2004).

Finally, within the same illness, different symptoms raise different reactions and concerns, which strongly depend on culture. In India, sadness is stigmatized more than physical complaints, and the difficulty to find somebody to marry because of mental illness is a major concern (Chowdhury et al., 2001), while in Western countries the concern is more about losing a job and inability of performing.

For all these reasons anti-stigma campaigns need to be tailored on cultural determinants of stigma (Angermeyer, Buyangtus, Kenzine, & Matschinger, 2004).

Whichever the set of mainstream assumptions contextual to the culture of people with mental illness live in, they are defined in terms of their disorder. This process of dehumanization creates a behavioral stereotype believed to be truly and accurately reflective of persons with mental illness, who become therefore marginalized through the culture of blame (Farber & Azar, 1999). Also, symptomatic experiences of persons with mental illness are reframed by mental health professionals through a technical jargon that is not always respectful of their feelings and experiential knowledge. This is an implementation of the culture of blame typical of stigma related to mental illness (Burton & Kagan, 2005). Because of past episodes (Hayward, Wong, Bright, & Lam, 2002), acute symptoms, medication side-effects, labeling due to diagnosis (in some cultures), changes in social status, and feelings of disempowerment, persons with mental illness internalize the stigma, which leads them to think to be the way the behavioral stereotype defines them to be. This internalized belief results in giving up on hope for the future and in feelings of impossibility to change their lives to attain people's respect and self-respect. According to Dinos, Stevens, Serfaty, Weich, and King (2004), stigma also influences acceptance of the diagnosis, adherence to treatment, and even daily functioning.

Regardless of causal mechanisms differing from developed to developing countries, there exists a very well-known association between poverty and mental illness (Saraceno & Barbui, 1997). Mental illness can induce economical marginalization, through disability and job loss, but also social marginalization, due to incarceration, stigma, symptoms, and low income. In developed countries, the lack of community ties and support is considered to be responsible for the worse outcome mental illness tends to have (Jilek, 2001). In developing countries, where societies are not so much fragmented and attention to social relationships is higher because of poverty (Payne, 2001), outcomes tend to be better (Sartorius et al., 1986). Marginalization prevents excluded people with mental illness from full personal, interpersonal, and social lives (Burton & Kagan, 2005), thereby lowering their self-esteem even further. Social status, to which marginalization is partly linked, shifts according to degree of activity and type of illness. Marginalization itself can lead to low self-esteem and isolation other than disempowerment (Burton & Kagan, 2005). Marginalization due to joblessness, low-income, lower education, and disability also lowers connectivity in developed countries like the United States (Pew Internet and American Life Project, 2003). Limited access to resources also diminishes greatly the opportunity of exiting this vicious cycle and creating meaningful relationships.

Internalized stigma and the shame elicited by symptoms are reinforced by lack (or unavailability) of accurate de-stigmatizing information and resources on the illness and its course. That leads to possible symptom hiding in the attempt to avoid facing a hopeless illness that, in turn, enhances stress levels and worsens the illness itself. A recrudescence of the illness endangers the characteristics of available social networks, possibly severing such availability. Internalized stigma also correlates negatively with self-esteem (Link, Struening, Neese-Todd, Asmussen, & Phelan, 2001), empowerment, and recovery orientation (Ritsher, Otilingam, & Grajales, 2003). For all these reasons, "existential loneliness" is often the hallmark of the person with mental illness (Nystrom, Dahlberg, & Segesten, 2002).

On a personal level, any illness has the potential to make a person feel defective, more so mental illness, especially in those societies whose fulcrum is mind, intelligence, control, power, performance, and rationality. Though culture can influence the weight such constructs have, self-esteem and life satisfaction are cross-cultural (Diener & Diener, 1995). This, too, endangers pre-existing social networks and these persons' trust in their capability of forming new ties, generating a self-fulfilling prophecy.

There are many kinds of initiatives to attenuate stigma, marginalization, isolation, and low-self esteem. Anti-stigma campaigns are based on the contact hypothesis — that is, people who had previous contacts with persons with mental illness have less stigma. Many campaigns are also based on the biomedical model of mental illness. In both cases, however, results have been contrasting. The contact hypothesis has been confirmed in several studies conducted in Germany (Angermeyer & Matschinger, 1997), in the United States (Roth, Antony, Kerr, & Downie, 2000), and in Russian populations (Angermeyer et al., 2004), but refuted in those conducted in Nigeria (Ohaeri & Fido, 2001), Oman (Al-Adawi et al., 2002), and Hong Kong (Callaghan, Shan, Yu, Ching, & Kwan, 1997). Social distance, defined as unwillingness to get close to people with mental illness, has multiple components, each needing to be addressed (Lauber, Nordt, Falcato, & Rossler, 2004). Nevertheless, attributing mental illness to biological causes seems to be counterproductive, in that in most recent studies it was repeatedly found to increase social distance (Angermeyer, Beck, & Matschinger, 2003; Dietrich et al., 2004). A way to explain these findings is to think the biological metaphor associates persons with mental illness to some degree of unpredictability (Read & Law, 1999) or perceived lack of power over the condition itself. In some settings, the biological model of mental illness can enhance stigmatization by depicting persons with mental illness at par, resulting in patronization. At the same time for some other reasons, the sociological model of mental illness can reduce the sense of personal responsibility and control. Blaming the illness on something external to a person, such as trauma or society, does not address the psychological reaction to the illness and the person's role in reacting against the illness. However, in both the cases, persons with mental illness are reduced to by-products of a chemical imbalance or victims of society malfunctions.

Other ways to provide against stigma, isolation, and low-self esteem are self-help groups, some of which also address resource development and/or mental illness advocacy. The main characteristic of self-help groups is their aim at individual change (Kurtz, 1997, p. 4), and the mechanism by which the change happens is the individual actions to help themselves (Borkman, 1999, p. 4). The self-help movement was born in the United States with Alcoholics Anonymous, but it quickly spread to the rest of the world (Armstrong, 1993). Self-help groups have a defined goal or a set of defined goals (e.g., alcohol moderation for Moderation Management, sobriety with the help of a Higher Power for Alcoholics Anonymous, sobriety without the help of a Higher Power for Rational Recovery, etc). They also have a well-defined method to accomplish their goal that varies from ideology to ideology, each being unique of that group (e.g., a book of Abraham Low's for Recovery Inc., the Big Book for Alcoholic Anonymous, etc). Other than being change-oriented, self-help groups are led by peers, participation in them is voluntary and are free of charge (Kurtz, 1997, p. 4).

In developed countries, the interest in self-help groups was promoted by the high cost of medical services, the lack of alternative to the bio-medical model, and the emphasis on prevention (Reissman & Carroll, 1995). This model can be adopted in developing countries as they lack medical services due to high cost, lack alternatives to traditional healing models, and often focus on prevention as a requirement of mental health assistance programs. So, in 1997, the United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP), in an attempt to improve the condition of persons with disability by supporting the spreading of self-help to developing countries, published the *Guidelines on Establishing and Strengthening Self-Help Organizations of Disabled Persons*. Self-help ideology matches the sustainable development concept because "…the capacity to acquire and generate knowledge in all its forms, including the recovery and *upgrading of traditional knowledge*, is perhaps the most important factor in the improvement of the human condition" (Bezanson & Sagasti, 1995, pp. 5-6). It is useful to remember how the success rate in treating mental illness depends on the *number* of available options, *not* on the kind of options available (Halliburton, 2004). Among the many benefits of self-help groups are: (1) information sharing, (2) engaging in problem solving, and (3) developing social networks. Self-help groups also allow the ability to assist more persons with less money, a useful characteristic for both developing country setting and developed countries with managed care systems (Murray, 1996).

To date, there is no online community-based outreach program that systematically addresses reduction in marginalization, stigma, isolation, and low self-esteem of people with mental illness through the use of ICT and networking. Even if some of the self-help groups mentioned above have an online presence, there is not any online self-help group advocating for this kind of provision to be extended worldwide as first-line community-based intervention for people with mental illness.

Breaking the Cycle Through Community-Based ICT

There is at least one moment in the lives of persons with mental illness in which they come to experience isolation. All of a sudden, they become sick with basically no information on the illness, no idea on where to start searching, no social support, and no available community resource.

Community-based ICT solutions have to take into account possible hurdles in their implementation, mainly: (1) literacy and culture, (2) ideology and culture, and (3) peculiarities of online self-help and culture.

Literacy and culture. Online groups tend to be multicultural, because their members usually join from several countries. Unless the group is a regional one or speaking a language spoken in only limited parts of the world, online communities require higher literacy in participants (in terms of knowing a second language and being fluent in text-based communication). However, multicultural environments have better chances for a positive outcome in helping people with mental illness, because they offer more options than a one-culture-only would.

Ideology and culture. Any self-help group intervention needs to be designed in terms of ideology (approach to the illness and its drawbacks), shared values (kind of interaction among members that is being proactively fostered), and shared goals (ways to overcome everyday problems created by the illness). It is essential for a self-help group to trust members enough to share stories and experiential knowledge in an accepting and respectful climate, so the ideology has to encompass some provisions for encouraging people of all cultures to express themselves.

Peculiarities of online self-help and culture. Self-help is a "holistic approach to people's cultural, economic, and social needs" (Braithwaite, Waldron, & Finn, 1999) that is a strong asset in an intercultural, distributed environment. However, Internet groups differ from face-to-face groups in their dynamics (McKenna & Bargh, 1998, 1999; McKenna & Green, 2002) and present special challenges, because group member fluency, as well as their computer literacy, social condition, formal education, financial status, sex, ethnicity, and age, can vary greatly. Cultural barriers are not immediately apparent and when they become so, they can be surprising to deal with, especially because of the disinhibiting effect of the medium.

Of the many possible community-based ICT interventions, three are worthy of being suggested as starting points: (1) medical information retrieval, (2) online learning and training, and (3) online self-help.

Information retrieval. According to the Pew Internet and American Life Project (2003), more than 8 out of 10 Americans have searched specific information on the net and 93% of those surfers searched medical information at least once. Among medical disorders, 39% of people search information about psychiatric disorders (Pew Internet and American Life Project, 2002). In the developing world, kiosks have the purpose of connecting people to information. People with mental illness are in particular need of online information about their disorder, because even buying books on mental illness in a bookstore can be somewhat stigmatizing, since families and friends of persons with mental illness also get stigmatized, in a process called *courtesy stigma* (Angermeyer, Schulze, & Dietrich, 2003). Furthermore, asking the doctor for the exact diagnosis and explanations connected with it is a behavior that depends on culture and pre-existing patient-doctor relationship (Bensing, Roter, & Hulsman, 2003; Ohtaki, Ohtaki, & Fetters, 2003).

Though most medical information on the Web is available for free, availability does not imply reliability, any more than reliability implies comprehensibility. The first quality initiative, Health On the Net (HON) foundation, was launched in 1995. HON has multiple aims: codifying a set of features that qualifies a site with medical information as reliable, issuing a certification of medical information quality (HON seal), defining a procedure to check on sites that display the HON seal, and doing some sort of *post-marketing surveillance* on certified sites (Nater, Boyer, & Eysenbach, 2000). Though the site assessment is a rigorous process, unfortunately HON certification is released on a voluntarily basis. That means it cannot be imposed on every site with medical information, but just revoked in case of non-compliance with HON Code of Conduct. For a review of all Internet health information quality initiatives, see Risk and Dzenowagis (2001). In searching the Net for medical information, some basic medical literacy skills are required, in order to also discriminate what is evidence-based from what is not (Sackett, Rosenberg, Muir Gray, Haynes, & Richardson, 1996). Medical information is of particular importance for people with mental illness, because psycho-educational interventions are shown effective in mental illnesses management and relapse prevention, especially bipolar disorder (Colom et al., 2003, 2004). Online psycho-education has been recently proven effective as well (Christensen, Griffiths, & Jorm, 2004). Furthermore, from an individual perspective, knowledge is empowering in itself, because it gives people the sensation that they are able to gain control over their lives.

The Internet is also useful in helping people with mental illness to gather information and connect to resources that can assist them in improving their income status (e.g., welfare, cash and energy assistance programs, housing, etc). To show all the importance of such provisions, Thomson, Petticrew, and Morrison (2001) report how housing improvement has a promising impact on physical and mental health outcomes. Many advocacy and governmental sites have useful link collections, nevertheless they rarely explain laws and regulations in layman terms, or offer assistance of some sort in filling out the paperwork. Some states in the U.S. and some European countries have benefit portals by which it is possible to check qualifications in advance, so not to apply uselessly. In developing countries, asynchronous broadband connectivity may ensure this same kind of service (Pentland et al., 2004).

Online training and learning. As Lundvall (1992) stated, "The most fundamental resource in our economy is knowledge" (p. 1). In the United States, over half of surfers used the Net to perform a search somehow related to online training and/or education (Pew Internet and American Life Project, 2003), and improved access to education is another strategy suggested by WHO (2004) to fight mental illness. Online training and/or learning can be a way to include persons with mental illness into a productive society. For this reason, its designers have to know which kind of training content they might need. Macdonald-Wilson et al. (2002) found that the most important functional limitations people with mental illness have regarded both interpersonal and cognitive skills. This implies that the training people with mental illness need more is on soft skills. Online learning, being self-paced and remote, is highly susceptible of accommodation, satisfying the most frequent kinds of accommodation requested by such people: flexible schedule (EEOC, 1997) and interpersonal facilitation with colleagues (Zuckerman, 1993). However, this is not always possible in developing countries, because of bandwidth and hardware limitations (Mansell & Wehn, 1998). Effective online learning environments are learner centered, knowledge centered, assessment centered, and community centered (Bransford, Brown, & Cocking, 1999). That means it takes into account the learning style of the learner but also of the discipline that is being taught, giving many opportunities for assessment, peer-assessment, and self-assessment, without neglecting the online social interactions upon which learning depends (Kreijns, Kirschner, Jochems, & Van Buuren, 2004). Staying at home even when regularly taking classes, asynchronous communication brainstorming and socialization, course material flexibility, and accommodation (in terms of colors, fonts, and formats) allow people with mental illness to better cope with symptoms, and possibly complete course materials and assignments even in conditions under which they would not be able to go to school and perform. More accommodations enhance performance, and that in turn increases self-efficacy (Bandura, 1994) and self-esteem. Online learning is more cost effective (Mansell & Wehn, 1998), especially when it takes cultural differences into account.

Online self-help. According to Moore (2000), social interaction is the primary use of home computers, and e-mail is the most used computer application in the United States (Pew Internet and American Life Project, 2003), as well as in Ghana (Dzidonu, 2004). Surfers do not limit their interactivity to private e-mails. In fact 84% of American Net surfers contacted an online community and 79% stayed in contact with it, while 26% used e-mail to contact a local group (Pew Internet and American Life Project, 2001). Social support and strengthening community networks are strategies recommended by WHO (2004) as primary prevention for mental illness. Creating an online community network is a sustainable target, because many search engines also offer free group membership, and that does not raise costs of this intervention, nor require special software. In order to make it an attainable goal as well, some considerations on online communication and dynamics are mandatory. Online written communication is a hybrid between oral and written communication (Bordia, 1996), more precisely a written substitute for oral communication (Jonsson, 1998). Pennebaker and Seagal (1999) studied written expression of emotions comparing it with talking, showing how they have the same effect, as long as emotions are involved in both cases. By storytelling, even in written form, people are helped in reframing and overcoming their stories. Also, people with invisible stigmatized conditions can benefit more from this intervention if they focus on what makes them different from non-stigmatized populations rather than what makes them similar. Computer-mediated communication is less inhibited than face-to-face (Suler, 2004) and has a hyper-personal effect (Walther, 1996). According to this theory, online communication can exceed face-to-face, especially in that personality cues that would normally go unnoticed assume stronger meanings. Furthermore, the information flow in computer-mediated interaction is slower (Liu, 2002) and less overwhelming, especially to people with mental illness. Having the possibility of thinking more before replying is a definitive advantage for people dealing with impulsivity. Some of these theories could be why computer-mediated communication has such an excellent impact on virtual groups for marginalized people (McKenna & Bargh, 1998). McKenna et al. (2002) also noticed how people experiencing social anxiety, difficulties with social interactions, and isolation prefer to locate more often their "real self" online.

There are a number of reasons why online settings could be an appropriate tool for people with mental illness as far as social support is considered. Because having a solid support network is essential to recovery and the need of socialization is part of the feeling of empowerment, the main problem to address is how to form a stable community for people with mental illness, and which type of community to form. Both support and self-help groups are available alternatives. Self-help groups are more suitable to this task, because their ideology and method are well-defined: theoretical foundations can be tested, effectiveness and outcome evaluated, they can be both reproduced in similar realities and adapted to different settings. On the contrary, support groups often do not have a codified method, rely too much on the individual that leads them, and outcomes are unpredictable and difficult to even detect since there is no expected behavioral change as a consequence of attendance. Peers volunteering to lead such a self-help community have to be trained in that specific ideology. They also have to be trained in online group dynamics and in online community design, because creating community networks cannot be done successfully without considering the community members right from the start (Andrews, Preece, & Turoff, 2001). Group size (Foth, 2003) and dynamics (McKenna & Green, 2002) are important variables to consider for successful community design.

FUTURE TRENDS

Using information technology and networking for the empowerment of people with mental illness worldwide is a very stimulating challenge. For what pertains to people with mental illness in low-income countries, it involves bridging the existing digital divide between developed and developing countries (Paua, 2003). It also involves bridging the many smaller digital divides embedded in the developed societies (Mansell & Wehn, 1998; Paua, 2003). Such a complex process is not only about finding funding sources to bridge those gaps, but it is about allocating resources and using existing technologies and infrastructures in a quite creative way (UNCSTD, 1997).

Some implementation ideas on how to move from this point on can be borrowed from a managerial perspective. In strategic planning, needs and nature of the services to be offered have to be assessed, along with its possible competitors.

Possible interventions for people with mental illness can be roughly divided in two categories: one provided by professionals (like psychiatrists, psychologists, and social workers, more present in developed countries, and Ayurvedic and religious healers, more present in developing countries) and another provided by peers (in the form of self-help or support groups). A third form would be group therapy, in which a professional leads a group of peers. Potential competitors of online self-help groups are face-to-face groups and e-therapy. The former have the same advantages of their online counterparts, save for timing and locations, plus the big advantage for people who are not literate enough to use a computer. In developing countries, where population is mainly rural, transportation costs for going to a self-help face-to-face meeting have to be considered. E-therapy does not have transportation costs, but fees are quite high and it could be culturally inappropriate for people in developing countries. Competitors of medical online information are face-to-face medical information and information following other paradigms of treatments for mental illness. The latter could be culturally more appropriate and therefore could be inserted in community-based ICT informational intervention as one more option to have, thereby enhancing outcome responses (Hulliburton, 2004). Finally, online learning/training's main competitor is an onsite one, which in developing countries is feasible mainly in urbanized areas.

Among risks and pitfalls of community-based ICT approaches, the refusal of this perspective by policymakers because of cultural shock has to be considered. Linguistic analysis of written communication nowadays allows us to distinguish persons with depression or prone to have it from persons without it (Rude, Gortner, & Pennebaker, 2004), persons with suicidal ideation from persons without it (Stirman & Pennebaker, 2001), and even liars from honest people (Newman, Pennebaker, Berry, & Richards, 2003). Therefore, existing tools should be able to understand written communication with much higher reliability than in the past. Though missing of physical and social cues (Sproull & Kiesler, 1991), computer-mediated communication does not lack in nonverbal cues. In fact, time (chronemics, see Walther & Tidwell, 1995) and distance (proxemics, Jeffrey & Mark, 1998; Krikorian, Lee, Makana Chock, & Harms, 2000) are nonverbal cues that do convey, for example, the importance of the conversation topic for a given person. Emoticons are used (Walther & D'Addario, 2001), as well as verbs, to transform equivocal into univocal statements, like for example during jokes. Therefore, the effective use of these tools allows the user to utilize computer-mediated communication in a far richer way than in the past. Nevertheless, many persons, also among information technology professionals, are still polarized on neglecting social communication nuances in favor of task orientation, and interpreting differences between computer-mediated and face-to-face communication as pre-mediated inferiority (or superiority) to one over the other. They often still insist in not modifying their communication style according to the media they use. Working with information technology and networking for empowering people with mental illness teaches both researchers and policymakers how to overcome old myths and rigid constructs. Ultimately, what technology is able to bring us strictly depends on how the medium is being used. Researchers can decide to shape the technology or be shaped by it (Schuler, 1996), and it is up to them to make a responsible choice, never forgetting how the choices will have repercussions on the rest of society, nevertheless.

In order for an information technology organization to be able to lead the change, it also needs a shift in its attitude toward leadership. Just recently, Sosik, Jung, Berson, Dionne, and Jaussi (2004) analyzed the differences between the older *strategic leadership* and the newer *strategy-focused leadership.* In the latter case, leaders do not focus on strategy, but on how to produce an organization that is focused on strategy. They connect people, technology, work processes, and business into a community. This construct has the big advantage of adding social capital to the leadership process. In ICT organizational settings, this is even a more precious skill, especially when dealing with initiatives for developing countries.

Convincing mainstream funders, donors, developers, and colleagues of the theoretical foundations, practical effectiveness, and cost effectiveness of such approaches will be necessary in order to create the required shift to use existing resources, technologies, and infrastructures in a different, more creative way. An excellent way to do so is to validate the theories with research, within an evidence-based framework. Mutated from medicine (Sackett et al., 1996), this approach to implementations and policies are, slowly but increasingly, substituting opinion-based and consensus-based guidelines and becoming *best practices.* With an evidence-based approach, the implementer may be able to have a theory, produce evidence of its appropriateness, and implement reproducible programs based on it. Moreover, the other users will be able to use the same words meaning the same definitions. These concepts will be portable from one situation to another.

In developed countries, information connectivity is already very high, so using this viable and available technology for people with mental illness and scientifically testing the outcome not only has its rationale, but is definitely more sustainable than insisting in reaching consensus-based approaches or in using old opinion-based methods. In developing countries, another reason to favor evidence-based approaches is because resources for marginalized people of any kind are spare and misallocated and cannot be unethically wasted in implementing solutions whose effectiveness is unproven.

In ICT in general, but especially in those programs designed to technologically assist developing countries, the social aspect of interaction is often neglected in favor of a more technology-based approach. Rather than exporting the "developed" approach, they could rather learn how to use the focus on community that is used in other countries. Because connectivity alone does not ensure community (Foth, 2003), the developer's attention has to shift from how to access information to how to *use* information (Lundvall, 1992; Menou, 2001) and assist this shift with careful community design as opposed to mere system design. Factors to take into account when designing a community are: provision of socio-cultural animation, population of the network, attention to sociability, and care for human-to-human ties (Foth, 2003). This is even more important for people with mental illness for whom empowerment, positive personal interactions, social support, and community networks are social protective factors (Hosman, Jané-Llopis, & Saxena, 2005). Therefore, interventions addressing ICT and network usage for people with mental illness have the potentiality to be the primary prevention, thus opening up a further unexplored scenario.

Implementation of online psycho-educational, self-help, and training projects, even in isolation from one another, could test the feasibility of this model and discover its possible pitfalls, with particular regard to gathering and packaging information for psycho-education, searching, and selecting possible old and new self-help models that want to try online delivery of service, training of peers providing online services, feasibility of cheap online training and learning for this population, and usability of online learning materials, also in terms of culture.

Given the high potential for the sustainability and effectiveness of this approach, researchers in the field should focus on: how to build fully functional and effective online self-help groups, how to promote standardization in medical information site user ratings, and how social learning can be integrated into online communities.

CONCLUSION

People with mental illness worldwide are a very large community, often marginalized for concomitant socio-economical factors. Among poorest members, not just access to medical care, but even access to reliable information about the illness, and social and network support — including training resources — are denied, impaired, or otherwise jeopardized by the absence of resources, infrastructures, and economical support to provide reliable Internet access to their population. Illness, poverty, but also lower social status, low self-esteem, stigma, and no hope for a better future make it virtually impossible for them to exit from this vicious marginalized cycle. To persons in these conditions, empowerment sounds like nothing but a word. And it is precisely for this reason that empowerment at personal, interpersonal, and societal levels becomes essential.

People with mental illness might have difficulties interacting with others, both other persons with mental illness and *normal* ones. This may be true whether they are actively symptomatic or not. If symptomatic, the Internet can be their only access to low-cost reliable medical information, as well as providing networking opportunities with other persons with mental illness. If not symptomatic, access to ICT and networking can be a valid answer to many practical problems to escape poverty and marginalization. With a computer and Net access, they can connect with resources of any kind, even when low self-efficacy and self-esteem rather than the illness itself do not allow them to venture into the world. This is especially true for persons with mental illness who also belong to another stigmatized population (e.g., LGBT, minorities, people with physical disabilities, jobless, HIV-positive, low-income, etc.). Therefore, in both developed and developing countries, ensuring low-cost connectivity is the starting point to move further in assisting people with mental illness, whenever literacy is high enough to utilize it.

But even when they have the financial and material resources to buy what they need in order to get connected, like in developed countries, they might not know where they need to go to address their informational, social, and learning needs. Financial provisions to consistently enhance the connectivity of people with mental illness would be a beginning, but that has to be followed by a package of designed resources to acquire information, social networks, and resources that could really make a difference in their life.

Many online interventions could empower them. Ensuring psycho-education, e-learning, and supportive facilitation interventions to people with mental illness is a challenge that will give them capacity and skills to overcome these challenges and participate equally, as they deserve.

In this context, bridging the digital divide has become a rehabilitative challenge, particularly bringing affordable ICT, network access, and sustainable interventions to jointly address stigma, marginalization, isolation, and low self-esteem. Unfortunately, the world cannot be changed as long as the initiators are reluctant to change their mindset, adopting an eclectic approach and being open to paradigm shifts on how to utilize the available resources.

REFERENCES

Al-Adawi, S., Dorvlo, A. S., Al-Ismaily, S. S., Al-Ghafry, D. A., Al-Noobi, B. Z., Al-Salmi, A., et al. (2002). Perception of and attitude towards mental illness in Oman. *International Journal of Social Psychiatry, 48*(4), 305-317.

Alcoholics Anonymous. (2001). *The big book online* (4th ed.). Retrieved November 4, 2004, from http://www.aa.org/bigbookonline/

Ameen, S. (2002). Transcultural psychiatry: A critical review. In S. Ameen (Ed.), *Mental ealth reviews*. Retrieved January 20, 2005, from http://www.psyplexus.com/excl/transcultural_psych.html/

Andrews, D. C., Preece, J., & Turoff, M. (2001). A conceptual framework for demographic groups resistant to online community interaction. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences (HICSS-34)* (Vol. 7, pp. 7013-7023). Retrieved March 30, 2005, from http://csdl.computer.org/comp/proceedings/hicss/2001/0981/07/09817013.pdf

Angermeyer, M. C., Beck, M., & Matschinger, H. (2003). Determinants of the public's preference for social distance from people with schizophrenia. *Canadian Journal of Psychiatry, 48*(10), 663-668.

Angermeyer, M. C., Buyantugs, L., Kenzine, D. V., & Matschinger, H. (2004). Effects of labeling on public attitudes towards people with schizophrenia: Are there cultural differences? *Acta Psychiatrica Scandinavia, 109*(6), 420-425.

Angermeyer, M. C., & Matschinger, H. (1997). Social distance towards the mentally ill: Results of representative surveys in the Federal Republic of Germany. *Psychological Medicine, 27*(1), 131-141.

Angermeyer, M. C., & Matschinger, H. (1996). The effect of violent attacks by schizophrenic persons on the attitude of the public towards the mentally ill. *Social Science and Medicine, 43*(12), 1721-1728.

Angermeyer, M. C., Schulze, B., & Dietrich, S. (2003). Courtesy stigma: A focus group study of relatives of schizophrenia patients. *Social Psychiatry and Psychiatric Epidemiology, 38*(10), 593-602.

Annan, K. (1997, May 12). *Secretary-general stresses international community's objective of harnessing informatics revolution for benefit of mankind. Inter-agency project on universal access to basic communication and information services* (E/CN.16/1997/Misc.3). United Nations Commission on Science and Technology for Development, 3rd session, Item 9, Secretary-General note, p. 1. Geneva: United Nations.

Armstrong, M. J. (1993). Disability self-help organizations in the developing world: A case study from Malaysia. *International Journal of Rehabilitation Research, 16*(3), 185-194.

Badshah, A., Khan, S., & Garrido, M. (Eds.). (2004). *Connected for development: Information kiosks and sustainability* (ICT Task Force Series 4). Retrieved January 26, 2005, from http://www.unicttaskforce.org/wsis/publications/Connected%20for%20Development.pdf

Bandura, A. (1994). Social cognitive theory of mass communication. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research* (pp. 61-90). Hillsdale, NJ: Lawrence Erlbaum.

Barrio, C. (2000). The cultural relevance of community support programs. *Psychiatric Services, 51*(7), 879-884. Retrieved March 30, 2005, from http://ps.psychiatryonline.org/cgi/content/full/51/7/879

Becker, A. E., & Lee, D. T. S. (2002). Indigenous models for attenuation of post partum depression: Case studies from Fiji and Hong Kong. In A. Cohen, A. Kleinman, & B. Saraceno (Eds.), *World mental health casebook: Social and mental programs in low-income countries* (pp. 221-236). New York: Kluwer.

Bensing, J. M., Roter, D. L., & Hulsman, R. L. (2003). Communication patterns of primary care physicians in the United States and The Netherlands. *Journal of General Internal Medicine, 18*(5), 335-342.

Bezanson, K., & Sagasti, F. (1995). *The elusive search: Development and progress in the transition to a new century.* Ottawa, Canada: International Development Research Centre and Lima, Peru: GRADE, mimeo.

Bhugra, D. S., & Matrogianni, A. (2004). Globalisation and mental disorders. *British Journal of Psychiatry, 184*, 10-20. Retrieved March 31, 2005, from http://bjp.rcpsych.org/cgi/content/full/184/1/10

Bordia, P. (1996). Studying verbal interaction on the Internet: The case of rumor transmission research. *Behavior Research Methods, Instruments & Computers, 28*(2), 149-151.

Borkman, T. J. (1999). *Understanding self-help/mutual aid: Experiential learning in the commons.* New Brunswick, NJ: Rutgers University Press.

Braithwaite, D. O., Waldron, V., & Finn, J. (1999). Communication of social support in computer-mediated groups for persons with disabilities. *Health Communication, 11*(2), 123-151.

Bransford, J., Brown, A., & Cocking, R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Burton, M., & Kagan, C. (2004). Marginalization. In I. Prilleltensky & G. Nelson (Eds.), *Community psychology: In pursuit of wellness and liberation* (pp. 293-308). London: Palgrave Macmillan.

Byrne, P. (1997). Psychiatric stigma: Past, passing and to come. *Journal of the Royal Society of Medicine, 90*(11), 618-620.

Callaghan, P., Shan, C. S., Yu, L. S., Ching, L. W., & Kwan, T. L. (1997). Attitudes towards mental illness: Testing the contact hypothesis among Chinese student nurses in Hong Kong. *Journal of Advanced Nursing, 26*(1), 33-40.

Castelnuovo, G., Gaggioli, A., Mantovani, F., & Riva, G. (2003). From psychotherapy to e-therapy: The integration of traditional techniques and new communication tools in clinical settings. *Cyberpsychology and Behaviour, 6*(4), 375-382.

Chandrasekhar, C. P., & Ghosh, J. (2001). Information and communication technologies and health in low income countries: The potential and the constraints. *Bulletin of the World Health Organization, 79*(9), 850-855. Retrieved January 4, 2005, from http://whqlibdoc.who.int/bulletin/2001/issue9/79(9)850-855.pdf

Chowdhury, A. N., Sanyal, D., Bhattacharya, A., Dutta, S. K., De, R., Banerjee, S. et al. (2001). Prominence of symptoms and level of stigma among depressed patients in Calcutta. *Journal of Indian Medical Association, 99*(1), 20-23.

Christensen, H., Griffiths, K. M., & Jorm, A. F. (2004). Delivering interventions for depression by using the Internet: Randomised controlled trial. *British Medical Journal*, *328*(7434), 265-270. Retrieved November 1, 2004, from http://bmj.bmjjournals.com/cgi/content/abridged/328/7434/265

Colom, F., Vieta, E., Reinares, M., Martinez-Aran, A., Torrent, C., Goikolea, J. M. et al. (2003). Psychoeducation efficacy in bipolar disorders: Beyond compliance enhancement. *The Journal of Clinical Psychiatry, 64*(9), 1101-1105.

Colom, F., Vieta, E., Sanchez-Moreno, J., Martinez-Aran, A., Torrent, C., Reinares, M., et al. (2004). Psychoeducation in bipolar patients with comorbid personality disorders. *Bipolar Disorders, 6*(4), 294-298.

Colwell, C., Petrie, H., Kornbrot, D., Hardwick, A., & Furner, S. (1998, April). Haptic virtual reality for blind computer users. In *Proceedings of the 3rd International ACM SIGACCESS Conference on Assistive Technologies* (ASSETS 1998), *4*, 92-99.

Daft, R., & Lengel, R. (1986). Organizational information requirements, media richness and structural design. *Management Science, 32*(5), 554-571.

Desrosiers, A., & St. Fleurose, S. (2002). Treating Haitian patients: Key cultural aspects. *American Journal of Psychotherapy, 56*(4), 508-521.

Diener, E., & Diener, M. (1995). Cross-cultural correlates of life satisfaction and self-esteem. *Journal of Personality and Social Psychology, 68*(4), 653-663.

Dietrich, S., Beck, M., Bujantugs, B., Kenzine, D., Matschinger, H., & Angermeyer, M. C. (2004). The relationship between public causal beliefs and social distance toward mentally ill people. *The Australian and New Zealand Journal of Psychiatry, 38*(5), 348-354.

Dinos, S., Stevens, S., Serfaty, M., Weich, S., & King, M. (2004). Stigma: The feelings and experiences of forty-eight people with mental illness. Qualitative study. *British Journal of Psychiatry, 184*(2), 176-181.

Draguns, J. (1990). Applications of cross-cultural psychology in the field of mental health. In R.W. Brislin (Ed.), *Applied cross-cultural psychology* (pp. 302-324). Newbury Park, CA: Sage Publications.

Draper, R., & Rigby, M. (2000). An ICT solution for community mental health: Putting virtuality into the vision. In A. Hasman, H.-U. Prokosch, B. Blobel, J. Dudeck, G. Gell, & R. Engelbrecht (Eds.), *Studies in Health Technology and Informatics: Vol. 77. 16th Congress Medical Infobahn for Europe* (MIE 2000, pp. 969-973). Amsterdam: IOS Press.

Dzidonu, C. (2004, September). Demand and supply for access and connectivity: The case of Ghana. *Low cost access and connectivity: Local solutions* (United Nations ICT Task Force Working Paper Series, pp. 1-20). New York: United Nations. Retrieved February 15, 2005, from http://195.218.115.39/pport/pdf/301238730.pdf

Eapen, V., & Ghubash, R. (2004). Help-seeking for mental health problems of children: Preferences and attitudes in the United Arab Emirates. *Psychological Reports, 94*(2), 663-667.

Equal Employment Opportunity Commission. (1997, March 25). *EEOC enforcement guidance: The Americans with Disabilities Act and psychiatric disabilities* (EEOC Notice No. 915.002). Washington, DC: U.S. Government Printing Office. Retrieved October 15, 2004, from http://www.eeoc.gov

European Council. (2000, March 23-24). Presidency conclusions (Lisbon European Council 2000). Brussels, Belgium: Author. Retrieved December 23, 2004, from http://www.europarl.eu.int/summits/lis1_en.htm

Eyber, C. (2002, October). The trauma debate. In *Psychosocial issues* (FMO Research Guide, chap. 2). Retrieved January 3, 2005, from http://www.forcedmigration.org/guides/fmo004/fmo004-2.htm

Farber, B. A., & Azar, S. T. (1999). Blaming the helpers: The marginalization of teachers and parents of the urban poor. *American Journal of Orthopsychiatry, 69*(4), 515-528.

Foth, M. (2003). Connectivity does not ensure community: On social capital, networks and communities of place. In S. Marshall & W. Taylor (Eds.), *Proceedings of the 5th International Information Technology in Regional Areas Conference* (ItiRA 2003, pp. 31-39). Rockhampton, Australia: Central Queensland University Press. Retrieved March 30, 2005, from http://itira.cqu.edu.au/proceedings_online/2003/ref_papers/foth.pdf

Glasgow University Media Group. (1996). *Media and mental distress* (G. Philo, Ed.). London: Longman.

Goldberg, D., & Lecrubier, Y. (1995). Form and frequency of mental disorders across cultures. In T. B. Ustun & N. Sartorius (Eds.), *Mental illness in general health care: An international study* (pp. 323-334). Chichester, UK: John Wiley & Sons.

Greenstein, S., Lizardo, M., & Spiller, P. (1997). *The evolution of advanced large scale information infrastructure in the United States* (NBER Working Paper No. 5929). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w5929

Gutierrez, G. (2001). Medicare, the Internet, and the future of telemedicine [Supplement]. *Critical Care Medicine, 29*(8), N144-N150.

Halliburton, M. (2004). Finding a fit: Psychiatric pluralism in south India and its implications for WHO studies of mental disorder. *Transcultural Psychiatry, 41*(1), 80-98.

Hayward, P., Wong, G., Bright, J. A., & Lam, D. (2002). Stigma and self-esteem in manic depression: An exploratory study. *Journal of Affective Disorders, 69*(1), 61-67.

Hopper, K., & Wanderling, J. (2000). Revisiting the developed versus developing country distinction in course and outcome in schizophrenia: Results from ISoS, the WHO collaborative follow-up project. *Schizophrenia Bulletin, 26*(4), 835-846.

Hosman, C., Jané-Llopis, E., & Saxena, S. (Eds.). (2005). *Prevention of mental disorders: Effective interventions and policy options.* Oxford: Oxford University Press.

Huggins, R., & Izushi, H. (2002). The digital divide and ICT learning in rural communities: Examples of good practice service delivery. *Local Economy, 17*(2), 111-122.

Hugo, C. J., Boshoff, D. E., Traut, A., Zungu-Dirwayi, N., & Stein, D. J. (2003). Community attitudes toward and knowledge of mental illness in South Africa. *Social Psychiatry and Psychiatric Epidemiology, 38*(12), 715-719.

ILO, UNESCO, & WHO (1994). Community-Based Rehabilitation, CBR for and with people with disabilities (Joint Position Paper). Geneva: ILO.

International Society for Mental Health Online & Psychiatric Society for Informatics. (2000). Suggested principles for the online provision of mental health services. Retrieved January 6, 2005, from http://ismho.org/suggestions.html

ITU (International Telecommunication Union). (2004, September). *Building digital bridges with emerging technologies* (ITU Background Paper BDB/09). Retrieved March 22, 2005, from http://www.itu.int/digitalbridges/docs/background/BDB-technology.pdf

International Telecommunication Union. (1997, February). *World telecommunication development report 1996/97. Trade in telecommunications: World telecommunication indicators* (C.R. Blackman, Ed.). Retrieved from http://www.itu.int/osg/spu/wtpf/wtpf98/trade/wtdr/

Jeffrey, P., & Mark, G. (1998, March). Constructing social spaces in virtual environments: A study of navigation and interaction. In K. Höök, A. Munro, & D. Benyon (Eds.), *Workshop on personalised and social navigation in information space* (SICS Tech. Rep. T98:02, pp. 24-38). Stockholm: Swedish Institute of Computer Science.

Jilek, W. G. (2001). Cultural factors in psychiatric disorders. In *Proceedings of the 26th Congress of the World Federation for Mental Health.* Retrieved January 2, 2005, from http://www.mentalhealth.com/mag1/wolfgang.html

Jonsson, E. (1998). *Electronic discourse: On speech and writing on the Internet* (chap. 1). Retrieved September 24, 2004, from http://www.ludd.luth.se/users/jonsson/D-essay/Electronic-Discourse.html

Kabir, M., Iliyasu, Z., Abubakar, I. S., & Aliyu, M. H. (2004, August 20). Perception and beliefs about mental illness among adults in Karfi village, northern Nigeria. *BMC International Health and Human Rights, 4*, Article 0003. Retrieved March 30, 2005, from http://www.biomedcentral.com/content/pdf/1472-698X-4-3.pdf

Khandelwal, S. K., Jhingan, H. P., Ramesh, S., Gupta, R. K., & Srivastava, V. K. (2004). India mental health country profile. *International Review of Psychiatry, 16*(1-2), 126-141.

Karim, S., Saeed, K., Rana, M. H., Mubbashar, M. H., & Jenkins, R. (2004). Pakistan mental health country profile. *International Review of Psychiatry, 16*(1-2), 83-92.

Kreijns, K., Kirschner, P. A., Jochems, W., & Van Buuren, H. (2004). Determining sociability, social space, and social presence in (a)synchronous collaborative groups. *Cyberpsychology and Behaviour, 7*(2), 155-172.

Krikorian, D. H., Lee, J. S., Makana Chock, T., & Harms, C. (2000). Isn't that spatial? Distance and communication in a 2-D virtual environment. *Journal of Computer Mediated Communication, 5*(4). Retrieved March 21, 2005, from http://jcmc.indiana.edu/vol5/issue4/krikorian.html

Kurihara, T., Kato, M., Sakamoto, S., Reverger, R., & Kitamura, T. (2000). Public attitudes towards the mentally ill: A cross-cultural study between Bali and Tokyo. *Psychiatry and Clinical Neurosciences, 54*(5), 547-552.

Kurtz, L. F. (1997). *Self-help and support groups: A handbook for practitioners*. Thousand Oaks, CA: Sage Publications.

Lauber, C., Nordt, C., Falcato, L., & Rossler, W. (2004). Factors influencing social distance toward people with mental illness. *Community Mental Health Journal, 40*(3), 265-274.

Link, B. G., Struening, E. L., Neese-Todd, S., Asmussen, S., & Phelan, J. C. (2001). Stigma as a barrier to recovery: The consequences of stigma for the self-esteem of people with mental illnesses. *Psychiatric Services, 52*(12), 1621-1626. Retrieved December 8, 2004, from http://ps.psychiatryonline.org/cgi/content/full/52/12/1621

Liu, Y. (2002, April 1). What does research say about the nature of computer-mediated communication: Task-oriented or social-emotion-oriented? *Electronic Journal of Sociology, 6*(1), Article 0001. Retrieved March 21, 2005, from http://www.sociology.org/content/vol006.001/liu.html

Lonner, W. J., Dinnel, D. L., Hayes, S. A., & Sattler, D. N. (Eds.). (2004). *Online readings in psychology and culture*. Retrieved December 23, 2004, from http://www.ac.wwu.edu/~culture/contents_complete.htm

Low, A. A. (1997). *Mental health through will training: A system of self-help in psychotherapy as practiced by Recovery, Incorporated* (3rd ed.). Glencoe, IL: Willett.

Lundvall, B.-Å. (1992). Introduction. In B.-Å. Lundvall (Ed.), *National systems of innovation: Towards a theory of innovation and interactive learning* (pp. 1-19). London: Pinter Publishers.

MacDonald-Wilson, K. L., Rogers, E. S., Massaro, J. M., Lyass, A., & Crean, T. (2002). An investigation of reasonable workplace accommodations for people with psychiatric disabilities: Quantitative findings from a multi-site study. *Community Mental Health Journal, 38*(1), 35-50.

Mansell, R., & Wehn, U. (Eds.). (1998). *Knowledge societies: Information technology for sustainable development.* New York: Oxford University Press.

McKenna, K. Y. A., & Bargh, J. A. (1998). Coming out in the age of the Internet: Identity "demarginalization" through virtual group participation. *Journal of Personality & Social Psychology, 75*(3), 681-694.

McKenna, K. Y. A., & Bargh, J. A. (1999). Causes and consequences of social interactions on the Internet: A conceptual framework. *Media Psychology, 1*, 249-269.

McKenna, K. Y. A., & Green, A. S. (2002). Virtual group dynamics. *Group Dynamics: Theory, Research and Practice, 6*(1), 116-127.

McKenna, K. Y. A., Green, A. S., & Gleason, M. J. (2002). Relationship formation on the Internet: What's the big attraction? *Journal of Social Issues, 58*(1), 9-32.

McKenzie, K., Patel, V., & Araya, R. (2004). Learning from low income countries: Mental health. *British Medical Journal*, *329*(7475), 1138-1140. Retrieved December 21, 2004, from http://bmj.bmjjournals.com/cgi/content/full/329/7475/1138

Menou, M. J. (2001). The global digital divide; beyond hICTeria. *Aslib Proceedings, 53*(4), 112-114. Retrieved from http://www.aslib.co.uk/proceedings/

Mind. (2005). *About Mind*. Retrieved March 30, 2005, from http://www.mind.org.uk/About+Mind/

Moderation Management. (n.d.). What is Moderation Management? Retrieved November 3, 2004, from http://www.moderation.org/whatisMM.shtml

Moore, D.W. (2000, February 23). Americans say Internet makes their lives better. *Gallup News Service*. Princeton, NJ: The Gallup Organization.

Murray, P. (1996). Recovery, Inc. as an adjunct to treatment in an era of managed care. *Psychiatric Services*, *47*(12), 1378-1381.

Nater, T., Boyer, C., & Eysenbach, G. (2000, June 30). Debate about evaluation and monitoring of sites carrying the HON-Logo [Letter to the editor]. *Journal of Medical Internet Research*, *2*(2), e13. Retrieved November 3, 2004, from http://www.jmir.org/2000/2/e13/

National Alliance for the Mentally Ill. (2005). *About NAMI*. Retrieved March 30, 2005, from http://www.nami.org/Content/NavigationMenu/Inform_Yourself/About_NAMI/About_NAMI.htm

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Linguistic style and deception. *Personality and Social Psychology Bulletin*, *29*(5), 665-675.

Nystrom, M., Dahlberg, K., & Segesten, K. (2002). The enigma of severe mental illness: A Swedish perspective. *Issues in Mental Health Nursing, 23*(2), 121-134.

Ohaeri, J. U., & Fido, A. A. (2001). The opinion of caregivers on aspects of schizophrenia and major affective disorders in a Nigerian setting. *Social Psychiatry and Psychiatric Epidemiology, 36*(10), 493-499.

Ohtaki, S., Ohtaki, T., & Fetters, M. D. (2003). Doctor-patient communication: A comparison of the USA and Japan. *Family Practice, 20*(3), 276-282.

Paua, F. (2003). Global diffusion of ICT: A progress report. In S. Dutta, B. Lanvin, & F. Paua (Eds.), *Readiness for the networked world* (Global Information Technology Report [GITR] 2002-2003, chap. 2). Retrieved January 3, 2005, from http://www.weforum.org/pdf/Gcr/GITR_2003_2004/Progress_Chapter.pdf

Payne, R. K. (2003). *A framework for understanding poverty* (revised ed.). Bayton, TX: RFT Publishing.

Pennebaker, J. W., & Seagal, J. D. (1999). Forming a story: The health benefits of narrative. *Journal of Clinical Psychology, 55*(10), 1243-1254.

Pentland, A., Fletcher, R., & Hasson, A. (2004). DakNet: Rethinking connectivity in developing nations. *IEEE Computer Society, 37*(1), 78-83.

Pew Internet and American Life Project. (2003, December 22). America's online pursuits: The changing picture of who's online and what they do. Retrieved September 23, 2004, from http://207.21.232.103/pdfs/PIP_Online_Pursuits_Final.PDF

Pew Internet and American Life Project. (2001, October 31). Online communities: Networks that nurture long-distance relationships and local ties. Retrieved September 16, 2004, from http://207.21.232.103/pdfs/PIP_Communities_Report.pdf

Pew Internet and American Life Project. (2002, May). Vital decisions: How Internet users decide what information to trust when they or their loved ones are sick. Retrieved September 23, 2004, from http://207.21.232.103/pdfs/PIP_Vital_Decisions_May2002.pdf

Phipps, L. (2000). New communications technologies: A conduit for social inclusion. *Information, Communication and Society, 3*(1), 39-68.

Puskar, K. R., Aubrecht, J., Beamer, K., & Carozza, L. J. (2004). Implementing information technology in a behavioral health setting. *Issues in Mental Health Nursing, 25*(5), 439-450.

Raguram, R., Raghu, T. M., Vounatsou, P., & Weiss, M. G. (2004). Schizophrenia and the cultural epidemiology of stigma in Bangalore, India. *The Journal of Nervous and Mental Disease, 192*(11), 734-744.

Ralph, R. O. (2002). The dynamics of disclosure: Its impact on recovery and rehabilitation. *Psychiatric Rehabilitation Journal, 26*(2), 165-172.

Read, J., & Law, A. (1999). The relationship of causal beliefs and contact with users of mental health services to attitudes to the "mentally ill". *The International Journal of Social Psychiatry, 45*(3), 216-229.

Regmi, S. K., Pokharel, A., Ojha, S. P., Pradhan, S. N., & Chapagain, G. (2004). Nepal mental health country profile. *International Review of Psychiatry, 16*(1-2), 142-149.

Reissman, F., & Carroll, D. (1995). *Redefining self-help*. San Francisco: Jossey-Bass.

Rigby, M., Lindmark, J., & Furlan, P. M. (1998). The importance of developing an informatics framework for mental health. *Health Policy, 45*(1), 57-67.

Risk, A., & Dzenowagis, J. (2001, December 26). Review of Internet health information quality initiatives. *Journal of Medical Internet Research, 3*(4), Article e28. Retrieved November 2, 2004, from http://www.jmir.org/2001/4/e28/

Ritsher, J. B., Otilingam, P. G., & Grajales, M. (2003). Internalized stigma of mental illness: Psychometric properties of a new measure. *Psychiatry Research, 121*(1), 31-49.

Roth, D., Antony, M. M., Kerr, K. L., & Downie, F. (2000). Attitudes toward mental illness in medical students: Does personal and professional experience with mental illness make a difference? *Medical Education, 34*(3), 234-236.

Rude, S. S., Gortner, E. M., & Pennebaker, J. W. (2004). Language use of depressed and depression-vulnerable college students. *Cognition and Emotion, 18*(8), 1121-1133.

Sackett, D. L., Rosenberg, W. M. C., Muir Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996, January 13). Evidence based medicine: What it is and what it isn't [Editorial]. *British Medical Journal, 312*(7023), 71-72. Retrieved November 1, 2004, from http://bmj.bmjjournals.com/cgi/content/full/312/7023/71

Saraceno, B., & Barbui, C. (1997). Poverty and mental illness. *Canadian Journal of Psychiatry, 42*(3), 285-290. Retrieved March 22, 2004, from https://www.cpa-apc.org/Publications/Archives/CJP/1997/April/revpaper2_0497.htm

Sartorius, N., Jablensky, A., Korten, A., Ernberg, G., Anker, M., Cooper, J. E., et al. (1986). Early manifestations and first-contact incidence of schizophrenia in different cultures. A preliminary report on the initial evaluation phase of the WHO collaborative study on determinants of outcome of severe mental disorders. *Psychological Medicine, 16*(4), 909-928.

Schuler, D. (1996). *New community networks: Wired for change*. Reading, MA: Addison-Wesley.

Shibre, T., Negash, A., Kullgren, G., Kebede, D., Alem, A., Fekadu, A., et al. (2001). Perception of stigma among family members of individuals with schizophrenia and major affective disorders in rural Ethiopia. *Social Psychiatry and Psychiatric Epidemiology, 36*(6), 299-303.

Simmonds, S., Coid, J., Joseph, P., Marriott, S., & Tyrer, P. (2001). Community mental health team management in severe mental illness: A systematic review. *British Journal of Psychiatry, 178*(6), 497-502.

Sosik, J. J. (Ed.), Jung, D. I., Berson, Y., Dionne, S. D., & Jaussi, K. S. (2004). *The dream weavers: Strategy-focused leadership in technology-driven organizations.* Greenwich, CT: Information Age Publishing.

Sproull, L., & Kiesler, S. (1991). *Connections: New ways of working in the networked organization*. Cambridge, MA: MIT Press.

Stirman, S., & Pennebaker, J. (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine, 63*, 517-522.

Stuber, J., & Kronebusch, K. (2004). Stigma and other determinants of participation in TANF and Medicaid. *Journal of Policy Analysis and Management, 23*(3), 509-530.

Suler, J. R. (2004, July 28). The online disinhibition effect [Electronic version]. *Cyberpsychology and Behavior, 7*(3), 321-326. Retrieved September 1, 2004, from http://www.rider.edu/~suler/psycyber/disinhibit.html

Thomson, H., Petticrew, M., & Morrison, D. (2001). Housing interventions and health: A systematic review. *British Medical Journal, 323*(7306), 187-190. Retrieved March 30, 2005, from http://bmj.bmjjournals.com/cgi/content/full/323/7306/187

Toledo Piza Peluso de, E., & Blay, S. L. (2004). Community perception of mental disorders: A systematic review of Latin American and Caribbean studies. *Social Psychiatry and Psychiatric Epidemiology, 39*(12), 955-961.

Trimpey, J. (2003). Why self-recovery?. Retrieved November 4, 2004, from http://www.rational.org/why.html

UN (United Nations). (2000a, August). Bridging the digital divide — Global Volunteer Corps for information technology training (DPI/2147/C). Retrieved January 5, 2005, from http://www.un.org/millennium/media/UNITeS.htm

UN (United Nations). (2000b, September 18). Millennium Declaration (A/RES/55/2). General Assembly Resolution, 8th plenary meeting, 5th session, item 60(b). Retrieved January 17, 2005, from http://www.un.org/millennium/declaration/ares552e.pdf

UNCSTD (United Nations Commission on Science and Technology for Development). (1997, May 12). Action arising from the second session: Implementation of and progress made on decisions taken at the 2nd session of the Commission, including follow-up work on technology for basic needs, gender and sustainable development, and coalition of resources (E/CN.16/1997/8). UNCSTD 3rd session, item 4, UNCTAD secretariat note, suppl. 11. Geneva: United Nations.

UNDP (United Nations Development Program). (2001). UNDP thematic trust fund. Information and communication technology (ICT) for development. Retrieved January 11, 2005, from http://www.undp.org/trustfunds/TTF-ICTe.PDF

UNESCAP (United Nations Economics and Social Commission for Asia and the Pacific). (1997). Management of self-help organizations of people with disabilities (ST/ESCAP/1849). Retrieved November 4, 2004, from http://www.unescap.org/esid/psis/disability/decade/publications/z15006mg/z1500601.htm

UNITeS (United Nations Information Technology Service). (n.d.). A brief history of the UNITeS process. Retrieved January 17, 2005, from http://www.unites.org/html/unites/history.htm

UNV (United Nations Volunteers). (2005). FAQ: Frequently asked questions about UN Volunteers. Retrieved January 17, 2005, from http://www.unv.org/about/faq.htm

U.S. General Accounting Office. (2000). Community-based care increases for people with serious mental illness (GAO-01-224). Retrieved December 6, 2004, from http://www.gao.gov/new.items/d01224.pdf

Wagner, C., Cheung, K., Lee, F., & Ip, R. (2003). Enhancing e-government in developing countries: Managing knowledge through virtual communities. *The Electronic Journal of Information System in Developing Countries, 14*(4), 1-20.

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research, 23*(1), 3-43.

Walther, J. B., & D'Addario, K. P. (2001). The impacts of emoticons on message interpretation in computer-mediated communication. *Social Science Computer Review, 19*(3), 324-347.

Walther, J. B., & Tidwell, L. C. (1995). Nonverbal cues in computer-mediated communication, and the effect of chronemics on relational communication. *Journal of Organizational Computing, 5*(4), 355-378.

World Bank. (1993). World development report 1993: Investing in health. Retrieved from http://publications.worldbank.org/

World Bank. (1998). World development report 1998/1999: Knowledge for development. Retrieved January 5, 2005, from http://www.worldbank.org/wdr/wdr98/contents.htm

WHO (World Health Organization). (2001). Mental health: New understanding, new hope (World Health Report 2001). Retrieved November 22, 2004, from http://www.who.int/whr/2001/en/whr01_en.pdf

WHO (World Health Organization). (2004). Promoting mental health: Concepts, emerging evidence, practice (WHO Summary Report). Retrieved November 22, 2004, from http://www.who.int/entity/mental_health/evidence/en/promoting_mhh.pdf

WHO (World Health Organization) European Ministerial Conference on Mental Health. (2005, January 14). Mental health declaration for Europe: Facing the challenges, building solutions (EUR/04/5047810/6). Retrieved January 19, 2005, from http://www.euro.who.int/document/mnh/edoc06.pdf

Zuckerman, D. (1993). Reasonable accommodations for people with mental illness under the ADA. *Mental and Physical Disability Law Reporter, 17*(3), 311-319.

KEY TERMS

Community-based approaches: Strategies that address problems of disadvantaged populations through the involvement of the community they belong to, as a whole but at different levels. Community-based interventions encourage promotion of service delivery, equitable opportunities, human rights, and quality of life of such populations through implementations taking place into venues familiar to the target populations (i.e., churches, libraries, schools). This kind of approach is becoming increasingly popular due to both its sustainability and trans-cultural effectiveness.

Computer-mediated communication: Communication between two or more persons that happens through technological devices, such as a computer and the Internet. These interactions differ from face-to-face ones in that they are written, remote, and lack physical cues. They can be synchronous (like chat rooms or instant messaging) or asynchronous (like e-mails). However, they are able to foster a sense of community, thereby generating social capital.

DALY: Method for calculating the health impact of a disease in terms of reported (or estimated) cases of premature death, disability, and days of infirmity due to illness from a specific disease or condition. For each disease or risk factor, DALY is calculated as the annual sum of the years of life lost due to premature mortality and the *years lived with disability* in the population for that health condition. DALY extends the concept of potential years of life lost due to premature death to include years of *healthy* life lost by virtue of being in states of ill health.

Online self-help: Online self-help groups aim at individual change by helping people help themselves, like traditional self-help groups but in online settings. They overcome the participation barriers traditional self-help groups have by taking advantage of the anonymity and disinhibition characteristic of online settings, in order to foster deeper disclosure and acceptance, with particular respect to embarrassing situations, non-mainstream preferences, and stigmatizing health conditions.

Patient education: Interventions for providing persons with an illness with some medical education about their disorder in order to improve compliance to hygiene measures and medications, recurrence prevention, management of the illness, short-term and long-term outcomes, and quality of life. Health care professionals, qualified publications, and/or trained peer-counselors usually distribute medical information. For many illnesses this is a low-cost, evidence-based intervention.

People with mental illness: Mental illness displays itself as difficulties in behaving, thinking, and/or interacting with other persons, and has very high rates worldwide. Symptoms vary from accessional relational impairment, along with chronic external and internalized stigmas, to actual psychiatric disability that needs accommodation. Furthermore, people with mental illness also experience stigma, marginalization, isolation, and low self-esteem that influence one other and the illness course deeply.

Virtual learning: Consists of acquiring formal education through the Internet, generally in a self-paced way. A variety of online environments are available (Web interface, e-books, e-mails, e-groups, tele-classes, chat rooms, instant messaging, one-to-one tutoring, study buddy, group work, and others), and all are learner centered, knowledge centered, assessment centered, and community centered. Even though virtual learning is not suitable for all kind of learners, it is usually more cost effective than traditional onsite learning.

This work was previously published in Empowering Marginal Communities with Information Networking, edited by H. Rahman, pp. 191-226 copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global)

Chapter 5.25
The Impact of Professional Certifications on Healthcare Information Technology Use

Neset Hikmet
University of South Florida, Sarasota-Manatee, USA

Anol Bhattacherjee
University of South Florida, Tampa, USA

ABSTRACT

This study examines the effects of professional certifications such as JCAHO on healthcare information technology (HIT) usage in healthcare organizations and user satisfaction with such usage. Using survey data collected from healthcare administrators in a nation-wide sample of 347 hospitals and long-term care facilities, we provide evidence that professional certifications do indeed enhance HIT usage and user satisfaction, at least within specialized user groups such as healthcare administrators. We further demonstrate that this increase in HIT usage due to professional certifications increases with facility size and is more prominent for larger hospitals than for smaller long-term care facilities, though the same cannot be said of user satisfaction. Our study suggests that professional certifications can be used as a valuable tool for motivating HIT usage, while also drawing attention to an under-examined area of HIT research.

INTRODUCTION

As the cost of healthcare has soared in the United States, rising to $1.7 trillion or 15.3% of GDP in 2003 (CMS, 2005), the role and use of healthcare information technology (HIT) has come into increased focus. HIT, in this context, refers to a wide range of clinical systems such as electronic medical records (EMR), computerized patient order entry (CPOE), and pharmacy information systems, and administrative systems such as patient billing systems, budgeting systems, and scheduling systems, that are expected to streamline healthcare delivery to patients, improve healthcare quality, and reduce delivery costs. In the report presented to the U.S. Congress by the Medicare Payment Advisory Commission (MEDPAC, 2004), HIT was identified as having the potential to significantly improve the quality, safety, and efficiency of healthcare. A similar report by the National Health Leadership Council (NBCH, 2005) identified HIT as the critical foundation for promoting health system reform, generating productivity and performance improvement, and producing significant cost reduction in healthcare expenditures.

As healthcare organizations face increasing pressure to invest in HIT, many healthcare managers are struggling to find ways to motivate physicians, nurses, and administrators to use the implemented HIT. Clearly, technology deployment is futile if users do not use the technology, use it inappropriately, or find ways to circumvent its usage. For instance, in 2003 doctors at the prestigious Cedars-Sinai Medical Center at Los Angeles rebelled against their newly installed CPOE system, complaining that the system was too great a distraction from their medical duties, forcing the withdrawal of a system that was already online in two-thirds of the 870-bed hospital (Freudenheim, 2004). The Leapfrog Group (an advisory group associated with the National Academy of Sciences) estimated that, of the nation's 300 non-governmental hospitals (6% of all hospitals in the U.S.) that have implemented comprehensive HIT systems, only 40 of these systems (less than 1%) are routinely used by of doctors for ordering prescriptions and laboratory tests (Freudenheim, 2004).

One technique that is expected to motivate HIT use among hospital administrators is professional certifications (Chow, 2001; Watcher, 2004). Many hospitals nationwide seek certification from the Joint Commission of Accreditation of Healthcare Organizations (JCAHO) to demonstrate their commitment to quality in healthcare delivery through the use of HIT and to meet eligibility requirements for participation in the government's Medicare and Medicaid programs (Associated Press, 2004). A significant requirement of JCAHO certification is continuous monitoring and tracking of a variety of operational statistics such as design of new services, implementation of safety plans, and infection control, which require the use of HIT. Furthermore, information management, potentially using HIT, is a key concern in the JCAHO certification process, including management of patient-related information, use of comparative information, and the transmission of national hospital quality measure data (JCAHO, n.d.). Hence, it is incumbent upon healthcare administrators to use the available HIT appropriately in order to meet and continue JCAHO certification status for their facility.

In this paper, we examine whether professional certifications such as JCAHO are indeed effective in motivating HIT use among administrators in healthcare settings. Additionally, we are interested in knowing whether JCAHO-induced HIT usage patterns vary across different types of healthcare organizations such as hospitals and long-term care facilities (nursing homes). Given the size and revenue differences between hospitals and long-term care facilities and the consequent availability of slack financial and technological resources to devote to HIT-driven quality initiatives such as JCAHO certifications, one may expect the motivation for JCAHO and HIT usage to be different across these organizations. Furthermore, given the broader scale and scope of hospital operations relative to long-term care facilities, the benefits of JCAHO may be disproportionately larger for the former than the latter.

Examining these issues is important for both practical and research reasons. From a practical standpoint, though JCAHO certification is eagerly sought by healthcare facilities nationwide, there is no evidence yet that such certification indeed yields significant quality and performance gains through the use of HIT. Though professional certification may be one potential tool in a manager's arsenal to motivate organizational HIT use, at least within certain user groups such as healthcare administrators, to the best of our knowledge, no prior study has yet examined whether such certifications are effective in the first place. Our study addresses this gap by examining the effect of certifications on HIT use by using field survey data collected from a random national sample of healthcare executives working for JCAHO certified and non-certified hospitals and long-term care facilities. If professional certifications are indeed effective in enhancing HIT usage and improving healthcare delivery, as expected, then organizational investments in such certifications and/or similar quality initiatives will be cost-justified. Furthermore, our study may also provide some evidence as to which type of healthcare organization stands to benefit most from JCAHO certification and related HIT usage.

From a research perspective, understanding the role of professional certifications is interesting also because much of the prior research on HIT usage has focused almost exclusively on user perceptions of the usefulness and ease of use of a given HIT, users' personal attributes such as information-seeking preference and Internet dependence, and situational characteristics such as healthcare needs (e.g., Wilson & Lankton, 2004). However, currently we know little about what organizational and/or structural factors can motivate given HIT deployment and usage among healthcare organizations to improve healthcare delivery. For instance, one of the stated concerns of the U.S. government is to understand organizational factors that enable or hinder HIT use among healthcare organizations (AHCPR, 1998). The ultimate goal of our work is to provoke enough interest among academics to study the underlying structural factors, such as professional certifications, that may help explain why HIT deployment is more successful in some organizations and less successful in others.

The rest of the paper proceeds as follows. In the next section, we present our research hypotheses and theoretical rationale for the same. In the third section, we describe our empirical methods for data collection and analysis to test the our hypotheses. Following this, we discuss our observed findings and their implications and limitations. The paper concludes with a summary of our key findings and suggestions for future research.

RESEARCH HYPOTHESES

JCAHO is a non-profit organization whose mission is "to continuously improve the safety and quality of care provided to the public through the provision of health care accreditation and related services that support performance improvement in health care organizations" (www.jcaho.org). As part of their accreditation procedure, JCAHO has been independently auditing the operational and quality performances of healthcare facilities in the U.S., once every three years, since 1994. Even though JCAHO certification is not mandatory for hospitals or other healthcare facilities, it is a prestigious award that implies that the facility is in compliance with the standards set by the accrediting organization and routinely implements and evaluates required quality indicators in all functional areas of the organization, including management of information (Watcher, 2004). Further, JCAHO certification is required for a healthcare facility to participate in Medicare and Medicaid programs, which are often large revenue sources for many healthcare organizations.

The JCAHO accreditation or re-accreditation process is based on a site visit, during which the JCAHO team evaluates how well a hospital meets more than 500 standards specified in the . This data is then aggregated into 46 "grid elements," 16 "performance areas," an overall performance score (on a 0 to 100 scale), and an accreditation decision (e.g., accreditation with commendation, conditional accreditation, accreditation denied, unaccredited, etc.). One key performance area recently added to the certification process is the "management of information," which includes grid elements such as information management planning, availability of patient-specific information, data collection and analysis, literature to support decision making, and use of comparative information. JCAHO certification requires formal evaluations of each of the metrics just mentioned and an ongoing organizational commitment to meeting JCAHO standards, performance elements, and scoring requirements.

JCAHO accreditation and re-accreditation procedures require hospital administrators to continuously monitor and document a wide variety of healthcare and patient safety statistics, such as adverse drug events and infection occurrences. Though JCAHO does not regulate what kind of HIT should be deployed or how it should be used or managed, it does encourage HIT usage as a reliable and timely means for tracking patient and healthcare quality information. For example, Standard IM.1.10 in JCAHO manual relates to information planning: "The hospital plans and designs information management processes to meet internal and external information needs"; and Standard IM.4.10 addresses information-based decision making: "The information management system provides information for use in decision making" (MLANET, 2006). The heavy record keeping, information processing, and periodical reporting needs imposed by JCAHO are therefore expected to motivate healthcare executives to aggressively deploy and utilize HIT to conform to JCAHO needs. Furthermore, meeting the JCAHO objectives by virtue of their HIT usage can be expected to enhance healthcare administrators' satisfaction with HIT usage, in consonance with positive associations between technology usage and user satisfaction reported in the information technology implementation literature (DeLone & McLean, 1992). These expectations lead us to hypothesize:

H1. Administrators in JCAHO certified facilities have (a) higher HIT use and (b) higher user satisfaction than those in non-JCAHO certified facilities.

JCAHO accreditation and re-accreditation require substantial financial and technological resources on the part of the accredited facility. Sophisticated HIT systems and applications are required to continuously track and monitor a wide variety of JCAHO performance and quality metrics and take remedial actions if things go wrong. A high level of financial resources and technological expertise is required to not only achieve such capability, but also to maintain it over the long-term. Larger organizations such as hospitals are likely to have more slack resources to devote to HIT investments than smaller organizations such as long-term care facilities. Once the appropriate HIT system is implemented, the better fit between the technology and administrators' tasks (e.g., tracking JCAHO metrics, etc.) may be expected to enhance their actual usage of HIT. Further, larger organizations, by virtue of their greater scale and scope of operations, are often able to distribute the benefits of JCAHO certifications over a wider array of services and divisions, resulting in greater satisfaction among healthcare administrators. Hence, we propose:

H2. The effect of JCAHO certification on (a) higher HIT use and (b) higher user satisfaction is greater for administrators of hospitals than for long-term care facilities.

RESEARCH METHODOLOGY

Survey Approach

Empirical data for testing our hypotheses was collected via a mail survey of healthcare administrators in hospitals and long- term care facilities throughout the United States. We focused specifically on healthcare administrators as the key informant in our study because these individuals were responsible for meeting JCAHO tracking and reporting needs and were expected to use HIT systems to do so. Our sampling frame consisted of a list of healthcare organizations purchased from a list broker that specializes in healthcare mailing lists. This list included the name, address, and type of healthcare facilities (e.g., hospital, long-term care) throughout the U.S., along with the names, addresses, phone numbers, and titles of senior executives (e.g., chief executive officers, presidents, managers, supervisors) in those organizations. A stratified sampling scheme was utilized, in which we categorized organizations in the sampling frame into three strata based on facility type: hospitals, long-term care facilities (e.g., nursing homes or assisted living facilities), and community health centers[1]. A sample of 6,713 respondents, divided proportionally (10% of each population stratum) among the three strata, was selected randomly as targets of our survey research.

Our survey followed the approach recommended by Dillman (1978), intended to maximize response rates and minimize non-response bias. Using this approach, we first sent out post-cards to our target respondents informing them of the pending arrival of a survey questionnaire regarding the role and use of IT in the healthcare sector and soliciting their participation in this survey. Twenty postcards were returned as "non-deliverable" due to invalid addresses or the addressee having moved to a different organization. Valid subjects were then mailed the questionnaire booklet, along with a personalized cover letter and a postage-paid envelope for mailing back responses. Five weeks later, the questionnaire was mailed to non-respondents again, followed five weeks later by a reminder card urging them for the third time to respond to the survey if they had not already done so.

Following this multi-round survey, 550 surveys were returned, for an overall response rate of 8.2%. To test for non-response bias, we conducted two multinomial distribution tests comparing the distribution of facility type (hospitals and long-term care, vs. community health centers) and respondents' position (upper management vs. administrative staff) in our sample with that of the population (as aggregated from the facility data provided by our mailing list provider). Chi-square analyses for these differences were non-significant for both facility type (χ^2=5.61, p=0.16) and respondent's position (χ^2=2.53, p=0.11), suggesting that our sample was reasonably representative of the target population. The population and sample distribution for these two dimensions are shown in Table 1.

The community health centers stratum was dropped from further analysis because very few of the responding centers were JCAHO certified or were interested in JCAHO certification. Additionally, being local, non-profit, and community-owned, and serving low income and other underserved communities, the vision and operational structure of community health centers were significantly different than that of hospitals and long-term care facilities. This divergence in vision and goals likely resulted in different motivations for HIT use or non-use in these facilities, further justifying their omission from our data sample. However, the remaining two strata were retained in our study because of higher levels of JCAHO certifications in these strata. Furthermore, hospitals were much larger in size (in bed count, number of employees), scope of operations (e.g., procedures served), and availability of resources than long-term care facilities. Hence, these facilities can respectively be considered to be proxies of large and mid-sized organizations. A comparative examination of these two facility types therefore not only helped us assess the generalizability of the effect of professional certifications, but also helped us examine whether organizational size might have had some influence on the hypothesized effect.

Variable Operationalization

Our two independent variables of interest were professional certification (presence vs. absence of JCAHO certification) and facility type (hospitals

Table 1. Population and sample distributions

	Sampling Frame	Percentage	Study Survey	Percentage
Position				
Upper Management	4642	69.4	398	72.2
Administrative Staff	2051	30.6	152	27.6
Facility				
Hospital	1666	24.9	144	26.1
Long-Term Care	3671	54.8	316	57.4
Community Health Centers and Others	1356	20.3	91	16.7

Table 2. Descriptive statistics

		Professional Certification		
Facility Type		JCAHO Certified	Non-Certified	Total
Hospital	Number of observations (%) Mean HIT use Mean user satisfaction	98 (85.2%) 59.82 10.10	17 (14.8%) 25.24 8.35	115 (100%)
Long-Term Care	Number of observations (%) Mean HIT use Mean user satisfaction	58 (25.1%) 25.34 9.89	174 (74.9%) 22.20 8.98	232 (100%)

Legend: HIT use scores were calculated as a sum of 10 items that captured the number of hours spent by the respondent on using 10 common HIT applications in the healthcare sector, such as reporting, billing systems, forecasting, etc. User satisfaction scores were calculated as a sum of two 1-7 Likert scales. Hence, overall satisfaction scores for each respondent ranged from 2 to 14.

vs. long-term care facilities), and our two dependent variables were HIT usage and user satisfaction. The independent variables were captured as single fill-in measures in our survey questionnaire (e.g., whether the facility was JCAHO certified or not, and whether it was a hospital or a long-term care facility).

HIT usage was measured as the number of hours respondents (healthcare administrators) used HIT per day. Specifically, our survey questionnaire asked respondents (healthcare executives) to enter the number of hours they spent per week at work using 10 common HIT applications, such as reporting systems, billing systems, and error tracking systems. Self-reported hours across these 10 categories were added to create an overall HIT use score for that respondent. This detailed elaboration of usage statistics across different HIT was expected to elicit the respondents' true level of usage, which is often masked in self-reported usage measures employing Likert and similar scales.

User satisfaction with HIT use was computed by summing two seven-point Likert-scaled items (anchored between "strongly disagree" and "strongly agree") that asked respondents the extent to which they were satisfied with the IT available at their workplace and the level of support they received from IT staff at work. This satisfaction scale was based on a similar scale developed and validated by Tan and Lo (1990). Likert scale was appropriate for this measure because satisfaction was a perceptual construct and is best measured as such.

DATA ANALYSIS AND FINDINGS

Sample Characteristics

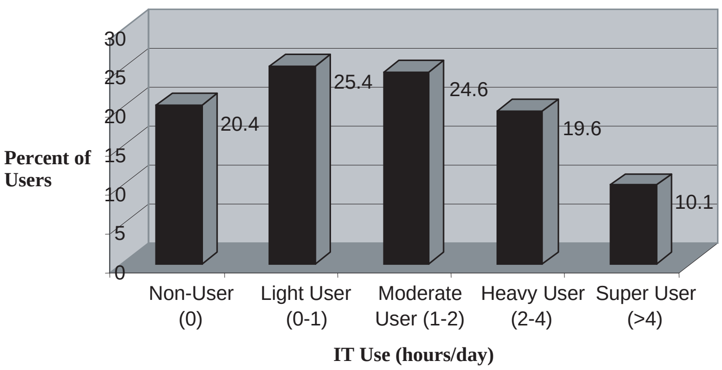
Following two rounds of reminders, our field survey resulted in an effective sample size of 347. This sample consisted of 115 hospitals and 232 long-term care facilities. Overall, 156 of the 347 facilities in our sample were JCAHO certified, and 191 were not. Ninety-eight of the 115 hospitals in our sample (85.2%) were JCAHO-certified, but only 58 of the 232 long-term care facilities (25.1%) enjoyed this certification, demonstrating a divergent pattern of JCAHO certification between these groups. This pattern is indicative of the organizational makeup of these two facility groups. Many long-term care facilities have fewer employees and perform tasks of less complexity in comparison to hospitals, thus making their recording and reporting activities a less important task for administrative staff and consequently resulting in the reduced dependence on JCAHO certification. Sample breakdown by groups, along with HIT use and user satisfaction means within each group, are reported in Table 2.

For descriptive analysis, we classified respondents into five categories based on their overall level of self-reported HIT use: (1) non-users: 0 hours per day, (2) light users: 1 hour or less per day, (3) moderate users: 1-2 hours per day, (4) heavy users: 2-4 hours per day, and (5) super users: more than 4 hours per day. This classification was based on an initial pilot study of healthcare executives (Hikmet & Burns, 2004). The frequency distribution of HIT use is shown graphically in Figure 1. It is interesting to note from this figure that more than 20% of the healthcare administrators in our sample did not use HIT at all, and more than 70% spend two hours or less per day on work-related HIT use. While this apparent low level of use may reflect respondents' time constraints at work, the required tracking and reporting by JCAHO and similar regulatory organizations are not bound by such constraints.

Hypotheses Testing

To test for the two hypotheses stated earlier, we first conducted a two-way multivariate analysis of variance (MANOVA) test to examine whether our two independent variables, professional certification (JCAHO vs. non-JCAHO) and facility type (hospitals v.s long-term care facilities), and their interaction had any significant effects on our two dependent variables of interest: HIT use

Figure 1. Extent of HIT use



and user satisfaction. MANOVA results, shown in Table 3, were significant for both professional certification (F=8.34, p<0.001) and facility type (F=3.67, p=0.026), while the interaction term between the two variables was weakly significant (F=0.97, p=0.053).

Next, we conducted follow-up analysis of variance (ANOVA) to examine which of the dependent variables experienced significant effects. As seen in Table 3, professional certification had a significant effect on both HIT use (F=7.17, p=0.008) and user satisfaction (F=9.46, p=0.002), when controlled for facility type. In contrast, facility type had a significant effect on HIT use (F=7.10, p=0.008), but not on user satisfaction (F=0.250, p=0.618), when professional certification was controlled. The interaction between professional certification and facility type also had a significant effect on HIT use (F=4.98, p=0.026), but not on satisfaction (F=0.94, p=0.332).

These analyses validate our first hypothesis that JCAHO certifications not only increases HIT use among healthcare administrators (Hypothesis H1a), but also increases user satisfaction with such use (Hypothesis H1b). HIT usage was more pronounced in larger facilities such as hospitals than in mid-sized facilities such as long-term care facilities, but user satisfaction experienced no such differential effect across facility types. Focusing on the interaction effect posited in the second hypothesis, we find that JCAHO certification had a stronger effect on motivating HIT use among hospital administrators than those in long-term care facilities (validating Hypothesis H2a), but did not increase user satisfaction among hospital administrators significantly more than those in long-term care facilities (failing to support Hypothesis H2b). Implications of these findings are discussed in the next section.

DISCUSSION AND CONCLUSION

Implications for Practice

The findings of our study confirm that healthcare managers can indeed motivate HIT usage within their organizations by proactively engaging in and obtaining professional certifications such as JCAHO. The incentive to maintain JCAHO accreditation works as an adequate incentive for healthcare administrators to use HIT systems appropriately to meet the monitoring and reporting requirements of JCAHO. Furthermore, administrators appear to be satisfied with their HIT usage, despite the demanding needs of JCAHO accreditation and re-accreditation. However, we do not have any evidence to examine whether this higher level of use among healthcare administrators translates into higher usage among other user groups such as physicians or nurses. Additional studies, with different subject populations, are required to examine those trends.

Table 3. MANOVA results

Independent	MANOVA		Dependent	ANOVA	
Variable	F-statistic	p-value	Variable	F-statistic	p-value
Prof. Certification	8.34	<0.001	HIT Use	7.17	0.008
			User Satisfaction	9.46	0.002
Facility Type	3.67	0.026	HIT Use	7.10	0.008
			User Satisfaction	0.25	0.618
Prof. Cert. x Facility Type	2.97	0.053	HIT Use	4.98	0.026
			User Satisfaction	0.94	0.332

Hospitals in our study experienced a larger increase in HIT use from JCAHO certification than long-term care facilities. Extrapolating this size effect, we expect that smaller-sized organizations such as doctors' offices stand to benefit least from professional certifications and, hence, professional certification programs may not be a worthwhile investment for smaller organizations such as long-term care facilities and community health centers. Furthermore, user satisfaction is expected to remain relatively invariant across facility types between JCAHO certified and non-certified groups. Hence, managers should measure HIT usage and not user satisfaction to evaluate the outcome of JCAHO certification process, and more so for smaller facilities.

Implications for Research

Our study contributes to HIT research by providing preliminary empirical evidence regarding the role of professional certifications on HIT use and user satisfaction in healthcare organizations. Though such certifications undoubtedly enhance an organization's stature and prestige in its community and signal its commitment on quality processes, their effect on HIT use in particular was unclear and unexamined prior to this study.

Our study also draws attention to organizational factors such as professional certifications that managers can control to motivate HIT usage within their organizations. While much of the prior HIT usage research has focused on personal or cognitive factors, such as perceived usefulness and perceived ease of use as antecedents of IT usage and user satisfaction (e.g., Wilson & Lankton, 2004), and has demonstrated good predictive abilities, it is worth noting that such factors cannot be controlled by managers and are therefore less relevant from a practitioner standpoint. In that sense, our study goes beyond predicting IT usage to the more relevant question of examining how organizations can proactively manipulate HIT use though professional certifications.

Though our study was conducted in a healthcare setting, our findings can be expected to be generalizable to other industry sectors. For instance, many firms in other industries seek ISO-9000 from the International Standards Organization to signal their commitment to similar quality improvement initiatives. Software firms often participate in CMM (Capability Maturity Model) certification initiatives from the Software Engineering Institute for similar reasons. However, additional studies are required to test for these effects.

Limitations of the Study

Finally, like most other empirical studies, our research was not without limitations. First, this study was conducted within the narrow context of healthcare organizations. Hence, our findings may not be readily generalizable to other industry sectors without conducting additional studies to test our reported effects in other industry contexts (e.g., manufacturing or service). Second, we viewed professional certification as an instance of a quality initiative, since such certification requires prior evaluation of organizational processes and procedures in accordance with the certification agency's norms and expectations. However, one may debate whether certification is an accurate or reasonable proxy of quality programs. Hence, the value and importance of our findings may be limited in instances where professional certification is not consistent with organizational quality initiatives. Third, the nature and substance of professional certifications vary widely across certification agencies, their scale and scope of certification, and certification metrics employed. Though JCAHO certification is viewed as being one of the most comprehensive and prestigious quality certifications in the healthcare industry, similar certifications from other agencies may be less valued. Such variance in certification quality may mitigate the nature of HIT use effects across certification agencies.

Future Research Directions

This study examined only two of several organizational factors, namely professional certifications and facility type (size), that can motivate HIT usage and user satisfaction among healthcare organizations. There may be more such organizational factors with comparable or greater predictive ability, and we encourage future research to uncover and investigate those factors. Second, we observed that the effects of some of these organizational factors are invariant for user satisfaction, while varying for HIT usage. Since HIT usage tends to be positively correlated with user satisfaction, this divergence of effects is theoretically perplexing. Additional research is required to examine potential reasons for such divergence. Third, future research can also examine the generalizability of our study's findings across other industry sectors (e.g., financial or technology), across other user groups (e.g., physicians or nurses), and across other forms of professional certification (e.g., ISO-9000 or CMM).

REFERENCES

Agency for Health Care Policy and Research (AHCPR). (1998). *Effective dissemination of health and clinical information and research findings.* Retrieved January 20, 1999, from http://www.ahrq.org

Associated Press. (2004, November 14). *Alta Bates risks loss of accreditation over negative audit.* Retrieved from http://www.signonsandiego.com/news/state/20041114-0931-ca-altabatesproblems.html

Chow, W.S. (2001). Ethical belief and behavior of managers using information technology for decision making in Hong Kong. *Journal of Managerial Psychology*, *16*(4), 258–267.

CMS (Center for Medicare & Medicaid Services). (2005). Brief summaries of Medicare and Medicaid. Retrieved January 20, 2006, from http://www.cms.hhs.gov/MedicareProgramRatesStats/downloads/MedicareMedicaidSummaries2005.pdf

DeLone, W. H., & McLean, E. (1992). Information system success: The quest for the dependent variable. *Information Systems Research*, *3*(1), 60–95.

Dillman, D.A. (1978). *Mail and telephone surveys.* New York: John Wiley & Sons.

Freudenheim, M. (2004, April 6). Many hospitals resist computerized patient care. *New York Times.* Retrieved February 2, 2006, from http://www.nytimes.com/2004/04/06/technology/06errors.html

Hikmet, N., & Burns, M. (2004, August). Health care executives: The association between external factors, use, and their perceptions of health information technology. In *Proceedings of the Tenth Americas Conference on Information Systems*, New York (pp. 220-227).

JCAHO (Joint Commission on Accreditation of Healthcare Organizations. (n.d.). *Comprehensive Accreditation Manual for Hospitals* (CAMH). Retrieved January 22, 2006, from http://www.jointcommission.org/AccreditationPrograms/Hospitals/Standards/FAQs/default.htm

MLANET *(Medical Library Association librarian's guide to JCAHO survey).* (2006). Retrieved March 1, 2006, from http://www.mlanet.org/resources/jcaho.html#Q1

MEDPAC (Medicare Payment Advisory Commission). (2004). *Report to the Congress: New approaches in Medicare*. Retrieved January 22, 2006, from http://www.medpac.gov/publications/congressional_reports/June04_ch7.pdf

NBCH (National Business Coalition on Health). (2005). *Health information technology: A founda-*

tion for health system reform. Retrieved January 22, 2006, from http://www.nbch.org/documents/wp_corporatepress0506.pdf

Tan, B.W., & Lo, T.W. (1990). Validation of a user satisfaction instrument for office automation success. *Information & Management*, *18*(4), 203–208.

Watcher, R.M. (2004, November 30). The end of the beginning: Patient safety five years after 'to err is human'. *Health Affairs*. [Web Exclusive, November 30, 534-545]

Wilson, E.V., & Lankton, N.K. (2004). Modeling patients' acceptance of provider-delivered e-health. *Journal of the American Medical Informatics Association*, *11*(4), 241–248.

ENDNOTE

1 Community health centers were later dropped from our sample because very few of these facilities had or desired JCAHO certification.

International Journal of Healthcare Information Systems and Informatics, Vol. 1, Issue 3, edited by J. Tan, pp. 58-68, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Section VI
Managerial Impact

This section presents contemporary coverage of the managerial implications of medical informatics, more specifically related to successful IT governance in the healthcare sector. Particular contributions address the implementation of large IT projects in healthcare institutions and the effective measurement of cost efficiency in healthcare. The managerial research provided in this section allows executives, practitioners, and researchers to gain a better sense of the role of IT in both healthcare and medical research.

Chapter 6.1
Governing Health Care with IT

Reima Suomi
Turku School of Economics and Business Administration, Finland

INTRODUCTION

The pressures for the health care industry are well known and very similar in all developed countries: altering populations, shortage of resources as it comes to staff and financial resources from the taxpayers, higher sensitivity of the population for health issues, new and emerging diseases, just to name a few. Underdeveloped countries have different problems, but they also have the advantage of being able to learn from the lessons and actions the developed countries made already, maybe decades ago. On the other hand, many solutions also exist, but they all make the environment even more difficult to manage: possibilities of networking, booming medical and health-related research and knowledge produced by it, alternative care-taking solutions, new and expensive treats and medicines, and promises of the biotechnology.

From the public authorities point of view, the solution might be easy: outsource as much as you can out of this mess. Usually, the first ones to go are marginal operational activities, such as laundry, cleaning, and catering services. It is easy to add information systems to this list, but we believe this is often done without a careful enough consideration. Outsourcing is often seen as a trendy, obvious, and easy solution, which has been supported by financial facts on the short run. Many examples show that even in the case of operational information systems outsourcing can become a costly option, not to speak of lost possibilities for organizational learning and competitive positioning through mastering of information technology.

In this article, we discuss how information technology and health care industry work together. Information technology is a valuable resource that must be managed within the health care industry. At the same time, information technology has the potential to renew the whole industry. Good practices in both must be supported by good IT governance.

Health care is a big resource user in every country. In Table 1 we have percentages of health care expenditures in relation to gross domestic product in selected countries, where the percentage is very high (WHO, 2004). As one can see, the cost explosion phenomenon hits both rich and poor countries, even though the wealthiest countries are well presented in the list.

Health care costs can be born by different parties within a national economy. Shares of different potential cost carriers vary from national economy to economy:

- The national government, directly or through different indirect arrangements such as separate funds or public insurance institutions
- Municipalities or other local public actors
- Private insurance institutions
- Employers
- The patients themselves

For example, in the United States, the raising costs of health care born by the employers have been a topic of much academic and industry discussion (Berry, Mirabito, & Berwick, 2004). Sadly enough, there is controversial evidence whether information technology can lower the total costs of running health services (Ammenwerth, Gräber, Herrmann, Bürkle, & König, 2003; Ko & Osei-Bryson, 2004).

There are few other forces than modern information technology that could cut down costs in the health care industry. In addition to cost cutting, information technology can provide extended productivity and is an ingredient in the processes that cumulate towards better care practices. But advantages from information technology are not to be harvested without constant focus on IT governance issues in the industry.

Table 1. Top 20 percentages of health care expenditure in relation to gross domestic product in 2001 in selected countries (WHO, 2004)

USA	13.9
Lebanon	12.2
Cambodia	11.8
Switzerland	11.0
Uruguay	10.9
Germany	10.8
Timor-Leste	9.8
Marshall Islands	9.8
France	9.6
Jordan	9.5
Argentina	9.5
Canada	9.5
Greece	9.4
Surimare	9.4
Australia	9.2
Palau	9.2
Portugal	9.2
Iceland	9.2
Croatia	9.0
Belgium	8.9

BACKGROUND

We have found the following reasons for the late adoption of modern information technology in the health care sector (Suomi, 2000):

- Fragmented industry structure
- Weak customers
- Strong professional culture of medical care personnel
- Hierarchical organization structures
- Handcrafting traditions
- One-sided education
- Big national differences in processes

We will next discuss each of these issues in greater detail.

Fragmented Industry Structure

Good competitors and customers are a key to success for any company and industry (Porter, 1990). Unfortunately, the health care sector has not been able to enjoy from neither of them. For a long time health care has been considered as a faceless public service, where normal competitive forces are not in effect. Health care organizations have not felt each other as competitors, and neither have they documented productive cooperative behavior. First with penetrating privatization the situation is starting to change.

Weak Customers

As it comes to customers, most often they get into touch with the industry when in a critical and sensitive situation, where bargaining power is very low. Bad service has just to be suffered. First during the last few years the concept "customer" has started to substitute the word "patient." Regulative bodies have become active in this respect, and for example in Finland a special patient-ombudsman has been institutionalized and legislation on patient reclamation and insurance has been introduced. In general, new technology is seen as a method to empower patients (Beun, 2003).

Strong Professional Culture of Medical Care Personnel

Professional cultures can have a profound outcome on organizational outcome. Within the health care sector, there are many strong professional cultures, the strongest of them being those of doctors and those of the nurses. People seeking to these professions usually value human interaction, and are not much up for abstract systems such as computers.

Hierarchical Organization Structures

A part of hospital organization has always been a strong hierarchical, professional and specialized structure. Work on the computers, unfortunately, is low on the hierarchy list, especially of course in the activities of keying in patient data that would be a natural thing to do for the doctors. As EITO (1995, p. 46) put it: "*for many Health Care applications, the most difficult obstacles can be social and cultural*." It is well known that information system development and application can be very difficult or at least different from less bureaucratic organizations than the health care.

Handcrafting Traditions

Even when we conclude that health care is a very information intensive industry, it has not been considered as such one. A good doctor is valued because of his handcrafting skills, especially in surgery, and it is not being understood that behind the handcraft operations a vast amount of knowledge is needed. Some, anyway, have understood that human body is the most complex entity in the world and of which information and knowledge has been collected over thousands of years.

One-Sided Education

Education of health care personnel has traditionally not focused on computer skills. Even the classical university tradition has kept medical and natural science (and thus computer) faculties apart from each other. Fortunately, during the last years, the drive for deeper cooperation between different science fields has begun to bear fruit.

Big National Differences in Processes

Patient care is very culturally bound, and especially the administrative processes behind vary greatly from one country to another. This, of course, makes standardization very difficult and the industry a bad target for suppliers of standard software and platforms. Neither do we have any dominant players in the field that would behave in the market as strong customers and trend setters.

Currently, however, not even the health care industry can not escape the tsunami of modern IT.

IT has to be governed within the industry and IT and governance structures meet in two ways. At one side, IT enables new governance structures for the health care industry. On the other, it is an object needing governing. As both sectors offer a multitude of new possibilities, innovations are called for in the industry.

IT governance thinking matures in organizations as any other discipline. Van Grembergen, De Haes, and Guldentops (2003) defined the following stages in their IT Governance Maturity Model:

- Nonexistent
- Initial/ad-hoc
- Repeatable but Intuitive
- Defined Process
- Managed and Measurable
- Optimized

Needless to say, in the health care industry, IT governance thinking is nonexistent or initial/ad hoc in the best situation.

HEALTH CARE, GOVERNANCE, AND IT

The Meaning of IT Governance Structure in Health Care

IT is an old acronym for information technology. Very often it is now replaced with the term *ICT,* referring to information and communication technology. This is to emphasize the communication services that are developing very fast, such as the Internet and mobile services. The letter "C" is often upgraded to the second dimension—alongside communication it can refer to contents. IT or IT governance is defined (IT Governance Institute, 2001) as follows:

IT governance is the responsibility of the board of directors and executive management. It is an integral part of enterprise governance and consists of the leadership and organizational structures and processes that ensure that the organization's IT sustains and extends the organization's strategies and objectives. (p. 10)

For many there is a temptation to understand governance as just a synonym for *management*. This is an oversimplification. Management is a goal-oriented activity, whereas governance is often given from outside, and organizations just have to live with it. This is not to say, that all governance structures would be beyond management control: most governance structures management can influence—at least on the long run. The long run is a key term in many aspects: When referring to governance structures, we talk about structures that are semipermanent and are not changed very frequently. *Structure* is a term closer to *architecture* than to *infrastructure*: governance structures are architectural terms, and are then implemented into infrastructures through different organizational forms. Therefore, the terms *organization form* and *governance structure* are

Table 2. Comparison of terms management, organizational form, and governance structure

	Management	Organizational Form	Governance Structure
Time perspective	Short	Medium	Long
Focus	Action	Internal organization	Interorganizational structures
Management Control	In action	Easy	Difficult
Metaphor	Communication channels	Infrastructure	Architecture
Character	Concrete	Formal	Abstract

not synonyms. Organizational forms are more formal and touch upon one organization, whereas governance structures are found in a richer selection of forms and organize themselves over a number of organizations. Table 2 summarizes our discussion here.

Governance structures are present in almost any human decision making situation. In Table 3, we have a collection of key aspects of IT governance structure issues in health care.

IT AS A TOOL TO BE GOVERNED IN HEALTH CARE

Information and communication technology needs management in health care organizations as in any other organization. Yet the issue seems to be very difficult for the health care organizations. Morrissey (2003, p. 6) paid attention to the fact that health care organizational culture is often hostile to new information technology: "*The expression 'Culture eats Strategy for lunch' has never been more accurate than with physician order entry.*" At the same time he documents the frustrations many hospitals and health care organizations have felt in the case of information technology: "*We have bought enough technology and we're not getting the expected value out of it.*" Haux, Ammenwerth, Herzog, and Knaup (2002, p. 19) have paid attention to the weak management of information technology and systems in health care organizations: "*The health care institutions, especially hospitals, must emphasize professional information management more strongly in their organizations.*"

Ross and Rockart (1996) have defined the following success factors for successful IT management:

- Achieve two-way strategic alignment
- Develop effective relationships with line management
- Deliver and implement new systems
- Build and manage infrastructure
- Reskill the IT organization
- Manage vendor partnerships
- Build high performance
- Redesign and manage the federal IT organization

Next, we discuss these issues in the health care IT context.

Two-way strategic alignment means that the organizational strategy affects the IT strategy, and vice versa. The current tumult of health care IT is partly a result of the fact that the effect of IT on health care processes has been denied for too long.

Effective relationship with line management means that communication is fluent and the parties aim at common goals. Effective communication needs a common language. Unfortunately, medical people and IT professionals do not always speak the same language. The applications the staff members use are the face of the information technology to the whole organization. A deliberate balance with investments to the infrastructure as well is needed. Functional applications can not be built on a flimsy base, but overinvestment in infrastructure is always a danger threatening organizations (Broadbent, 1999). Typical key applications of a hospital are electronic patient records (EPRs) and picture archiving and communication systems (PACSs).

Even in the health care organizational environment, IT professionals are under constant educational and skill renewal pressure. One of the big challenges meeting them is that patient data must move over traditional organizational boundaries and therefore must be able to deliver professional services in networked and often virtual organizational settings. The same issue is central in the management of vendor relationships: health care organization's IT staff must be able to integrate solutions from several vendors or, even better, they must be able to put enough pressure to vendors to self integrate their solutions.

Table 3. IT governance structure issues in health care

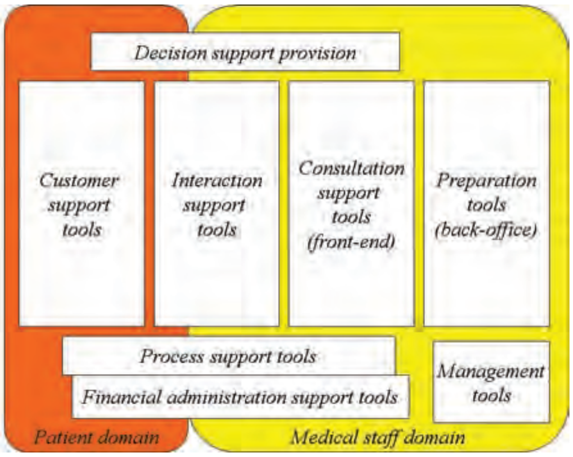
IT as an enabler

- Health-related information on the Web
- Private-public sector cooperation
- Allocation of patients to different levels of care
- Customer contacts distribution between electronic and classical means
- Ownership, structure and allocation of patient, population-level and other critical data
- Electronic forums for patients to interact
- Electronic prescription systems

IT as an object to be governed

- New legislation needs because of the new data processing possibilities
- Data privacy and security
- Structure and status of the information resource management in health care units
- IT-general management partnership
- Sourcing decisions of IT
- Charging arrangements on IT-services

Figure 1. Classification of IT tools in health care



Performance of IT solutions is dependent on many factors. Typical critical resources are time and money. Lack of money can be seen in undersized hardware and software solutions, and lack of IT-professionals time culminates in bad designs for the solutions.

Federal IT organization means that there are both centralized and decentralized IT units in the organizations. Even in big hospitals there is usually just one big IT unit. The federal structure problematic comes more into the foreground in the interorganizational settings of systems.

IT AS A TOOL FOR GOVERNANCE IN HEALTH CARE

How can information technology help in governing health care? We can approach this issue by taking a look at the application areas of IT in health care organizations. We propose the classification as shown in Figure 1, as adapted from (Suomi, 2000). The classification is a conceptual one, and actual information systems met in health care settings might fit several categories of the framework's systems.

The heart of our classification is in the interaction between the medical personnel and the patient, a kind of basic value-adding chain. In addition to the basic value chain in the middle of the figure, there are certain supporting functions and systems needed.

The figure divides the systems into systems for the use by the patient/customer (patient domain) and by the medical staff (medical staff domain).

We further differentiate between two basic use scenarios of these systems:

1. (patient) self-care scenario
2. (patient-medical staff) interaction scenario

In addition we have the activities of the health care professional or organization, that are not connected to any specific individual patient interaction, but they are of less interest for our analysis here.

SELF-CARE SCENARIO

The self-care scenario means that the patient tries to manage his or her disease or illness without individual professional help. Decision support provision in the self-care scenario helps in self-diagnosis and in deciding whether professional help is needed or not. Internet-provided information is a key component in this category.

Customer support tools help the patient with the daily management of the disease or illness. For example, easy-to-use tools to make different measurements (such as blood pressure) belong to this category. Process support tools are related tools, but focus in the self-care scenario more on the total life cycle of the disease. An example might be computerized tools for keeping track of different measurements. Financial administration tools help in managing the disease or illness condition financially.

INTERACTION SCENARIO

In the patient-medical staff interaction scenario, decision support helps to diagnose the patient and to decide what kind of care he or she needs. Now the decision is made in cooperation between the medical staff and the patient, not just by the patient, as in the self-care scenario.

Customer support tools are used in the same way as in the self-care scenario, but there they also help the patient to prepare for the interaction with the medical professionals.

Interaction support tools are active in the situations where the medical staff and the patient meet. This meeting can take many forms: it can be a physical meeting, a virtual meeting through electronic means, or some combination of these. Doctors often want to keep physical meetings technology free, but technology can help even in these cases, say to facilitate communication in the case of sense disabilities or language problems. In a virtual meeting, different systems such as phone or e-mail services belong to this category as well as those different applications of telemedicine.

Consultation support tools and preparation tools are to be used by the medical staff. Consultation support tools help the medical staff members during an individual customer consultation, whereas preparation tools are active outside the actual customer interaction. An example of consultation support systems might be a system used to support surgery operations.

In our analysis, the information domains are important. Customer support tools contain information for the patient and are used by him or her. Interaction support tools contain joint information and are used by both patients and medical staff. Consultation support tools contain patient-specific information that is targeted just for the medical staff.

To clarify the proposed value chain, take the example of a laboratory test. For the test, the customer might have to prepare himself or herself by not eating anything. The customer might obtain this information the hospital's Web site or a customer-support tool. An electronic patient card that would help in identifying the patient during the testing would be an interaction support tool. Analyzing the test would be a back-office function for the preparation tools. Analyzing the laboratory results and working out how to tell the results of the test to the customer in a comprehensive way could be facilitated by a consultation support tool.

Process support tools in the interaction scenario help in organizing the interactions between the patient and the medical staff (e.g., different consultation time reservation, queue management systems). Financial administration of the patient-medical staff interaction is one part of this. Electronic systems for handling prescriptions also belong to this category.

Finally, a health care organization has to maintain a lot of management routines that are not connected to any individual patients. Resource planning and management systems and statistics keeping are examples of this domain.

The information in Figure 1 is not to say that the different health care systems would be independent isles of automation. All the systems included in the figure should be integrated by a comprehensive electronic patient record.

FUTURE TRENDS

In Table 4, we list the most dramatic changes we will see in the health care sector because of modern IT and the challenges these changes cast on the IT governance.

One of the biggest changes in the industry is that information related to health, diseases,

Table 4. Future changes in the field of health care introduced by modern IT and the challenges to IT governance produced by them

Future changes in the field of health care introduced by modern IT	**Challenges to IT governance**
The amount of free medical information will grow very fast	Mechanisms to rate different information sources have to be developed
Patient self-care will grow in importance	Systems to support self-care must be developed
Patient empowerment will gain in importance	Customer rights and decisions must be tracked in medical information systems
Interaction between patients and medical staff will increasingly turn away from physical meetings to electronic means	Electronic means to facilitate patient-medical staff communication must be developed
Data collected about patients will grow in amount and quality	Data privacy needs even more concern
Medicalization of the society	Medical and other information systems will have to be linked sophistically
Costs of medical care will continue to raise	Cost monitoring will grow in importance
Focus should turn from individual care-taking episodes and consultations to long-term care-taking relationships and processes	Pressures to develop electronic patient records
Resources of the area will not grow as fast as demand	Information systems must be used for process redesign

sickness, and medicines is not scarce. Internet is a rich source of such information, at different levels of expertise, and in different languages. The gap between what information is available and what a health care professional should know is growing fast (Weaver, 2002); similarly, the pressures for laymen to know about medical issues grow. Decision support is needed more than ever in this plentitude of information. Different solutions should be found to differentiate right information from wrong, especially in the case of information targeted for laymen, such as most of the information found on the Internet.

Free information will shift the power balance between health professionals and patients: More often, the patients are the best experts on their disease, and self-care will grow in importance. Systems to support self-care must be developed. Different electronic forums or virtual communities will offer the patients forums to share experiences and peer advice, to the healthy ones as well as to the chronic and acute sick (Utbult, 2000).

If the patients get more empowered, they should too get more responsibility. Health care information systems should not just keep track of medical staff decisions, the decisions taken by the patients should also be recorded to the systems.

Similarly, the interaction between the patients and health care professionals is going to change: Electronic means are going to take share from face-to-face meetings (Cain, Sarasohn-Kahn, & Wayne, 2000; Gilson, 2003). A key survival factor for all the electronic sources of health information and all communication channels to be used by the patients is how they can build and sustain customer/patient trust (Luo & Najdawi, 2004).

Medicalisation of the society (Conrad, 1992) is a strong trend. Increasingly more issues in human life are seen as belonging to the sphere of medical expertise. If and when we accept this trend, medical decision makers need information about our lives in many aspects. As patient data can be electronically cumulated into huge databases, these databases can be used for different statistical, research, and other purposes. This calls for care and proper legislation giving the ramifications for data usage. For example, often the need of integrating data about one's health and social life arises, if effective care is looked for. Needless to say, we enter a difficult area where data privacy issues gain in importance.

Science and industry will continue to produce increasingly effective cure methods and medicines, but not always without increased costs. It is clear that not everyone can be granted the best possible care with public finances. Systems for keeping track of costs and making cost-informed care-taking decisions will be needed in the future.

For organizing patient flows through the health care system modern IT offers many possibilities. A key issue is to follow the long-term development of the patient, not to focus just on short-term episodes of care. This puts pressures on the electronic patient record systems.

Resources can be saved or wasted through care-taking decisions. It is for sure that in many nations the total share of health care costs is at its top already, so current resources should be used more efficiently. Efficiency, among other things, means that patients are taken care of in the best and most effective places, be they public or private, and of right level of expertise. This all calls for extensive process redesign and system support for that. Should patient data be all the time available anywhere though electronic means, would the Healthcare Supply Chain be much more effective (More & McGrath, 2001).

CONCLUSION

Health is undoubtedly among the most important issues for all people. In a modern society, the threats towards health are changing all the time, but at the same time the possibilities to maintain health and to cure illnesses also grow exponentially. The task is to make needs and solutions to meet in an effective way. This is about information and communication technologies and governance structures.

Modern IT allows health care organizations to structure themselves in new, innovative ways and simultaneously to empower the customers to interact with the organizations, with fellow patients and information sources in revolutionary new ways. In this environment, health care professionals too have to adjust their roles.

Managing and building governance structures for IT in health care organizations is not that much different from other organizations. Even in health care organizations the scope and status of information resource management has to be decided. Issues such as sourcing decisions, charging arrangements and data privacy and security issues all deserve their attention. There are, anyway, certain problems that need to be solved in this area:

- Data privacy and security needs are extremely important and might sometimes conflict with optimal care.
- As the area is new, legislation is often lagging behind.
- The field is a meeting place for two strong professional cultures, that of medical doctors and IT professionals, which might bring along difficulties.

REFERENCES

Ammenwerth, E., Gräber, S., Herrmann, G., Bürkle, T., & König, J. (2003). Evaluation of health information systems—Problems and challenges. *International Journal of Medical Informatics, 71*(2-3), 125-135.

Berry, L. L., Mirabito, A. M., & Berwick, D. M. (2004, Summer). A health care agenda for business. *MIT Sloan Management Review,* 56-64.

Beun, J. G. (2003). Electronic healthcare record; a way to empower the patient. *International journal of Medical Informatics, 69*, 191-196.

Broadbent, M., Weill, P., & St. Clair, D. (1999). The implications of information technology infrastructure for business process redesign. *MIS Quarterly*, *23*(2), 159-182.

Cain, M. M., Sarasohn-Kahn, J., & Wayne, J. C. (2000). *Health e-people: The online consumer experience*. Oakland: California Health Care Foundation.

Conrad, P. (1992). Medicalisation and social control. *Annu. Rev.Sociol., 18*, 209-232.

EITO. (1995). *European Information Technology Observatory.*

Gilson, L. (2003). Trustnext term and the development of health care as a social institution. *Social Science & Medicine*, *56*(7), 1453-1468.

Haux, R., Ammenwerth, E., Herzog, W., & Knaup, P. (2002). Health care in the information society. A prognosis for the year 2013. *International Journal of Medical Informatics*, *66*(1-3), 3-21.

IT Governance Institute. (2001). *Board Briefing on IT Governance* (2nd ed.). Retrieved from www.itgi.org

Ko, M., & Osei-Bryson, K.-M. (2004). Using regression spline to assess the impact of of information technology investments on productivity in the health care industry. *Information Systems Journal*, *14*(1), 43-63.

Luo, W., & Najdawi, M. (2004). Trust-building measures: A review of consumer health portals. *Communications of the ACM*, *47*(1), 109-113.

More, E., & McGrath, G. M. (2001). Reengineering the healthcare supply chain in Australia: The PeCC Initiative. In R. Stegwee & T. Spil (Eds.), *Strategies for healthcare information systems* (pp. 114-125). Hershey, PA: Idea Group Publishing.

Morrissey, J. (2003). An info-tech disconnect. *Modern Healthcare, 33*(6), 6-48.

Porter, M. E. (1990). *The competitive advantage of nations*. New York: Free Press.

Ross, J. W., & Rockart, J. F. (Eds.). (1996, *December 16-18*). Enabling new organizational forms: A changing perspective on infrastructure. *Proceedings of the 17th International Conference on Information Sytems*, Cleveland, OH.

Suomi, R. (2000). *Leapfrogging for modern ICT ssage in the health care sector.* Paper presented at the Proceedings of the Eighth European Conference on Information Systems, Vienna, Austria.

Utbult, M. (2000). *Näthälsa. Internetpatiente möter surfande doktore—uppstår konfrontation eller samarbete?* Stockholm, Sweden: Reldok och KFB—Kommunikationsforskninsgberedningen.

Van Grembergen, W., De Haes, S., & Guldentops, E. (2003). Structures, processes and relational mechanisms for IT governance. In W. Van Grembergen (Ed.), *Strategies for information technology governance* (pp. 1-36). Hershey, PA: Idea Group Publishing.

Weaver, R. R. (2002, March). Resistance to computer innovation: Knowledge coupling in clinical practice. *Computers and Society*, 16-21.

WHO. (2004). *The world health report 2004.* Geneve, Switzerland.

KEY TERMS

Electronic Patient Record: All health-related information related to a patient in electronic form, assembled as a single entity.

Electronic Prescription: Prescriptions created and handled in electronic form in an integrated information system.

Health Care Supply Chain: A managed set of activities related to the health care activity of a patient organized so that all information needed is all the time available and that the participants in the chain have a complete picture of the total process.

Picture Archiving and Communication System: A system for capturing, storing and distributing medical images. These systems are fast turning from storing analog images to storing digital images.

Sourcing Decision: Whether to buy goods/services from the market or to make them self—alone or in different alliances.

Virtual Community: A social aggregation on the Internet when people interact long enough to form personal relationships.

This work was previously published in Encyclopedia of E-Commerce, E-Government, and Mobile Commerce, edited by M. Khosrow-Pour, pp. 537-544, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 6.2
Entrepreneurial IT Governance:
Electronic Medical Records in Rural Healthcare

Carla Wiggins
Idaho State University, USA

John Beachboard
Idaho State University, USA

Kenneth Trimmer
Idaho State University, USA

Lela "Kitty" Pumphrey
Idaho State University, USA

ABSTRACT

Governance is traditionally viewed as a formal authority structure. Information Technology (IT) governance parallels this in that it refers to the patterns of authority over IT. However, traditional control theories of governance, particularly those applied to IT, may not apply in rural healthcare settings. Healthcare has significantly underinvested in IT. Even today, the vast majority of healthcare transactions occur via telephone, fax, paper, and EDI; much data is not captured, is captured incorrectly or inefficiently, and is difficult to retrieve and use. Employing an established IT governance framework, and working from the fundamental assumption that IT is a vital tool in the capture and use of healthcare data, the authors explore the appropriateness of the traditional IT governance framework within a rural healthcare setting. We present an innovative, non-authoritative, relationship-oriented approach for those seeking successful adoption of IT as a means of improving healthcare in rural settings.

INTRODUCTION

Healthcare is arguably the most transaction-intense industry in our society. Yet compared to other industries, healthcare has significantly underinvested in information technology (IT). Even today, the vast majority of healthcare transactions occur via telephone, fax, paper, and electronic data interface (EDI). The result of this archaic information communication system is that much data is not captured, is captured incorrectly or inefficiently, and is difficult to retrieve and use (Barber, Caillouet, Ciotti, & Lohman, 1994; Wager, Lee, & Glaser, 2005).

Health information is typically spread throughout the healthcare organization and held in incompatible legacy systems with little or no interconnectivity or interoperability (Pendharkar, Khosrow-pour, & Rodger, 2001). Healthcare executives are focused on improving the quality, reducing the cost, and expanding access to healthcare, but cannot improve what cannot be measured and cannot measure inconsistently captured or inaccessible information that is reported and held in non-compatible homegrown systems and databases.

In an industry where the paper medical record has been considered the "gold standard," Electronic Medical Record systems (EMR) are believed to be an increasingly vital facet for improving patient safety and quality of care, operational efficiency, and compliance with regulations while reducing medical errors and decreasing the risk of lawsuits. Still, EMR is perceived by many to be a money pit rather than a source of efficiency, income, and enhanced quality of care.

United States healthcare is struggling with decision-making, implementation, standardization, and connectivity surrounding the EMR. This is indicative of the unsystematic and independent nature of healthcare organizations in the United States. "The medical industry still has no clear common goals for IT and very few universally accepted standards" (Sobol & Prater, 2006, p. 74). Fewer than one in five hospital information technology (IT) executives report that their organizations have a fully operational EMR. Only 8% of physicians report using computerized order entry systems (CPOE) and only about one-third of U.S. hospital emergency and outpatient departments use EMR ("Study Shows Limited", 2005). In fact, the number of healthcare organizations reporting a functional EMR actually decreased from 19% in 2004 to 18% in 2005 (Lawrence, 2005).

Perhaps no single industry is as complex and convoluted in its structure, process, and "product" as the U.S. healthcare industry. It is constituted from a tremendously diverse set of public, private, and quasi-public organizations and agencies ranging in size from very small (i.e., solo physician offices) to very large (i.e., integrated health systems), cost reimbursement governmental programs (Medicare and Medicaid), and private organizations (Blue Cross/Blue Shield, and other private insurers). Additionally, it is often said that only the nuclear power industry is more heavily regulated than U.S. healthcare. In this schizophrenic environment, other healthcare organizations are both partners and competitors. Yet, within this complicated and multifaceted industry, organizations strive to meet their missions and serve their patients, constituents, and communities. How are decisions made in such an environment? More specifically, given the expected benefits of modernizing healthcare with information technology, how are IT governance decisions made in what is often referred to as a constant state of chaos?

This manuscript describes a rural family practice residency program that implemented an EMR. The residency program, which trains primary care physicians and provides primary care services to widely disbursed rural communities, received a federal grant for the acquisition and implementation of the EMR, with the simple initial goal of enhancing the practice's clinical research capabilities. As the purchase and implementation of the EMR progressed, however, the practice's simple research goal mutated and morphed into a much larger goal of extending the system throughout rural clinics and providers in the region.

IT GOVERNANCE IN HEALTHCARE

Traditionally, governance is "defining and realizing missions and goals, establishing strategic direction, policies and objectives to that end, and monitoring implementation" (McNally, 2003). Governance commonly concerns the patterns of authority that determine the use of organization resources and the integration of differences among organizational interests (Daily, Dalton, & Cannella, 2003; Sundaramurthy & Lewis, 2003).

Viewed as a set of formal authority relationships, an organization's Board of Directors or Trustees governs through its relationship with top management, and in turn, top management governs departments through its relationships among various subunits. IT governance parallels corporate governance in that it refers specifically to the patterns of authority over IT resources and the means for integrating IT interests. IT governance decisions determine the design of the technical infrastructure, the form of application management, and the alignment of the organization's corporate strategy and integrated information practices (Sambamurthy & Zmud, 1999).

Although the terms are often used interchangeably, IT governance and IT management are related, yet separate, endeavors. "Whereas the domain of IT management focuses on the efficient and effective supply of IT services and products, and on the management of IT operations, IT governance faces the dual demand of (1) contributing to present business operations and performance, and (2) transforming and positioning IT for meeting future business challenges" (Peterson, 2004, p. 44). Perhaps the trickiest aspect of IT governance is leaping that fence that divides the technical from the managerial world. While IT professionals certainly must go a long way toward being able to positively communicate with business process owners, functional managers need to invest a similar effort in understanding the terminology used within IT with a goal of recognizing how some seemingly technically focused decisions have real business consequences.

Weill and Ross (2004) regard the establishment of IT principles as the most fundamental governance decision an organization must make regarding the employment of IT. IT principles embody executive level-attitudes regarding the role information technology is to play in supporting the organization's mission and strategy. For example, while some healthcare professionals may view IT as a "necessary but evil" expense, other healthcare professionals may view IT as providing a strategic investment that can not only reduce the cost of providing healthcare but also ultimately improve the quality of healthcare provided. These attitudes are directly reflected in the organization's IT governance principles and policies and establish a basis for the identification of applications that the enterprise desires to implement. In the healthcare industry, applications directly contribute to patient care as well as support the organization's administrative functions. Decisions concerning IT principles and application needs drive the technical decisions concerning the organization's IT architecture and IT infrastructure.

Finally, the organization must make IT investment and prioritization decisions. Not only must management decide which software applications will best serve their organization's needs, but they must also determine how much to invest in the IT infrastructure.

These five areas (establishment of IT principles, identification of applications, IT architecture, IT infrastructure, and IT investment and prioritization) are the most important decisions made under the conceptualization of IT governance employed in this manuscript. While architecture and infrastructure decisions may be quite technical, organizations with well-designed IT governance mechanisms ensure that the IT principles and application needs drive design and infrastructure investment.

It is important to note that Weill and Ross (2004) are less concerned with what decisions are made than with ensuring that organizations have effective mechanisms in place to make these decisions. In their study of IT governance processes at 256 enterprises, they found the ability of senior management to accurately describe how these decisions were made proved to be the best predictor of high governance performance (Weill, 2004). Similarly, Schweiger, Melcher, Ranganathan, and Wen (2006) conclude from their Family Medical case that "communication to employees, both implicit and explicit, plays a significant role in the diffusion process" of IT (p. 90).

The need for management to describe decision-making processes is similar to the McGinnis, Pumphrey, Trimmer, and Wiggins (2004) argument that traditional control theories of governance applied to IT may not apply in rural healthcare settings. Rather, these authors posit that governance is a pattern of social relations integrating organizational activities. Contrary to the dominant paradigms of hierarchy, power, and resource based governance, coordination and communication appear to provide the most effective mechanism for IT governance for at least some healthcare organizations.

Healthcare organizations face uniquely challenging circumstances in establishing effective IT governance mechanisms required for the infusion of information technology into the day-to-day provision of medical services. Cost remains an almost insurmountable barrier to IT adoption, particularly in medical groups; many practices simply cannot afford the upfront costs. Medical practices also face physician and staff resistance to the use of EMR systems (Darr, Harrison, Shakked, & Shalom, 2003). EMRs "have been over-engineered and are not intuitive, forcing physicians to spend more time clicking through screens and menus to get their work done" (Brown, 2005, p. 48).

While healthcare lags other U.S. industries in IT adoption, IT is increasingly vital to healthcare's efforts to fulfil its missions and reach performance and productivity goals (Goldsmith, Blumenthal, & Rishel, 2003). The provision of high quality care can be enhanced and better monitored by the effective and appropriate use of IT (Zabada, Singh, & Munchus, 2001). According to the Institute of Medicine (1998), safety, effectiveness, patient centeredness, timeliness, efficiency, and equitability are the characteristics necessary for delivering excellent patient care. Each of these vital characteristics can be enhanced, monitored, and improved by the appropriate use of IT via effective governance.

Within the context of IT, an entity must continuously adjust to a rapidly changing environment or risk falling behind the competition. Healthcare entities have turned to, and have become increasingly dependent upon, IT as a means of creating efficiencies and providing more effective service.

METHODOLOGY: AN INTERPRETIVIST CASE STUDY DESIGN

This study employs a single-site case study design to provide a fine-grained examination of the creation and operation of an EMR system in a university affiliated medical practice providing healthcare services to a rural population. We take an interpretive philosophical approach illustrated by an analogy developed by Slife and Williams (1995). In describing interpretive ways of knowing, they ask the reader to consider the difference between a map of a city and an informal account of that city provided by a resident. The map, while admittedly an interpretation, represents an abstraction of an objective reality, depicting "only those features of the place that would remain unchanged even *if no one lived there* (e.g., patterns of streets, layouts of buildings)." The informal account from the city resident is quite different although ultimately as informative, perhaps more so. While "necessarily personal, incomplete and biased…" the personal account might describe the best places to eat or sections of town to avoid. By providing such description, the informal account "gives *meaning* to the town, from a native's point of view." Certainly it is not the only description or possible interpretation but it is legitimate nonetheless.

As are many scholars, we are interested in understanding IT within the healthcare industry. The selection of a single-site case study is particularly appropriate when investigating complex social phenomena where establishing appropriate boundary conditions between the phenomenon of interest and its environment is problematic (March, Sproull, & Tamuz, 1991; Stake, 1994; Yin, 1994). The selection of a single-site case study was further warranted given the family practice residency's status as an early adopter of IT within its segment of the healthcare industry.

Research activities and processes included in this study can be grouped into the following categories:

- **Data collection:** While documentary evidence was examined, the primary data sources were 14 face-to-face interviews with seven key informants at the practice. The selection of participants was purposive, and snowball techniques identified additional informants. Key informants are listed in Table 1. Interviews were organized around the research question but were conducted using the "active interview" approach (Fontana & Frey, 1994; Holstein & Gubrium, 1994).
- **Data analysis:** Analysis consisted primarily of the creation of a case narrative that seeks to accurately reflect interviewee perceptions. The objective was to develop an accurate and rich description of a phenomenon as seen through the eyes of the study participants. Related literature is referenced to provide useful context.
- **Validity and reliability assessments:** The authors used the results of the unstructured interviews to develop narrative themes representing participants' varied perspectives. Where possible, we used participant checks to test whether our interpretation of interview data fairly represents participant views (Altheide & Johnson, 1994; Miles & Hubrman, 1994; Patton, 2002).

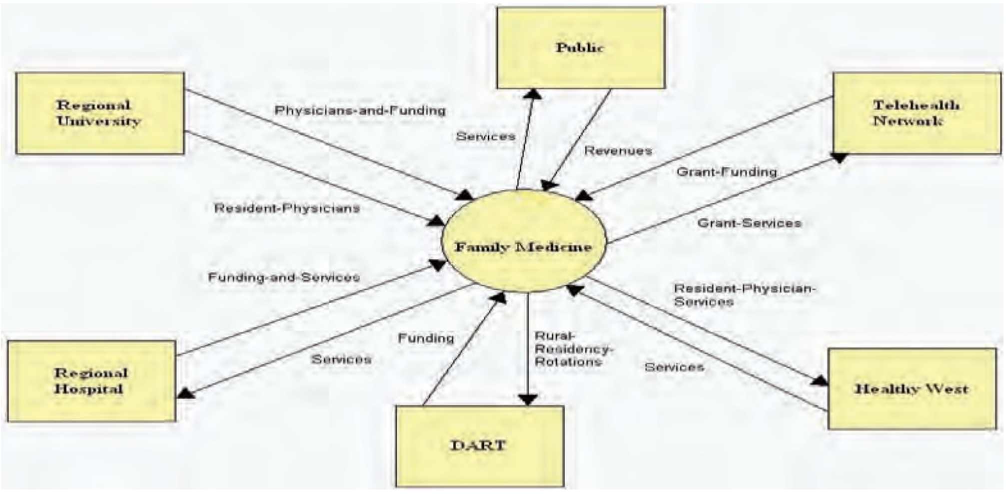
Rather than conducting some variation of discourse analysis typically associated with interpretive studies, the authors' construction of the case narrative presented in Section 4 should be read as a case history and Section 6 as historical interpretation or analysis consistent with a historical and ante-narrative method recommended by Dalcher (2004), which is particularly appropriate for the study of information system failures. While we do not perceive the subject implementation effort as a failure, the knowledge as it concerns IT gov-

*Table 1. List of key informants**

Informant	Position
Joseph (Joe) Clark, M.D.	Director of FM
Robert (Bob) Wood, M.D.	Associate Director of FM
Rita Ford, Pharm D.	Director of Grants, Research and Information Systems
Anne Wright	Director of Information Technology
John (Jack) Hopkins, M.D.	Third Year Resident Physician
James (Jim) Wilson, M.D.	First Year Resident Physician
Mary Miller	Accounts Receivable Manager

** Fictitious names used to preserve participant's anonymity.*

Figure 1. Family medicine relationships



ernance practices is "fragmented, distributed and hidden with the context" (Dalcher, 2004, p. 306). Dalcher argued that traditional research methods are unsuitable for studying information system failures because of the emergent properties inherent in such failures and an inability to "clearly delineate causes and effects" (Dalcher, 2004, p. 306). We simply are extending Dalcher's methodological recommendation to include a wider variety of IS implementation experiences.

These procedures demonstrate the authors' commitment to employing disciplined data elicitation and analysis techniques consistent with recommendations of leading qualitative researchers and methodologists.

FAMILY MEDICINE

Family Medicine (FM) is a health provider in the intermountain region of the Rocky Mountains that also serves as a Family Practice Residency Program (RP). RPs for family medical practice typically require three years of post-doctoral experience. FM's RP is a department in the College of Health Professions at the state university. FM is located in a city of 55,000, a regional hub for farming and ranching communities. In addition to providing family medicine, the residency physicians also participate in outreach to rural healthcare clinics within the region, rotating within the five rural clinics managed by another partner, Healthy West (HW). The residency program at FM is also part of the Demonstration: Assistance in Rural Training (DART) project, a regional funding program for residency rotation in the rural environment.

FM is managed by a Director, Joe Clark, M.D. who is on the faculty of the university and on the medical staff at Regional Medical Center (RMC). All physicians and residents at FM are on both the university faculty and hospital staff.

There are two additional upper management positions at FM. Bob Wood, M.D., Associate Director and Rita Ford, Pharm D. is Director of Grants, Research, and Information Systems. Wood and Ford had been pursuing acquisition of an electronic medical record (EMR) system with numerous unsuccessful grants. A principal justification for the EMR was quality data for research purposes. An opportunity arose for FM to address its EMR needs by participating in a grant request by the Telehealth Network (TLN) of the university, as TLN also shared in the rural healthcare mission. The environment that FM operates within is represented in Figure 1.

A time line of the events involved in the selection and implementation of the system is represented in Table 2. A brief discussion follows.

After receiving the grant, Clark initiated an investigation into EMR vendors. A committee of physicians experienced with EMR evaluated a number of systems, narrowing them to three viable, wireless enabled, EMR systems. Following system selection, Clark hired Anne Wright who designed, acquired, and built the wireless network that would serve as the systems backbone. Next, a third-year resident, Dr. Jack Hopkins, spent a week at a medical practice that had recently implemented the selected system, General Electric's Centricity.

After Hopkins return, Wright planned the implementation of the EMR. First, she began to build interfaces with the hospital's system to enable Mary Miller, accounts receivable manager, to phase in the billing system. Miller spent approximately three months making certain that charges generated by both FM and the hospital for FM patient services were accurately captured and billed. At the end of this reconciliation period, FM began to invoice patients and insurance providers and maintain accounts receivable independent of the hospital billing system. Next, Wright began to build interfaces between the hospital's laboratory, radiology, and pharmacy to provide the systems charting functions with supporting data. Third, patient data was input as data entry personnel were being trained on the system. Finally, Hopkins assisted in providing training on using the EMR to FM's other physicians.

As the initial EMR implementation at FM was commencing, Wright began to extend the system to the homes of physicians and residents. For example, first-year resident Jim Wilson received training in the EMR at the clinic from Hopkins, and had his home system configured for terminal server services by Wright. Wilson, a native of the region, will have received almost all of his residency training with an EMR, and can do his paperwork and patient research from his home as well as the clinic.

Table 2. Sequential events in electronic medical record implementation at family medicine

Event	**FM Participants in study**	**Actions**
Grant Process	Wood, Ford	Grant received for systems in both statewide Residency Programs.
System Selection	Clark	Advising group formed, EMRs assessed, EMR software selected. IT Director hired.
System Implementation	Wright, Hopkins, Miller, Wilson, Clark	Hardware selected, visits made, system installed, users trained, data input, system maintained, common system configuration implemented, chart room eliminated, transcription eliminated. Advising group formed for system implementation conflict resolution.
Implementation Follow-up	Wood, Wright, Wilson	Quality of data source, addition of information technology assistant, benefits of broad system availability, additional time for file maintenance.

An overriding strategy for systems implementation governance was initiated by Clark. The advisors from system selection had warned Clark about conflicts that frequently arose during the implementation of EMRs. To alleviate this, Clark formed an implementation advisory committee which consisted of community members, hospital representatives, and a network of associates who had been "down the path" of EMR implementation. This committee enabled Wright to mitigate opposition to specific system implementation issues by providing firm guidance in areas of physician and staff systems conflict.

All paper medical records and charts were eliminated from FM six months after the initiation of system conversion. One year after the initiation of the conversion, Wood noted that a major benefit of the EMR was chart availability; no longer did physical charts need to be located and physically shared by one caregiver at a time. A downside to this, mitigated by the elimination of the transcription department, was the time it took to input the documentation and assessment of patient visits. Wood also commented that the conversion from the paper system to the EMR made them aware of inaccuracies in the previous chart system.

Twelve months after initiating the system conversion, Wright saw herself in an enhanced role. The remaining funding from the grant was being used to implement a similar system in the state's other Residency Program. Wright found herself as a consultant for this project in addition to supervising an IT support person at FM.

DISCUSSION

While this study was originally conceived of as a traditional IT implementation case, the authors' attention gradually was drawn to the *ad hoc*, almost serendipitous, manner in which the EMR project was conceived, developed, and implemented. Having a familiarity with the IT governance literature, as well as recognizing the common perception that the health industry has been slow in its adoption of information technology, we refocused our analysis and conducted follow-on interviews to obtain a better understanding of the influence of formal and informal IT governance mechanisms on the formulation and execution of the EMR initiative. In this section, we extend our discussion of relevant IT governance literature and applicable U.S. healthcare policy to further lay the groundwork for our identification and analysis of four issues to be addressed in the development of IT governance mechanisms particularly relevant to the diffusion of IT-enabled healthcare services in the rural United States.

The benefits expected to be derived from information technology are so significant that in 1996, the U.S. government enacted the Health Insurance Portability and Accountability Act (HIPAA). The act directed the U.S. Department of Health and Human Services to provide a set of rules governing health information. The rules have been created to standardize the communication of electronic health information between and among healthcare providers to protect and secure private individually identifiable health information. Notably, the legislation does not recognize potential risks associated with establishing an EMR system, nor does it include significant provisions intended to safeguard patient privacy (Fedorowicz & Ray, 2004).

Despite the legislative mandate of HIPAA, progress in the implementation of EMR capabilities has been slow (Lawrence, 2005) and resistance exists to the use of EMR systems, particularly on the part of physicians (Darr et al., 2004). Furthermore, studies exist indicating that practicing physicians and other healthcare professionals have reason to be cautious concerning much of the hyperbole expounding the cost savings and healthcare benefits to be derived from EMR systems. In a seminal article, Eric Brynjolfsson failed to find any significant correlation between increased IT investments and multiple measures of financial success indicating that IT investments do not always deliver the anticipated benefits (1994). While there have been many explanations offered for his unexpected and counter-intuitive finding, the conclusion that must be drawn is that organizations are capable of making good and not so good investments in information technology.

The research on IT governance has provided important insights into an organization's ability to effectively translate IT investment into desired outcomes. Weill et al. (2004) declared "IT value depends on more than good technology" and have concluded "effective IT governance is the single most important predictor of the value an organization generates from IT" (Weill et al., 2004, p. 3-4). Accordingly, an examination of the IT governance function associated with the implementation of EMR at FM practice should prove useful.

As previously discussed, Weill et al. (2004) identify key IT governance decisions: IT principles, IT architecture, IT infrastructure, business application needs, and IT investment priorities. Without intending criticism of Weill et al.'s work, the FM case does not easily conform to the model depicted in their governance arrangement matrix. An analysis of this lack of fit provides insight into the unique challenges faced within the healthcare industry and suggests a need for further elaboration of IT governance theory to better accommodate the needs of enterprises operating in complex inter-organizational environments.

To explore this issue further within the context of our case study, we focus our discussion on four important areas: defining the enterprise, scale and resources, social governance, concluding with some cautionary observations.

1. Defining the Enterprise

While not discounting the emergence of complex inter-organizational relationships emerging in many commercial industries, establishing or defining pragmatic organizational boundaries within the health industry is uniquely difficult. As previously outlined, FM is a cooperative venture jointly supported by a public university and county owned regional medical center. Furthermore, FM, by hosting Family Practice residents, must maintain a relationship with and support the standards of the broader medical education community and the DART regional medical consortium.

The governance models proposed by Ross and Weill apply to a more cohesive organizational form where enterprise boundaries, while complex, are still generally identifiable. The quasi-public and heavily regulated organizational environment in which FM operates simultaneously permits a degree of independence and entrepreneurship with respect to FM internal operations and initiatives yet FM remains answerable to and under the jurisdiction of two much larger and independent institutions (i.e., the regional medical center and the state university). Despite the complex set of inter-relationships, FM was able to act in an entrepreneurial manner in initiating this EMR project. Lacking many of the formal policies and decision-making mechanisms associated with larger enterprises, FM established an advisory committee to participate in making substantive project decisions. The use of this *ad hoc* advisory committee additionally served to promote the active cooperation of its larger governing institutions.

As enterprises increasingly cooperate in the development of standardized IT-enabled processes, IT governance models will likely require elaboration to reflect the external entities (e.g., industry standards organizations) that may strongly influence application and infrastructure development dictating the magnitude and allocation of IT resources. We do not mean to imply that organizational management will be able to abrogate its IT governance responsibilities. On the contrary, greater effort will be required to monitor and perhaps attempt to influence external developments so that organizational needs can be met. While much of this effort will necessarily lie within the technical domain, the significant intertwining of technology and business (healthcare) processes will necessitate the involvement of medical and administrative professionals.

2. Scale and Resources

Despite its close association with two multi-million dollar enterprises, FM is a relatively small, independent organization that, on a general scale and scope, resembles many private physician practices. On average, FM sees 60 patients per day with an average of 15,660 patient visits per year. FM's current practice size is 12,000 patients.

There are 14 staff physicians at FM, whose salaries are funded by the university and RMC, as is Clark's salary. The number of program graduates is 50, half of whom are practicing in rural communities. Presently, there are 17 Family Practice residents in the Residency Program. The remaining staff of 20 employees is funded by RMC. Without the availability of the TLN grant, this initiative could not have been undertaken. There was no anticipated financial return on EMR investment for the FM. Certainly, there are various financial savings resulting from reducing some of the labour intensive administrative processes (e.g., transcription of physician notes and maintaining paper medical records). However, these savings were not expected to cover the full cost of the software application and expanded IT infrastructure.

In addition to supporting FM's research mission, the benefits of EMR are expected to accrue over time in the form of improved healthcare for patients for a variety of reasons more fully described in other literature (see for example Amatayakul, 2001; Brown, 2005; Chung, Choi, & Moon, 2003; Hough, Chen, & Lin, 2005; Snyder, Paulson, & McGrath, 2005; Rosenbloom, Grande, Geissbuhler, & Miller, 2004; Stausberg, Koch, Ingenerf, & Betzler, 2003). The expected quality of healthcare benefits are certainly non-trivial. However, the extent and timing of improved healthcare are not known with certainty and have not been quantified. Furthermore, given the structure and rigidity of Medicare, Medicaid, and third party reimbursement policies, the cost of improving healthcare cannot be passed on to the majority of FM's patients.

In short, it is difficult for small and independent practices to justify making investments in EMR particularly until the full costs of implementing and operating an IT infrastructure with requisite degrees of performance, reliability, and security are truly understood. The Weill and Ross IT governance model primarily reflects the profit-making motivation of modern commercial enterprises. Given the diverse participants, in both size and profit motivation, and given the public good anticipated to result from more effective IT employment within the health services industry, the establishment of effective IT principles becomes paramount. Healthcare participation in the development of IT principles will likely be expanded beyond the decision archetypes identified.

3. Social Governance

FM faces complex governance challenges. It operates under the supervision of two significant yet independent institutions, the university, and the regional medical center. It is obliged to support the disparate needs of these two institutions, yet operates largely as an independent practice. The convoluted governance structure has contributed to the evolution of socially constructed entrepreneurial governance processes focused on identification and accomplishment of jointly held objectives in a manner that virtually defies clear explanation. The social collaborative processes conducted among members of these distinct institutions and within the FM practice provided the basis for the entrepreneurial approach observed in the launching and execution of the EMR project.

The initial motivating idea for launching the EMR project at FM was straightforward and needs driven. A practicing physician with research and educational responsibilities desired a better way to analyze data generated by the FM practice. Lacking resources, a search for grant funds resulted. In seeking the grant funds, the researcher's needs were aligned with state and federal initiatives to improve rural healthcare, which legitimately related to the mission of the practice. In evolving its project goals to meet this broader objective, FM management displayed flexibility and opportunism reminiscent of the "garbage can" model of decision making described by Cohen, March, and Olsen (1972).

The social collaborative culture of the enterprise was further demonstrated by the *ad hoc* creation of the advisory board that assisted in the refinement of technical and functional requirements, evaluated alternatives and recommended the product that was finally purchased. In addition, a separate *ad hoc* group was utilized to assist in the actual system implementation. As internal participants negotiated the configuration and use of the application, a group of physician practitioners experienced in using EMR was recruited. This second committee counselled the practice concerning implementation and use of specific features and provided an independent third party to deflect, depersonalize, and resolve conflicts among the practice's staff.

The pattern of decision-making appears to fit with the McGinnis et al. (2004) work specific to the healthcare industry. In that study of rural healthcare providers, the authors posited that overall institutional governance reflected a pattern of social relations employed to integrate organizational activities. Both cases emphasize the importance of informal organizational mechanisms and are consistent with other research investigating strategic IT alignment (Chan, 2002). The importance of emergent social relationships contrasts with the structural determinants typically described in governance research. Tellingly, Weill and Ross concluded that the specific pattern of IT governance mattered less than the ability of study participants to accurately describe the governance processes used within their respective organizations (2004). While not explicitly addressing the issue of informal organizational relationships, their findings appear to indicate that a strong degree of "transparency" within the governance processes is indicative of more effective IT governance. It becomes a question for future research as to whether such transparency is associated with the development of strong social relationships identified by McGinnis et al. (2004).

4. Cautionary Notes

Thus far we have focused on the positive aspects of the FM's EMR initiative. Yet the ad hoc approach to the formulation of this initiative has engendered risks as well.

The two physicians most responsible for the initiation of this effort, Clark and Wood, confess to having a relatively limited knowledge of information technology. While this undermines any argument that technology is being chased for technology's sake, there was a risk that the resource and management commitments required to maintain reliable and secure services, once implemented, would not be fully understood.

A technical lead was not hired until after the grant had been awarded. Fortunately, FM was able to bring on an exceptionally talented individual. Admittedly, the previous statement smacks of hyperbole, but we shall explain. One person assumed overall technical responsibility for:

- Design, specification, and implementation of required IT infrastructure.
- Installation, configuration, and customization of EMR applications and data base management systems (DBMS).

- Complete oversight of the execution of migration from paper to electronic records.
- Coordination and development of customized interfaces with regional medical center applications (e.g., billing, scheduling, and laboratory applications).
- Development and implementation of security architecture including compliance with HIPAA guidelines and regulations.
- Provision of user training and being FM's help desk.

Essentially, one individual is serving as strategic IT planner, systems (including servers, desktops, tablet PCs, network, DBMS, and applications) architect, project manager, systems administrator (for all systems previously identified), trainer, and help desk.

Therein lies a serious problem. Beyond the apparent difficulty in identifying an individual possessing requisite technical and managerial knowledge and skills to accomplish these tasks, there is an important contingency that a single individual is unable to address. That is, Anne Wright constitutes a single and absolutely critical point of failure. While this type of problem is not unheard of even in very large enterprises, it far too common in small healthcare practices. This "single point of failure" structure is evidence of weaknesses in the IT governance processes in this study and has significant implications for the healthcare community as discussed next.

CONCLUSION AND LIMITATIONS

In summary, governance serves to coordinate organization decisions and activities both vertically from top to bottom and laterally across organization functions and departments. Thus, governance is the mechanism that assures that strategy formulation and implementation produce desired organizational performance. However, in the chaotic U.S. healthcare environment, where local and regional health organizations are both competitors and partners, we suggest that IT governance theory should be extended to explicitly address the implications of evolving industry technology standards and practices. Furthermore, governance theory must also recognize the "public good" nature of health services and the impact of government legislation and regulation on the development and evolution of enterprise IT principles. Finally, even within the context of strong industry and legislative influence, governance theory should recognize that an entrepreneurial style of governance exist among and between organizations.

Further research is needed to determine whether this governance approach is suited to other U.S. healthcare organizations or other nations. It is generally accepted that the U.S. healthcare system differs from that of almost all other industrialized nations in that the rest of the industrialized world has socialized healthcare or socialized insurance. However, the lessons learned regarding the strategies employed by the Family Medicine residency program in IT governance may fit other rural healthcare organizations.

The exploratory nature of this work is a limitation. This is a preliminary assessment of IT governance in one specialized setting—a family practice residency program associated with a state university. Using qualitative date from a descriptive case study means care must be taken in generalizing our findings and conclusions to other rural healthcare facilities. However, this preliminary work will pave the way for other studies of IT governance in rural healthcare. Using a variety of qualitative and quantitative techniques, researchers may wish to further examine issues of leadership, decision making, coordination, and entrepreneurship in governance.

The literature discussing EMR identifies significant potential benefits to be derived from effective EMR implementation. IT governance research strongly suggests that a lack of executive involvement in IT governance decisions (i.e., deferring too many decisions to the IT staff, significantly increases the risk of not deriving expected value from IT investments). Healthcare professionals expend tremendous time and energy in developing the knowledge base required to provide healthcare services. There is an understandable reluctance on the part of many such professionals to invest even more time and effort in learning the capabilities and limitations of information technology. However, as information technology becomes increasingly embedded in the day-to-day processes of providing health services, such an investment will be required. The fact is that not only does information technology have the potential to significantly improve patient care; misuse can be expensive, dangerous, and potentially deadly.

The impact of this governance strategy on other rural healthcare providers in the region needs to be assessed in the future. Perspectives on improved patient healthcare provided by the EMR should be assessed from the physician, professional/non-clinical staff, and patients involved in the rural environment.

REFERENCES

Altheide, D. L., & Johnson, J. M. (1994). Criteria for assessing interpretive validity in qualitative research. In N. K. Denzin & Y. S. Lincoln (eds.), *Handbook of qualitative research* (pp. 485-499). Thousand Oaks, CA: Sage.

Amatayakul, M. (2001). EMR: The coming of age. *IT Healthcare Strategist, 3*(10), 1-4.

Barber, N. F., Caillouet, L. P., Ciotti, V. G., & Lohman, P. M. (1994). Experts debate the future of healthcare computing. *Healthcare Financial Management, 48*, 66-74.

Brown, N. (2005). Driving EMR adoptions: Making EMRs a sustainable, profitable investment. *Health Management Technology, 26*(5), 48-49.

Brynjolfsson, E. (1994). The productivity paradox of information technology. *Communications of the ACM, 36*(12), 67-77.

Chan, Y. E. (2002). Why haven't we mastered alignment? The importance of informal organizational structure. *MIS Quarterly Executive, 1*(2), 97-112.

Chung, K., Choi, Y. B., & Moon, S. (2003). Toward efficient medication error reductions: Error-reducing information management systems. *Journal of Medical Systems, 27*(6), 553-61.

Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A garbage can model of organizational choice. *Administrative Science Quarterly, 17*(1), 1-25.

Daily, C. M., Dalton, D. R., & Cannella, A. A. (2003). Corporate governance: Decades of dialogue and data. *Academy of Management Review, 28*(3), 371-382.

Dalcher, D. (2004). Stories and histories: Case study research (and beyond) in information systems failures. In M. E. Whitman & A. B. Woszczynski (eds.), *The handbook of information systems research* (pp. 305-322). Hershey, PA: Idea Group Publishing.

Darr, A., Harrison, M. I., Shakked, L., & Shalom, N. (2003). Physicians' and Nurses' reactions to electronic medical records: Managerial and occupational implications. *Journal of Health Organization and Management, 17*(5), 349-359.

Fedorowicz, J., & Ray, A. W. (2004). Impact of HIPAA on the integrity of healthcare information. *International Journal of Healthcare Technology and Management, 6*(2), 142-157.

Fontana, A., & Frey, J. H. (1994). Interviewing: The art of science. In N. K. Denzin & Y. S. Lincoln (eds.), *Handbook of qualitative research* (pp. 262-272). Thousand Oaks, CA: Sage.

Geertz, C. Local (1985). *Knowledge: Further essays in interpretive anthropology*. New York: Basic Books.

Goldsmith, J., Blumenthal, D., & Rishel, W. (2003). Federal health information policy: A case of arrested development. *Health Affairs, 22*, 44-55.

Holstein, J. A., & Gubrium, J. F. (1994). Phenomenology, ethnomethodology, and interpretive practice. In N. K. Denzin & Y. S. Lincoln (eds.), *Handbook of qualitative research* (pp. 262-272). Thousand Oaks, CA: Sage.

Hough, C. B. H., Chen, J. C. H., & Lin, B. (2005). Virtual health/electronic medical record: Current status and perspective. *International Journal of Healthcare Technology and Management, 6*(3), 257-275.

Lawrence, S. (2005). *Portable EMRs are still to come*. Retrieved from Eweek.com/print_article2/0,2533,a=147019

March, J. G., Sproull, L. S., & Tamuz, M. (1991). Learning from samples of one or fewer. *Organization Science, 2*(1), 1-13.

McGinnis, S. K., Pumphrey, L., Trimmer, K., & Wiggins, C. (2004). Sustaining and extending organization strategy via information technology governance. *Hawaii International Conference on System Sciences 37*, Big Island of Hawaii, January 5-8.

McNally, D. J. (2003). The board: The roles and responsibilities of nonprofit board membership. *Pennsylvania CPA Journal, 73*(4), 46-49.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.

Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd Ed.).Thousand Oaks, CA: Sage.

Pendharkar, P. C., Khosrow-pour, M., & Rodger, J. A. (2001). Development and testing of an instrument for measuring the user evaluations of information technology in healthcare. *Journal of Computer Information Systems, 41*(4), 84-90.

Peterson, R. R. (2004). Integration strategies and tactics for information technology governance. In W. Van Grembergen (ed.), *Strategies for information technology governance* (pp. 37-80). Hershey, PA: Idea Group Publishing.

Rosenbloom, S. T., Grande, J., Geissbuhler, A., & Miller, R. A. (2004). Experience in implementing impatient clinical note capture via a provider order entry system. *Journal of American Medical Information Association*, *11*(4), 310-15.

Sambamurthy, V., & Zmud, R. W. (1999). Arrangements for information technology governance: A theory of multiple contingencies. *MIS Quarterly, 23*(2), 261-290.

Schweiger, D., Melcher, A., Ranganathan, C., & Wen, H. J. (2006). Applying adaptive structuration theory to health information systems adoption: A case study. *International Journal of Health Information Systems and Informatics*, *1*(1), 78 -91.

Slife, B. D., & Williams, R. N. (1995). *What's behind the research? Discovering hidden assumptions in the behavioural sciences.* Thousand Oaks: Sage.

Snyder, K. D., Paulson, P., & McGrath, P. (2005). Improving processes in a small health care network: A value-mapping case study. *Business Process Management Journal, 11*(1), 87-99.

Sobol, M., & Prater, E. (2006). Differences in computer usage of U.S. group medical practices: 1994 vs. 2003. International Journal of Health Information Systems and Informatics, *1*(1), 64-77.

Stake, R. E. (1994). Case studies. In N. K. Denzin & Y. S. Lincoln (eds.), *Handbook of qualitative research* (pp. 236-247). Thousand Oaks, CA: Sage.

Stausberg, J., Koch, D., Ingenerf, J., & Betzler, M. (2003). Comparing paper-based with electronic patient records: Lessons learned during a study on diagnosis and procedure codes. *Journal of*

American Medical Information Association, 10(5), 470-77.

Study Shows Limited Use of Electronic Medical Records. (2005). *Healthcare financial management, 59*(5), 27.

Sundaramurthy, D., & Lewis, M. (2003). Control and collaboration: Paradoxes of governance. *Academy of Management Review*, *28*(3), 397-415.

Wager, K. A., Lee, F.W., & Glaser, J. P. (2005). *Managing healthcare information systems.* San Francisco, CA: Jossey-Bass.

Weill, P. (2004). Don't just lead, govern: How top-performing firms govern IT. *MIS Quarterly Executive*, *3*(1), 1-17.

Weill, P., & Ross, J. W. (2004). *IT governance: How top performers manage IT decision rights for superior results.* Boston, MA: Harvard Business School Press.

Wiggins, C., Pumphrey, L., Beachboard, J. C., & Trimmer, K. (2006). Entrepreneurial governance in a rural family practice residency program. *Hawaii International Conference on System Sciences 39*, Kauai Island of Hawaii, January 4-7.

Yin, R. K. (1994). *Case study research: Design and methods* (2nd Ed.). Thousand Oaks, CA: Sage.

Zabada, D., Singh, S., & Munchus, G. (2001). The role of information technology in enhancing patient satisfaction. *The British Journal of Clinical Governance, 6*(1), 9-16.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 1, Issue 4, edited by J.N.D. Gupta, S.K. Sharma, and J. Hsu, pp. 1-28, copyright 2004 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 6.3
Governance Structures for IT in the Health Care Industry

Reima Suomi
Turku School of Economics and Business Administration, Finland

INTRODUCTION

The pressures for the health care industry are well known and very similar in all developed countries (i.e., altering population, shortage of resources for staff and from taxpayers, higher sensitivity of the population for health issues, new and emerging diseases, etc.). Underdeveloped countries experience different problems, but they have the advantage of learning from the lessons and actions that developed countries underwent perhaps decades ago. On the other hand, many solutions also exist, but they all make the environment even more difficult to manage (i.e., possibilities of networking, booming medical and health-related research and knowledge produced by it, alternative caretaking solutions, new and expensive treatments and medicines, promises of biotechnology, etc.).

From the public authorities' points of view, the solution might be easy—outsource as much as you can out of this mess. Usually, the first services to go are marginal operational activities, such as laundry, cleaning, and catering services. It is easy to add information systems to this list, but we believe this is often done without a careful enough consideration. Outsourcing is often seen as a trendy, obvious, and easy solution, which has been supported by financial facts on the short run. Many examples show that even in the case of operational information systems, outsourcing can become a costly option, not to mention lost possibilities for organizational learning and competitive positioning through mastering of information technology.

BACKGROUND

We have found the following reasons for the late adoption of modern information technology in the health care sector (Suomi, 2000):

- Fragmented industry structure
- Considerable national differences in processes
- Strong professional culture of medical care personnel
- One-sided education
- Handcrafting traditions

- Weak customers
- Hierarchical organization structures

ICT and governance structures meet in two ways. On one side, ICT enables new governance structures for the health care industry. On the other, it is an object in need of governing. As both sectors offer a multitude of new possibilities, innovations are called for in the industry (Christensen, Bohmer, & Kenagy, 2000).

IT governance thinking matures in organizations as any other discipline. Van Grembergen, De Haes, and Guldentops (2003) have defined the following stages in their IT Governance Maturity Model:

- Non-existent
- Initial/ad-hoc
- Repeatable but intuitive
- Defined process
- Managed and measurable
- Optimized

Needless to say, in the health care industry, IT Governance thinking is non-existent or initial/ad hoc in the best situation.

THE MEANING OF ICT GOVERNANCE STRUCTURE IN HEALTH CARE

IT is an old acronym for information technology. Nowadays, it is replaced often with the term ICT, referring to information and communication technology. This emphasizes the communication services that are developing very quickly, such as the Internet and mobile services. The letter *C* is often upgraded to the second dimension: alongside communication it can refer to contents. IT or ICT governance is defined (IT Governance Institute, 2001) as follows:

IT governance is the responsibility of the board of directors and executive management. It is an integral part of enterprise governance and consists of the leadership and organizational structures and processes that ensure that the organization's IT sustains and extends the organization's strategies and objectives.

For many, there is a temptation to understand governance as just a synonym for management. This is an oversimplification. Management is a goal-oriented activity, whereas governance is

Table 1. Comparison of terms management, organizational form, and governance structure

	Management	Organizational Form	Governance Structure
Time perspective	Short	Medium	Long
Focus	Action	Internal organization	Inter-organizational structures
Management Control	In action	Easy	Difficult
Metaphor	Communication channels	Infrastructure	Architecture
Character	Concrete	Formal	Abstract

Table 2. ICT governance structure issues in health care

ICT as an enabler
- Health-related information on the web
- Private-public sector co-operation
- Allocation of patients to different levels of care
- Customer contacts distribution between electronic and classical means
- Ownership, structure and allocation of patient, population-level and other critical data
- Electronic forums for patients to interact
- Electronic prescription systems

ICT as an object to be governed
- New legislation needs because of the new data processing possibilities
- Data privacy and security
- Structure and status of the information resource management in health care units
- ICT-general management partnership
- Sourcing decisions of ICT
- Charging arrangements on ICT-services

often given from the outside, and organizations just have to live with it. This is not to say that all governance structures would be beyond management control; management can influence most governance structures, at least in the long run. The long run is a key term in many aspects. We talk about structures that are semi-permanent and not changed very frequently. Structure is a term closer to architecture than to infrastructure; governance structures are architectural terms and are then implemented into infrastructures through different organizational forms. The terms *organization form* and *governance structure* are not synonyms. Organizational forms are more formal and touch upon one organization, whereas governance structures are found in a richer selection of forms and organize themselves over a number of organizations. Table 1 summarizes our discussion.

Governance structures are present in almost any human decision-making situation. In Table 2, we have a collection of key aspects of governance structure issues in health care.

FUTURE TRENDS

One of the biggest changes in the industry is that information related to health, sickness and medicines is not scarce. Internet is a rich source of such information, at different levels of expertise and at different languages. The gap between what information is available and what a health-care professional should know is growing very fast (Weaver 2002). This will shift the power balance between health professionals and patients: increasingly often the patients are the best experts on their disease. Different electronic forums or Virtual Communities (Rheingold 1993) related to health are born on the Internet. They have different services and values to offer to the healthy ones and to the chronic and acute sick (Utbult, 2000). Similarly, the interaction between the patients and health care professionals is going to change: electronic means are going to take share from face-to-face meetings (Cain, Sarasohn-Kahn & Wayne 2000; Gibson 2003).

For organizing patient flows through the health care system modern ICT offers many possibilities. Should patient data be all the time available anywhere though electronic means, would the

Healthcare Supply Chains be much more effective (More & McGrath 2001). Effectiveness means that patients are taken care of in the best and most effective places, be they public or private, and of right level of expertise. As patient data can be electronically cumulated into huge databases, these databases can be used for different statistical, research and other purposes. This calls for care and proper legislation giving the principles.

Managing and building governance structures for ICT in health care organizations is not that much different from other organizations. Even in health care organizations, the scope and status of information resource management has to be decided. Issues such as sourcing decisions, charging arrangements, data privacy, and security issues all deserve their attention. There are certain problems that need to be solved in this area:

- Data privacy and security needs are extremely important and might sometimes conflict with optimal care.
- As the area is new, legislation is often lagging behind.
- The field is a meeting place for two strong professional cultures—medical doctors and ICT-professionals—that might bring along difficulties.

CONCLUSION

Health is undoubtedly among the most important issues for all of us. In a modern society, the threats towards health are changing all the time, but at the same time, the possibilities to maintain health and cure illnesses grow exponentially. The task is to make needs and solutions meet in an effective way. This is about information and communication technologies and governance structures.

Modern ICT allows health care organizations to structure themselves in new, innovative ways, and simultaneously to empower the customers to interact with the organizations, with fellow patients, and with information sources in revolutionary new ways. In this environment, health care professionals also have to adjust their roles.

REFERENCES

Cain, Mary M., Sarasohn-Kahn, Jane, & Wayne, Jennifer C. (2000). *Health e-people: The online consumer experience*. California Health Care Foundation, Oakland, CA.

Christensen, Clayton M., Bohmer, Richard, & Kenagy, John (2000, September-October). Will distruptive innovations cure health care? *Harvard Business Review*, 102-112.

Gilson, L. (2003). Trustnext term and the development of health care as a social institution. *Social Science & Medicine*, *56*(7), 1453-1468.

IT Governance Institute (2001). IT governance executive summary. Retrieved January 10, 2004, from www.itgi.org

More, Elizabeth, & McGrath, G. Mike (2001). Reengineering the healthcare supply chain in Australia: The PeCC initiative. In Robert Stegwee & Ton Spil (Eds.), *Strategies for healthcare information systems* (pp. 114-125). Hershey, PA: Idea Group Publishing.

Rheingold, Howard (1993). *The virtual community – Homesteading on the electronic frontier.* Addison-Wesley, New York.

Suomi, Reima (2000). Leapfrogging for modern ICT usage in the health care sector. *Proceedings of the 8th ECIS conference*, Vienna, Australia.

Utbult, Mats (2000). Näthälsä. *Internetpatienter möter surfande doktorer – uppstår confrontation eller samarbete*. TELDOK Rapport 138. Stockholm.

Van Grenbergen, Wim, De Haes, Steven, & Guldentops, Erik (2003). Structures, processes and relational mechanisms for IT governance. In Wim Van Grembergen (Ed.), *Strategies for information technology governance* (pp. 1-36). Hershey, PA: Idea Group Publishing.

Weawer, Robert R. (2002, March). Resistance to computer innovation: Knowledge coupling in clinical practice. *Computers and Society*, 16-21.

KEY TERMS

Electronic Patient Record: All health-related information related to a patient in electronic form, assembled as a single entity.

Electronic Prescription: Prescriptions created and handled in electronic form in an integrated information system.

Healthcare Supply Chain: A managed set of activities related to the health care activity of a patient, organized so that all necessary information is available all the time and the participants in the chain have a complete picture of the total process.

Sourcing Decision: Decision whether to buy goods/services from the market or to make them self. Sourcing decision can be made as an independent decision unit or as a part of a bigger group.

Virtual Community: A social aggregation on the Internet when people interact long enough to form personal relationships.

This work was previously published in Encyclopedia of Information Science and Technology, Vol. 2, edited by M. Khosrow-Pour, pp. 1305-1308, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 6.4
Changing Healthcare Institutions with Large Information Technology Projects

Matthew W. Guah
Erasmus University Rotterdam, The Netherlands

ABSTRACT

This article reviews the development of institutional theory in direct relations to historical changes within the UK's National Health Service (NHS) with an eye to contributing to the theoretical specification of healthcare information processes. This is done partly by extending certain paradigms (see Meyer & Rowan, 1991; Powell & DiMaggio, 1991; Tolbert & Zucker, 1994) through a proposed model of causes and consequences of variations in levels of institutionalisation in the healthcare industry. It reports findings from a 5-year study on the NHS implementation of the largest civil ISs worldwide at an estimated cost of $10 billion over a 10-year period. The theoretical basis for analysis is developed, using concepts drawn from neo-institutionalism, realisation of business value, and organisational logic, as well as mixed empirical results about the lack of IT investments value in the NHS. The findings suggest that large scale, IT change imposed upon a highly institutionalised healthcare industry is fraught with difficulty mainly because culturally embedded norms, values, and behavioural patterns serve to impede centrally imposed initiatives to automate clinical working practices. It concludes with a discussion about the nature of evaluation procedures in relation to the process of institutionalising IS in healthcare.

INTRODUCTION

An historical overview of IT projects in the UK's National Health Service (NHS) during the last six decades is presented here with the intention to both clarify the links between institutional theory and previous traditions of sociological work on organisational structure. The initial exposition of this theory by works of established institutionalists (Meyer & Rowan, 1991; Scott, Ruef, Mendel, & Caronna, 2000; Tolbert & Zucker, 1994) focuses on the ways of challenging dominant theoretical and empirical traditions in organisational research. While this article clarifies some ambiguity and elaborates on the logical and empirical implications of a phenomenologically based version of institutional theory, the primary aims are to clarify the independent theoretical contributions of institutional theory to analyses of the NHS and to develop this theoretical perspective further in order to enhance its use in empirical research in other healthcare environments (internationally and globally).

Markus (1983) claims that interaction theory draws together three principal strands of resistance: (1) internal factors, (2) technical problems, and (3) political context. This theory has been highly influential in IS strategy and other social sciences generally since Markus first developed the ideas over two decades ago. The focus here (see Table 1) is on how interaction theory offers a new way of looking at IS implementation in the healthcare industry.

Much has been researched in the last few decades about the major lack of a coherent implementation strategy for IS (Sambamurthy & Zmud, 1994) in the healthcare industry (Stevens, Schade, Chalk, & Slevin, 1993; Vogel, 2003). Most of such claims have been levelled against an apparent "productivity paradox" with respect to investments in healthcare management (in general) and IS (in particular). The Wanless report (2002) and Committee on Quality Health Care in America assessment report by Institute of Management (2002)—both national government's mandated investigations into the UK and USA national healthcare systems respectfully—among others, have failed to find a convincing body of evidence that investment in healthcare IS is associated with increased output (refuting the productivity paradox), but not with healthcare value as measured by patient satisfaction.

WHAT IS INSTITUTIONALISM?

Institutionalism is continuously being used to mean different things by researchers of political science, economics, and sociology. Lowndes (1996, p. 182) presents institutionalism as informal codes of behaviour, written contracts, and complex organisations with four elements:

- A middle-level concept. Institutions are devised by individuals and therefore constrain individuals' actions. Institutions here are seen as part of the broad social fabric and medium for individuals' day-to-day decisions and other activities. DiMaggio and Powell (1994) argue that institutions shape human actions, imposing constraints while providing opportunities for individuals.
- Having formal and informal aspects. Lowndes (1996) views institutions to involve formal rules or laws, which allows informal norms and customs to be practiced. That is because some institutions are not

Table 1. Implementation theory: Usage, fitness, relationship and sufficiency

Authors	IS Implementation	Theory Description
Lucas, 1993	Appropriate use of IS	Process theory explaining appropriate IS use Variance theory linking use with business value
Grabowski & Lee, 1993	Strategic fitness of IS	Process-type relationship between strategic fit and performance of IS
Markus, 1983	Relationship of IS assets	How IS investment do or do not become IS assets How IS assets do or do not yield improved organisational performance
Sambamurthy & Zmud, 1994	Insufficient to produce impacts	Process model connecting raw material inputs to outputs Variance theory of IS management competencies and IS impacts Variance theory linking impacts and business value

consciously designed nor neatly specified, yet part of habitual actions by its members. Such institutions may be expressed in organisational form and relate to the processes within.

- Having legitimacy. Legitimacy in institutions goes beyond the preferences of individual actors. Such preferences are valued in them and go beyond their immediate purpose and outputs.
- Showing stability over time. Lowndes (1996) views institutions as gaining their legitimacy due to their relative stability over time, and their links with a "sense of place."

New institutionalists generally view institutions to have "the humanly devised constraints that shape human interaction" (North, 1990, p. 3) what March and Olsen, (1989) refer to as "rules of the game" (p. 162) that organisations and individuals are constantly expected to play the game. Another stand taken by new institutionalists sees informal institutions (tradition, custom, culture, and habit) are embedded in culture and conventions defined as behaviour structuring rules (March & Olsen, 1989; North, 1990). New institutionalists stress embodied values and power relations of institutions together with interaction between individuals and institutions (Lowndes, 1996). They attempt to distinguish between informal institutional rules and personal habits. Such distinction forms the basis for the definition of institution in this research where informal conventions and their impact upon the NHS and its partners are being explored.

Research Methodology

The research study began in 2001, with the initial interest of conducting an exploratory-descriptive study in 10 NHS hospitals to explore why, "historically, the NHS has not used or developed IT as a strategic asset in delivering and managing healthcare" (Department of Health [DoH], 2000). Intensive literature review unveiled few longitudinal studies, which systematically and rigorously examined how IT systems were introduced and changed over time. There were very limited studies that examined inter-organisational relationships between different constituents in the adoption and diffusion of IT systems (NHS directorship, hospital management systems, or IT suppliers and patients). Not only were most of these studies descriptive and lacked a historical dimension, they presented IS in healthcare as largely theoretical with most contributions reporting the findings of a specific IT project implementation using simple success and failure criteria—Scott et al. (2000) being among the most significant contributions.

Using such a relevant and wide-ranging backdrop this research study recognised that it was important to extend the empirical enquiry for two reasons: (1) exploratory-descriptive case studies on a single organisation (or one hospital) would not elicit in-depth and rich data to develop any meaningful analysis and conclusions on how IT was being deployed and managed; and (2) the introduction of a large scale, IT-enabled change program needed to be researched at the wider societal, organisational field and individual levels, covering an extended period of time, to understand the processes of institutionalisation (Tolbert & Zucker, 1994). The research study was therefore designed to capture the myriad of views and opinions about the national program over a 5-year period to build a rich picture of such processes underpinning large scale IT change.

Three methods of data collection were adopted: (1) a range of academic, government, and industry studies on the healthcare sector were assembled—both UK and healthcare services in other countries. The materials proved invaluable for understanding the societal, economic, political, cultural, and technical differences in healthcare nationally and internationally; (2) participation in trade fairs, conferences, workshops, and exhibitions on healthcare—focusing on general or more specific healthcare activities. These events

Table 2. Numbers of interviews conducted

Categories of Interviewees	Year 1		Year 2		Year 3	
	Contacts Made	Persons Interviewed	Contacts Made	Persons Interviewed	Contacts Made	Persons Interviewed
NHS Information Authority	32	5	30	10	10	15
Major IT Service Providers	90	56	60	45	17	12
Primary Care Trusts Admin	15	5	25	12	22	12
Secondary Care Trust Admin	0	0	9	3	7	4
Local NHS IT Managers	15	6	20	11	60	42
Medical Consultants	3	1	8	4	9	6
Nurses & Junior Doctors	13	3	15	3	11	4
Healthcare Researchers	35	20	20	8	10	7
Total Interviews		105		96		102

also generated many useful research contacts that proved invaluable for targeting interviews. (3) A semi-structured interview (see Table 2) schedule was used to enable interviewees to expand on their answers. While most interviews lasted for about 90 minutes, a limited number lasted just under an hour, with nearly all interviews being tape recorded and subsequently transcribed. Respondents were later contacted with feedback from the interviews and, where necessary, errors were corrected. This method of data collection was critical for allowing interviewees to raise additional themes, issues, and concerns that they felt were important to the research study. As a result of the political contention of some of the interview content, some interviewees asked that names of individuals and hospitals be anonymous.

After the initial 6 months of interviews, the scope of the study had to be extended, as it was important to elicit data and information from a wider range of respondents engaged in the implementation of the national program. These included IT service providers bidding for public sector IT contracts and doctors in general practices around the country. Most IT service providers offered critical insights into the political and procurement processes within the NHS and public sector more generally (Guah & Currie, 2004). General practitioners, on the other hand, offered useful insights about the communication channels underpinning the institutional processes underpinning the national program. Given the range of constituents involved, the resulting data were evaluated and interview schedules refined, ensuring questionnaires be more closely targeted to the professional and personal situation of the individual, as generic questions were less meaningful. The final questionnaire—consisting of 15 questions—was ultimately divided into the following major themes:

- ***Vision* for the national program:** overall vision and how it was compatible with individual hospital objectives.
- ***Strategy* for the national program:** Who was engaged with and how the strategy

was being communicated within different organisations.

- ***Implementation* of the national program:** What professional, managerial, and technical skills or capabilities were needed to implement various elements of the national program.
- ***Value delivery* for the national program:** The main risks identified by each hospital and how past IT failure could be avoided, as well as looking at the cost/benefit choices and issues for each organisation.
- ***Risk\analysis* for the national program:** The value being derived from the national program.

The aim was to get the perspectives of a number of different informants using structured interviewing, by building up intensive longitudinal cases which would, nevertheless, be amenable to statistical analysis. In this method, differences of perception of informants become part of the data, not an inconvenience to be explained away in the search for some objective truth.

DATA ANALYSIS

Content analysis was used to surface themes in the interview data that reflected participants' understandings related to systems implementation. The approach suggested by Weber (1990) was used to code the interview data. A set of codes used to classify the data was developed, based on concepts from the research literature and augmented with major additional concepts discovered by the researchers during the coding. We used a content analysis form where each sentence from the interview transcripts was assigned one or more codes. Each data element was coded with an assessment of the level of agreement in code assignments, involving a certain degree of recoding of data sources. As this was the first study that uses content analysis about modelling of system implementation in the NHS, a certain degree of recoding was considered acceptable.

Table 3 contains a list of the most frequently cited attributes and benefits of the system implementation model. The audiotapes were fully transcribed and individual hospital and service provider summaries were produced before conducting a content analysis of each transcript. After a complete review of all summaries, issues describing IS implementation strategies by iterative examination were identified. Certain themes emerged, which were explored using the competing values framework as an interpretive framework where appropriate (see Table 3). The trustworthiness of such analysis has been assessed by triangulation between data sources and exploring any differences in the researcher' interpretations during a couple of follow-up meetings with selected interviewees.

During the period of the field study, there was a continuing, vigorous, informal debate within NHS Information Authority as to the merits of establishing a fault proof IS implementation framework in healthcare, particular for the NHS, during this period of healthcare reform. Benefits in terms of improved quality, greater structure, and more discipline were widely accepted.

THE NHS CASE STUDY

The NHS is the institution responsible for all healthcare and services in the UK with the goal of undertaking this responsibility at no costs to the public, at the point of delivery. The NHS was created in 1948 by a parliamentary art of the UK government of Mr. Howard Wilson, after a national healthcare review by Mr. Black immediately after World War II. While the NHS operating environment has changed radically within the last six decades, only a few periods of strategic importance to the objective of this article will be revisited by the author.

Table 3. Frequently described implementation attributes and benefits

Implementation Attributes		
Item	**Count**	**% of Cat**
Applications work together	40	13
Data sharing	173	57
Common database	127	42
Real-time processing	106	35
Record once, use everywhere	121	40

Implementation Benefits		
Item	**Count**	**% of Cat**
Improved data accuracy/reliability	61	20
Lower costs of support, maintenance	212	70
Greater efficiency and productivity	167	55
New or increased functionality	106	35
Better management, decisions, analysis	136	45

The period from late 1980s to early 1990s brought in the advent of competitive bidding bringing long-term increase costs to the management of the NHS, as well as a feeling of internal market within the NHS. By the mid-1990s, management of IS in the NHS was division based. Divisions were spread across several sites and medical functions were centrally controlled. Computing services and IS development projects were beginning to be contracted to external private businesses, and staff at the NHS were beginning to feel disgruntled and unappreciated. The increasing influence of global communications, Internet, and other new technologies demanded a response from the NHS.

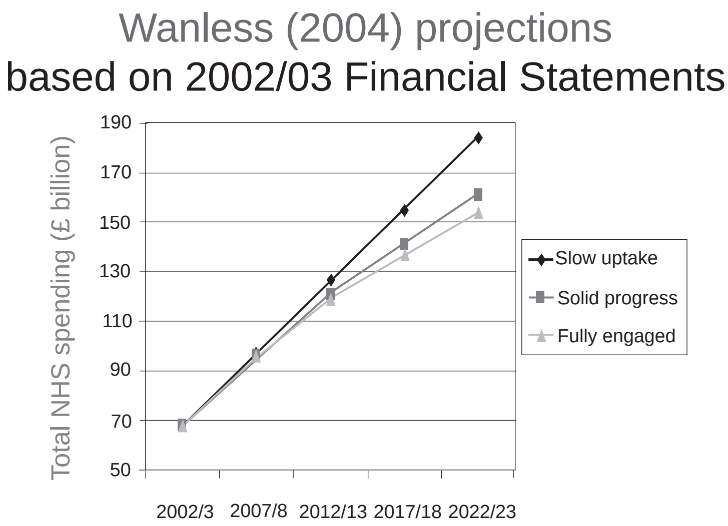
In the late 1990s the government increasingly recognised the opportunity to use IT to improve the delivery of service within the NHS. After a series of reviews of NHS IT service delivery, a more integrated and seamless IT organisation was recommended (DoH, 2000; Wanless, 2002). The NHS Information Authority embarked on the Integrated Care Report Service (ICRS) project to provide, among other services, a nationwide electronic patient database. The result was a document called "Information for Health" that specified the need for the complete automation and integration of various patient information databases in the country (DoH, 2000). The system was commissioned to selected IS service providers at a combined price of $10 billion.

In spite of its vision—to transformation IT—the NHS has a history of introducing large-scale IT development projects that have not been overall successes, with some suggesting failure rates of between 60 to 80% (Brown, 2001). Though the UK public sector spent around $3.5 billion annually on IT, the failure of major IT-enabled projects were characterised by delay, overspending, poor performance, and abandonment (National Audit Office [NAO], 2004, p. 3). At the political level, it is argued that "better IT is needed in the NHS because the demand for high-quality healthcare continues to rise and the care now provided is much more complex, both technically and organizationally" (Connecting for Health, 2004, p. 7). About $250 million is spent on management and administration in the NHS, a controversial figure, as many believe more doctors and nurses should be recruited (see Figure 1).

THEORETICAL ANALYSIS OF THE CASE

The NHS case study illustrates the dynamic nature of a national healthcare IS implementation, set within the context of a rapidly changing organisation. As with all large IT-enabled programs, the success or otherwise of the strategic plan is in its implementation (Doolin, 2004; Hendy, Reeves, Fulop, Huchings, & Masseria, 2005; Herzlinger,

Figure 1. NHS projected expenditures over two decades (Wanless, 2004)



1989). The lessons of IT costs versus medical decision making are well documented in the literature and involve the lack of alignment between the business and IT strategy (Luftman, 2000); a lack of ownership and leadership of the IT project among senior decision makers (Brown, 2001); poor risk assessment skills (Heathfield, Pitty, & Hanka, 1998); over-scoping of the functional and technical specification leading to IT projects becoming over-budget and late (NAO, 2004); poor communication between program/project managers and potential users of the system (Guah & Currie, 2005); inadequate resources to deliver/implement IT systems (Currie & Guah, 2006).

The empirical research found that issues of project risk were at the forefront of the minds of clinicians, general practitioners (GPs), hospital managers, and IT staff. Formal project management methods and tools were perceived as offering only part of the solution to mitigate the considerable risks from introducing the national program. The fragmentation was not just about the diversity of IT systems within the NHS, but also about the political geographical, social, organisational, and financial complexity of delivering healthcare.

The overriding view was for the national program to become an integrated IS across and beyond the NHS. The threats to achieving this were perceived by many clinicians to fall within the control of politicians and IT service providers rather than from NHS staff. Project risk mitigation was a complicated issue, compounded by the political and ideological considerations, such as the public-private partnership funding initiative (PFI), which facilitated the increasing use of private sector firms. While the NHS is often characterised as a top-down bureaucracy (Mohan, 2002), past achievements in IT development and implementation had often been initiated at a decentralised (hospital, departmental, unit) level. Although this was now discouraged by the centrist approach of the national program, staff participating in the research expressed concerns that any failures associated with IT project implementation may be labelled *staff resistance* rather than the shortcomings of external constituents, such as politicians, management consultants, or IT suppliers.

The success or failure of IS is inextricably linked with the dynamics of the organisation within which they exist. Miranda and Saunders (2002) have demonstrated the complex interaction of technical, social, cultural, and political elements that result in a failed IS. Equally, IS success depends on more than technical competence. The cultural and political environment of the NHS is difficult to study as it depends not only on the tangible organisational structure but also on the tacit knowledge and the perceptions of the participants (Guah & Currie, 2005). This is in addition to the cultural and political environment of an organisation that is not static but rather in a state of constant flux and dynamic change.

Institutionalism of IS in the NHS is concerned with processes of cultural persistence and change of healthcare processes. The survival of an organisation depends as much on conforming to societal norms of acceptable practice as to achieving high levels of production efficiency and effectiveness (Covaleski, Dirsmith, & Michelman, 1993). Prior work has shown that an organisation's formal structure, policies, and procedures serve to demonstrate conformity with the institutionalised rules and requirements of external constituents (DiMaggio & Powell, 1983; Meyer & Rowan, 1991). In light of these concerns healthcare in the UK showed that the national program was intended to play a high profile role within the heavily institutionalised environment of hospitals (Scott et al., 2000).

The UK healthcare system is infused with institutional logics emanating from various sectors across the organisational field. Healthcare is politically contentious where societal level logics developed by government are embedded into policies and procedures that cascade down to organisations where they are interpreted by various stakeholders including clinicians, managers, administrators, and patients. How these logics are interpreted varies according to the degree to which they affect changes to the perceived or real material resource environment of the institutional actors. The vision for the national program was infused with the institutional logics more commonly associated with the private sector, as an innovation that would contribute to greater productivity, efficiency, cost control, and customer satisfaction in healthcare delivery. Paradoxically, this externally directed institutional logic served to under represent and simplify the vast complexities and contradictions in how it was perceived, and reacted to, by those affected by government-led IT-enabled changed. Within the NHS, staff were increasingly sceptical about the merits of private sector logics, such as the PFI initiative, as their values, norms, and goals invariably placed financial considerations secondary to choices about patient care.

The proliferation of new entrants into the healthcare organisational field was a consequence of changing government policies over six decades. During this era of professional dominance, healthcare workers, particularly clinicians, enjoyed a level of freedom to define and structure their working practices. This extended to choices about the types of technology adopted and diffused across the NHS. As a new era emerged in the 1970s, which embraced managerialism as a way to enhance efficiency and performance, the European healthcare system was increasingly inundated with various managerial fads and panaceas, like BPR and change management (Herzlinger, 1989). An outcome of such interventions was that isomorphic structures across the NHS were increasingly threatened, as NHS managers were keen to demonstrate "best practice" examples through the adoption of the latest management ideas. Implicit in this logic was that NHS organisations that had not embraced "new ideas" ran the risk of being labeled as "against modernisation" or, at worst, "failing institutions."

Moving from an era of managerialism to one which increasingly advocates the use of "market mechanisms" to regulate and monitor healthcare services, efforts to differentiate NHS organisations still needs to be intensified further. The political rhetoric surrounding the right of patients to "choose" between one provider and the next is not likely to be based upon anything more than a crude assessment of the number of "stars" awarded a primary care trust (PCT), with those holding a low number becoming labeled as offering a less than adequate quality of service to patients. This will further fragment the organisational field of healthcare as the status of individual NHS organisations becomes increasingly differentiated on the basis of current and future evaluation criteria to measure performance. The topic of performance measurement was highly contentious, however, as respondents offering both a clinical and technical perspective believed that the emphasis upon target-setting was carefully designed to absolve politicians from responsibility by accentuating the role and accountability of professional groups.

The concepts of the organisational field, institutional logics, and governance systems are central to our understanding of how the healthcare system adopts and adapts to changes in the material-resource environment and the beliefs, rules, and ideas that comprise the value system. How and why these eras have emerged underpins our investigation into the UK NHS and facilitates our understanding of the nature and scope of large-scale change programs, and the extent to which they signal an institutional change within healthcare.

The institutional logics and governance systems must therefore be understood to explain why this type of change has occurred. In our investigation, we are concerned to apply our theoretical framework to help us understand how a large–scale, nation-wide technical change program is being adopted and diffused throughout the healthcare system and, more specifically, the changes in the organisational field, institutional logics, and governance systems that serve to encourage or inhibit such change.

TOWARDS BETTER VALUE IN HEALTHCARE DECISION MAKING

The national program is intended to help the situation healthcare managers presently face—with almost no say over the crucial factors which most managers anywhere else in the Western world and in other industries need to have in order to be effective (Wanless, 2004). The situation was brought about by a mixture of a quasi-medieval system and a control approach to managing patient data. Unlike other industries (i.e., banking and airline industries) where decision making leads to a good outcome, requires adaptation, and matching of the process to the individual customer, the healthcare industry may not necessarily provide a gold standard for a process that guarantees good outcomes for patients (Grabowski & Lee, 1993; Martin, 2003).

The previous is partly a result of humans being biological creatures and biological systems are inherently variable. As individuals we all have our own copies of genetic material—these materials mutate and evolve randomly. Because of the variability, the number of formulas and data points required to document each instantiation of a biological system constantly increases by several folds. Consequently, the number of conditions that need to be handled by uniform data standards is much greater than that required by standards for simpler physical, production or materials handling systems, however large or geographically widespread they may be. It takes into consideration that the patient's state of health is the result of the complex interaction between his/her unique genetics makeup, brain capacity, environment, and habits. Thus explaining why some individuals may be able to carry on fairly normal activities despite severe loss of 80% of their pulmonary capacity, others may be disabled by minor arthritis. Careful reproduction of a healthcare process that results in a good outcome for the former may not help the later.

Such scenarios dictate that the delivery of healthcare is not only unique in variety but also in the range of services and products. This explains why healthcare services are not typically chosen by the patient but by a more professional and knowledgeable representative (although these medics are increasingly being influenced by the pharmaceutical companies and/or government (Stevens et al., 1993). This shows just how the healthcare industry is different from others in the Western economy because the market in most industries is driven by the customer.

Health industry requirements are also exceptionally demanding in a number of areas. Most notable are the implications of violations of personal privacy while involving all those who need to know; dual responsibility for personal and public health; the complexity and expansion of the knowledge base and terminology; the high risk to the providers' livelihood combined with pressures to make critical decisions continuously and rapidly; and poorly defined outcomes. All of this is in the context of a "guild system" of responsibility, accountability, and power (Markus, 1983; Vogel, 2003). The healthcare industry also has to support personal and moral values, which in themselves are very complex. The judgments taken about personal attitudes to risk and potential benefit on interventions are all driven by our unique physical and mental makeup and local context—though our values are constantly changing over time.

The implementation of the national program demonstrates how healthcare services are perhaps the most complex large-scale business of UK's economy. More variability and uncertainty at the point of service, as to causality, processes, and to the outcome of that investment exists in healthcare delivery than in any other public sector. With such "variability and uncertainty" in the healthcare business, it is not surprising that identifying and measuring, let alone valuing, a financial return on investment (ROI) from computers in healthcare presents special challenges.

Value is a much broader concept than "benefits" as it implies the additional gain from one investment as opposed to another (Lucas, 1993; Sambamurthy & Zmud, 1994). The national program, like other IT projects, can generate value in many ways. As well as creating quality and process improvements, data from the system has utility that is much more subjective than that of any other resource (Guah & Currie, 2004). While data may be viewed as a commodity, asset, or resource, information is derived from the qualitative use of data and involves value judgments.

A number of the difficulties in "measuring" the national program's potential value the patients are that (Currie & Guah, 2006; Hendy et al., 2005):

- many infrastructure investments cannot be cost justified on an ROI basis;
- some of its sob-systems are being implemented to change difficult-to-measure NHS staff actions;
- many parts may be strategic systems, thus, eluding measurement;
- much of the new investment does not take into account the prior costs;
- efficiency (doing things right) is easier to measure than effectiveness (doing the right things); and
- since effectiveness (doing the right things) and innovation (doing new things) can not be readily quantified in terms of traditional outputs, improvements are not usually reflected in economic efficiency statistics.

While the effectiveness of a healthcare delivery process can be defined as "the extent to which a desirable outcome is achieved in a timely manner," the efficiency of healthcare delivery process could also be defined as "the extent to which healthcare delivery process is completed with the minimal consumption of resources." Consider the rather complex process of diagnosing and treating certain categories of patients in the NHS, most healthcare delivery processes can be effective (the patient achieves a full recovery and returns to his/her normal activities within 3 months) but relatively inefficient (say the patient had a 15-day inpatient stay, extensive ambulatory services, and consumed $300,000 in healthcare resources). The healthcare delivery process also can be ineffective (the patient has minimal subsequent capacity and never returns to his/her normal activities) but efficient (if the patient had a 3-day inpatient stay, a short post discharge ambulatory regimen, and consumed only $5,000 in healthcare resources). This demonstrates the need for healthcare managers to further strive for high effectiveness in combination with high efficiency and at the lowest possible costs.

Most components of the national program are intended for use in the NHS primarily to capture and manipulate data for improved—both clinical and administrative—decision making. Majeed (2003) suggests that part of the value anticipated from the national program derives from improvements in the effectiveness of the clinical decision-making process. It should enable physicians and nurses to make better, quicker decisions through mechanisms such as online access to evidence-based results for designated disease conditions, assistance in placing orders (detecting a drug-drug interaction before the order for a medication is actually placed), and receiving an alert electronically after a significantly abnormal test result (Keeler & Newman, 2001). Increasing the effectiveness of the clinical decision-making process should also lead to higher efficiency of that process, which should consequently lead to fewer errors being made and fewer resources being consumed. Among many problems highlighted by Majeed (2003), was the fact that necessary information was often difficult to obtain or simply unavailable, and what was available did not always support the clinical decision-making process.

Improved access to records means dramatically improved efficiencies in a variety of areas (Lucas, 1993). With the old paper-based system, the average turn-around time before laboratory and radiology reports reached the physician was several days. With the national program, diagnostic test results are available within seconds of being verified and has the potential of being brought to the physician's attention through their e-mails on the desktop or text message on a mobile (for telemedicine). In addition, the redundant orders that were often triggered by result delays have been virtually eliminated with the faster turnaround time, as well as with duplicate verifications for pharmaceutical orders (Majeed, 2003).

The national program will vastly improve communication between clinicians, particularly between nurses and physicians, and the emergency room and primary care physicians. It will give nurses instant access so that they can communicate with the doctors in the right way; the nurses are less frustrated because they do not have to go through the paper trail like they did before. These intangibles we know can be difficult to value but they are real enough to the participants.

CONCLUSION

The UK NHS is a highly institutionalised and complex system, which exists and operates both as a material-resource environment and a set of beliefs, rules, and ideas. Although these two environmental facets are conceptually distinct, material-resource environments are influenced by the institutional context. The selection of resources and how they are combined and deployed is determined by institutional beliefs and rule systems (Meyer & Rowan, 1991). By looking at the national program, this article has shown that implementing such a large IT project in the healthcare environment involves a supplier base that is seriously diverse, stretching from locally based specialists in particular applications and/or industry sectors to suppliers that are capable of applying a combination of sophisticated management techniques and technology investment to achieve new levels of process performance.

The primary contribution of this article has been to provide a theoretical basis drawing from institutional theory, which was used to analyse the NHS implementation of the national program. The theorisation goes beyond the relatively simplistic types of studies that dominate the IS literature today. Much to the contrary, it has been shown that an implementation strategy can accommodate elements such as the links between culture, contradiction and conflict, an analysis of detailed work patterns, and the dynamic and emergent nature of political involvement at national level.

The theory has been illustrated using limited empirical examples only, with a focus on the NHS systems, but it could be used to analyse any case study involving healthcare systems from any parts of the developed world. Viewed from a more critical perspective, however, any theory illuminates some elements of particular case situations and is relatively silent on others. The NHS has grown within an environmental niche that arose out of a complex interaction between the national healthcare environment, business environment, the organisational environment, and the people within the NHS. Changes within the organisation subsequently rendered the environment hostile to the national program, which was affected by its changing links with organisational structure and people, the changing responses of people within the NHS to the environment around them, and the changing individual and collective mindsets and understanding of those people. While a detailed discussion of ways in which this can be achieved is beyond the scope of this article, some broad approaches have been mentioned.

In the current environment of increasing demands for better quality of healthcare from patient and seemingly reduced amount of funding from national governments, the need for suitable institutional theory is increasingly common and the IS field must increase its understanding of the problematic issues involved and approaches to resolving them. It is hoped that this article makes a modest contribution to these goals.

REFERENCES

Brown, T. (2001). Modernization or failure? IT development projects in the UK public sector. *Financial Accountability & Management, 17*(4), 363-381.

Connecting for Health. (2004). *Business plan*. Retrieved May 2005, from http://www.connectingforhealth.nhs.uk

Covaleski, M. A., Dirsmith, M. W., & Michelman, J. (1993). An institutional theory perspective on the DRG framework, case mix accounting systems and healthcare organizations. *Accounting, Organizations and Society, 18*(1), 65-80.

Currie, W. L., & Guah, M. W. (2006). IT-enabled healthcare delivery: The UK National Health Service. *Information Systems Management, 23*(2), 7-22.

Department of Health (DoH) (UK). (2000). *NHS plan: An information strategy for the modern NHS*. London: Author.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review, 48*, 147-160.

DiMaggio, P. J., & Powell, W. W. (Eds.). (1994). *The new institutionalism in organizational analysis.* Chicago: University of Chicago Press.

Doolin, B. (2004). Power and resistance in the implementation of a medical management information systems. *Information Systems Journal, 14*, 343-362.

Grabowski, M., & Lee, S. (1993). Linking information systems application portfolios and organizational strategy. In R. D. Banker, R. J. Kauffman, & M. A. Mahmood (Eds.), *Strategic information technology management: Perspectives on organizational growth and competitive advantage* (pp. 33-54). Hershey, PA: IGI Global.

Guah, M. W., & Currie, W. L. (2004). Application service provision: A technology and working tool for healthcare organization in the knowledge age. *International Journal of Healthcare Technology and Management, 6*(1), 84-98.

Guah, M. W., & Currie, W. L. (2005). *Internet strategy: The road to Web services*. Hershey, PA: IGI Global.

Heathfield, H., Pitty, D., & Hanka, R. (1998). Evaluating information technology in healthcare barriers and challenges. *British Medical Journal, 316,* 1959-1961.

Hendy, J., Reeves, B. C., Fulop, N., Huchings, A., & Masseria, C. (2005). Challenges to implementing the national program for information technology: A qualitative study. *British Medical Journal, 420,* 1-6.

Herzlinger, R. E. (1989). The failed revolution in healthcare—The role of management. *Harvard Business Review,* 95-103.

Institute of Medicine. (2002). *Crossing the quality chasm: A new health system for the 21st century.* Washington, DC: National Academy Press.

Keeler, J., & Newman, J. (2001). Paperless success: The value of e-medical records. *HIMSS Proceedings, 2*(45).

Lowndes, V. (1996). Varieties of new institutionalism—A critical appraisal. *Public Administration, 74,* 181-197.

Lucas, H. C. (1993). The business value of information technology: A historical perspective and thoughts for future research. In R. D. Banker, R. J. Kauffman, & M. A. Mahmood (Eds.), *Strategic information technology management: Perspectives on organizational growth and competitive advantage* (pp. 359-3744). Hershey, PA: IGI Global.

Luftman, J. (2000). Assessment business IT alignment maturity, *Communications of the AIS, 4*(14), 1-51.

Majeed, A. (2003). Ten ways to improve information technology in the NHS. *British Medical Journal, 326,* 202-206.

March, J. G., & Olsen, J. P. (1989). *Rediscovering institutions: The organizational basis of politics.* New York: Free Press.

Markus, M. L. (1983). Power, politics and MIS implementation. *Communications of the ACM, 26*(6), 430-445.

Martin, J. B. (2003). Effectiveness, efficiency, and the value of IT. *Journal of Healthcare Information Management, 17*(2).

Meyer, J. W., & Rowan, B. (1991). Institutionalized organizations: Formal structure as myth and ceremony. In W. W. Powell & P. J. DiMaggio (Eds.), *The new institutionalism in organizational analysis.* Chicago: University of Chicago Press.

Miranda, S. M., & Saunders, C. S. (2002). The social construction of meaning: An alternative perspective on information sharing. *Information Systems Research.*

Mohan, J. (2002). *Planning, markets and hospitals* London: Routledge.

National Audit Office (NAO). (2004). *Improving IT procurement.* Report by the Comptroller and Auditor General, HC 877 Session, 2003. London: The Stationary Office.

North, D. (1990). *Institutions, institutional change and economic performance.* Cambridge, UK: Cambridge University Press.

Sambamurthy, V., & Zmud, R. W. (1994). *IT management competency assessment: A tool for creating business value through IT.* Working paper. Financial Executives Research Foundation.

Scott, W. R., Ruef, M., Mendel, P. J., & Caronna, C. A. (2000). *Institutional change and healthcare organizations: From professional dominance to managed care.* Chicago: University of Chicago Press.

Stevens, P., Schade, A., Chalk, B., & Slevin, O. (1993). *Understanding research: A scientific approach for health care professionals.* Edinburgh, UK: Campion Press.

Tolbert, P. S., & Zucker, L. G. (1994). Institutional analysis of organizations: Legitimate but not institutionalized. Institute for Social Science Research working paper. University of California, Los Angeles.

Vogel, L. (2003). Finding value from IT investments: Exploring the elusive ROI in healthcare. *Journal of Healthcare Information Management, 17*(4).

Wanless, D. (2002). *Securing our future health: Taking a long-term view.* London.

Wanless, D. (2004, February). *Securing good health for the whole population*. London.

Weber, R. P. (1990). *Basic content analysis* (2nd ed.). Newbury Park, CA: Sage.

This work was previously published in Journal of Information Technology Research, Vol. 1, Issue 1, edited by M. Khosrow-Pour, pp. 14-26, copyright 2008 by IGI Publishing.

Chapter 6.5
Information Assurance in E-Healthcare

Sherrie D. Cannoy
The University of North Carolina at Greensboro, USA

A. F. Salam
The University of North Carolina at Greensboro, USA

INTRODUCTION

There is growing concern that the healthcare industry has not adopted IT systems as widely and effectively as other industries. Healthcare technological advances generally emerge from the clinical and medical areas rather than clerical and administrative. The healthcare industry is perceived to be 10 to 15 years behind other industries in its use of information technology (Raghupathi & Tan, 1997). Incorporating new technology into the healthcare organization's processes is risky because of the potential for patient information being disclosed. The purpose of this study is to investigate the information assurance factors involved with security regulations and electronic medical record initiatives—a first necessary step in making the healthcare industry more efficient. Noncompliance of a healthcare organization's employees with security and privacy policies (i.e., information assurance) can result in legal and financial difficulties, as well as irreparable damage to an organization's reputation. To implement electronic medical initiatives, it is vital that an organization has compliance with security and privacy policies.

E-health technology is a relatively current phenomenon. There are two types of distance-related healthcare that are technology driven. Telehealth is known for involving telemedicine—medicine practiced over a distance, with the impetus of control being in the physician's hands (Maheu, 2000). E-health involves the patient or physician actively searching for information or a service, usually via the Internet (Maheu). Electronic medical records fall into the e-health category because the physician, healthcare partners, and patient would be able to access the information through an Internet connection.

Security and information assurance are critical factors in implementing e-health technologies. There is a lack of a well-developed theoretical framework in which to understand information assurance factors in e-healthcare. The theory of reasoned action (TRA) and technology acceptance model (TAM) enable a conceptual model of information assurance and compliance to be formed in the context of healthcare security and privacy policy. The relationship between behavior and intentions, attitudes, beliefs, and external factors has been supported in previous research and will provide a framework for ensuring compliance to security and privacy policies in healthcare organizations so that HIPAA (Health Insurance Portability and Accountability Act) regulations are enforced and electronic medical records (EMRs) can be securely implemented.

Traditionally, records in the healthcare industry have been paper based, enabling strict accessibility to records. This allowed for confidentiality of information to be practically ensured. The uniqueness of healthcare records and the sensitive information they contain is specific to the industry. Over the many years that medical records have been kept, those involved in the field have undertaken a self-imposed rule of stringently protecting the patient information while providing quality care.

The patient's expectation for confidentiality of personally identifiable medical records is also critical. According to Rindfleisch (1997, pp. 95-96), in his study of healthcare IT privacy, the threats to patient information confidentiality are inside the patient-care institution; from within secondary user settings which may exploit data; or from outsider intrusion into medical information. Rindfleisch (1997) examined specific disclosures which could release sensitive information such as emotional problems, fertility and abortions, sexually transmitted diseases, substance abuse, genetic predispositions to disease—all of which could cause embarrassment and could affect insurability, child custody cases, and employment.

The process of healthcare treatment includes not only the patient and physician but also nurses, office staff who send out bills and insurance claims, the insurance company, billing clearinghouses, pharmacies, and any other companies to which these processes can be outsourced. There is an estimate that states as many as 400 people may have access to your personal medical information throughout the typical care process (Mercuri, 2004). The government is also a partner in national health concerns, and also maintains databases containing information on contagious diseases, cancer registries, organ donations, and other healthcare information of national interest. (See http://www.fedstats.gov/programs/health.html for a listing of the databases.)

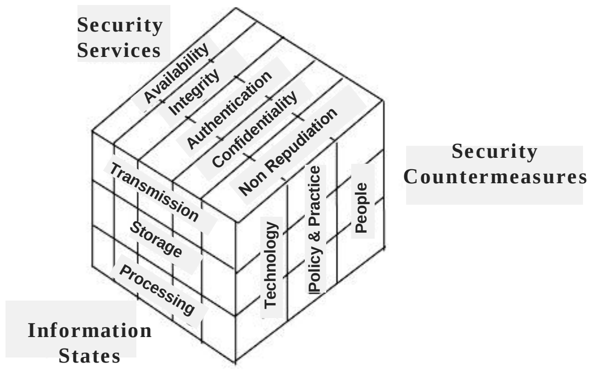
With the advent of government mandates such as EMRs and HIPAA regulations, the increased accessibility of sensitive records requires intense effort to create policies that limit access for those who are authorized. Although there is an area of information economics which views information as an asset that can be numerically valued for its benefit, the same perspective has not been adopted in healthcare. Especially in the United States, clinical information and patient care are considered proprietary (Hagland, 2004). There is no specific associated cost with one's medical information—what damage is done when one's medical information has been utilized improperly? Even though damages are ill-defined, there are regulations and standards for emerging technology in healthcare. The two most current important security and privacy issues involve HIPAA regulation and the government mandate for EMRs.

BACKGROUND

The HIPAA Regulation . . .

HIPAA was enacted in 1996 and covers insurance reform for ensuring preexisting coverage when changing jobs as well as the standardization of

Figure 1. Information assurance model (From Maconachy, Schou, Ragsdale, & Welch, 2001)



electronic transmissions. It consists of two components, the Security Rule and the Privacy Rule. If the rules are not enacted, there are severe financial penalties enforced by the government (Mercuri, 2004). Also, an organization risks having internal employees disclose information that would be of a confidential nature to patients, which could result in legal consequences.

The Privacy Rule (Markus, 2004) focuses on the use and disclosure of medical information, specifically that which is personally identifiable, also known as protected health information (PHI) in the industry. The goal of the Privacy Rule is to ensure protection of PHI across transmissions to health partners (insurance companies, billing clearinghouses, etc.). This requires the patient to fill out the Notice of Privacy Practices Patient Acknowledgment form, which suggests that the patient has read the HIPAA privacy information and allows the patient to determine the people to which one's PHI can be disclosed.

The Security Rule requires that PHI be protected specifically in electronic storage and transmissions. Implications for HIPAA compliance have been intense. Developing standards and security encryptions for existing software, as well as ensuring that third-party partners are compliant, has been time-consuming and costly. However, Privacy and Security Rule compliance will be critical for successful implementation of electronic medical record infrastructure.

Electronic Medical Records

An electronic medical record (EMR) contains a patient's medical history with a physician. The capture of one's medical information can be made available to authorized users such as other physicians, pharmacists, insurance companies, and the government. Due to the inherent virtualization of the record, the physician's office or hospital will not have physical control as they have in the past. Therefore, security measures, mainly technical components, are critical to EMR implementation. EMR records will also fall under the HIPAA Privacy and Security Rules. Since the healthcare industry has been reluctant to implement EMR plans for cost, security, or other reasons, the government has taken an active role to encourage development of EMRs through financial incentives. Healthcare and IT organizations are also collaborating to develop standards (Mercuri, 2004).

Measures Needed for Better Management

Handling of sensitive information will be vital to understanding the compliance for HIPAA regulation, as well as for the implementation of EMRs. With information now being stored and transmitted electronically, a new paradigm exists for power over the information. How organizations measure the success of HIPAA compliance will reflect on how sensitive information is handled. However, it is uncertain how the healthcare industry monitors this. The old adage of "what is not measured is not managed" comes to mind and one wonders if compliance will be monitored after training is administered and policies are implemented. The purpose of this study is to develop a preliminary framework of issues that determine compliance to information assurance and security policies.

Information Assurance

Figure 1 (from Maconachy, Schou, Ragsdale, & Welch, 2001) in which the aspects of information assurance are depicted. They discuss information assurance as an expansion of the "coverage, responsibilities, and accountability of security professionals" which includes "proactive offensive activities" (p. 307).

Aspects of security services, information, and security countermeasures fall under the information assurance (IA) umbrella. The focus of IA is integrating the relationships between these, since a weakness in one would result in a weakness in the entire system. The information can be either currently in storage, in processing, or in transmission mode. The Security Services are carried not only by technological details such as availability of the system, integrity of the data, authentication of users, and confidentiality, but by operations and procedures, and, most importantly, people. People "are the heart and soul of secure systems ... and require awareness, literacy, training, and education in sound security practices for the system to be secured" (Maconachy et al., p. 308). Whether or not a person follows a policy and uses the IA technology will determine how secure the system ultimately will be.

In terms of healthcare, information assurance provides a well-developed model in which to integrate the technological and behavioral issues in e-Healthcare security and privacy context. E-healthcare will be successful only if information assurance is carefully planned and monitored for compliance. In the premise offered for this research, information assurance is inherent in security and privacy policy compliance, and provides the structure for integrating the variety of research areas involved. Therefore, a theoretically-sound framework for incorporating TRA and TAM in the context of healthcare information assurance is proposed to bridge the gap between policy creation and policy compliance. This area has not been researched previously and will provide an important contribution to e-healthcare implementation in the industry, especially in bridging the gap between the technical and behavioral aspects of e-healthcare information assurance.

Government resources from Health and Human Services and associated National Institute for Standards and Technology (NIST) are helpful in framing the variables into a model for information assurance and compliance. The Department of Health and Human Services developed *Delivering Consumer-Centric and Info-Rich Healthcare: Framework for Strategic Action* (Thompson & Brailer, 2004), which discusses the readiness for change in healthcare, as well as a framework for action and leadership. NIST offers policy and security frameworks as well as healthcare tools and other assistance for both the clinical and administrative aspects (see http://www.nist.gov/public_affairs/healthcare.htm).

Healthcare Culture, Security, Compliance, and Information Assurance

It is often stated that an organization is only as strong as the weakest link. "When you start bringing in the users, your system is going to be as trustworthy as the least trustworthy person who touches it" (Baldwin, 2000). The opportunity with HIPAA and EMRs is to realize that privacy and security of PHI must be managed in a technologically driven environment. This supports our premise that, although technological tools can aid in monitoring security and privacy compliance, they cannot provide the comprehensive measure—comprehensive adoption includes people issues. Trish Markus (2004) questioned the establishment of a "culture of compliance," which advocates the management involvement and commitment through communication and training procedures. Mercuri (2004) quoted a chief information officer as stating that HIPAA "compliance is not sold in a bottle," therefore, "providing employees with policies and procedures for their job classification and requiring them to read and sign off on them" is not adequate (p. 27). Culture includes a shared vision and a positive link to company strategy. Indeed, organizational culture can determine if compliant attitudes, intentions, and behaviors will become second nature (Gue, 2002).

Theory-Based Research

Academic research in the healthcare area has been specific, usually technology-related versus theory building and testing. TRA and TAM are utilized for the purpose of this research because the factors including a person's attitudes, beliefs, and intention to adopt a technology such as EMRs and to examine one's intentions to comply with information assurance policy associated with those EMRs. Hu, Chau, and Tulu et al. (1999, 2002, 2003) researched TAM in the context of healthcare technology adoption, with mixed results. In their 1999 telemedicine study Hu et al. found that TAM is adequate, with exception to TAM's explanation of attitude and intention. Perceived usefulness was significant on intention to use telemedicine. For the physician to perceive telemedicine as being useful, Hu et al. (1999) suggested that proper user training is essential. Attitude also significantly influenced physician behavioral intention.

In similar studies, TRA was utilized and was extended in information technology usage. Amoako-Gyampah and Salam (2004) measured constructs for ERP (enterprise resource planning) project communication, training, belief in the project benefits, attitude towards ERP system, perceived usefulness, ease of use, and behavioral intention to predict usage of ERP systems. TAM was extended to external variables of project communication, beliefs in ERP system benefits, and training. Findings were significant for project communication and training affect on beliefs in the system benefits. Also, shared beliefs strongly affected both PU (perceived usefulness) and PEU (perceived ease of use) in the ERP setting.

An example of TRA utilization in another study by Salam, Iyer, Palvia, and Singh (2005) examined associations between the formation of trust, technology usage, and development of a relationship with a vendor website. It suggests that managers develop factors that affect the trustworthiness of their companies. This correlates to managing the factors that affect compliance with security and privacy policies.

Figure 2. Diagram of the theory of reasoned action

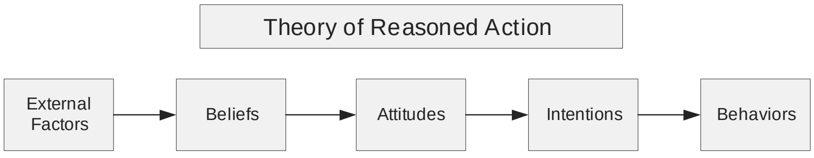
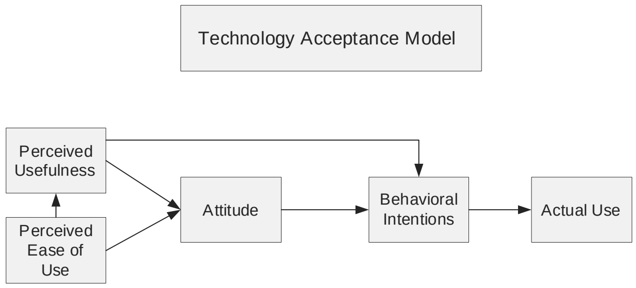


Figure 3. Diagram of the technology acceptance model



Capturing Information Assurance Issues in Healthcare IT Policy Implementation

The widely accepted theory of TRA has been utilized often to support the causal relationships between external factors, beliefs, attitudes, intentions, and behavior. TAM is a TRA spin-off and has supported the intention to use, perceived usefulness, and behavior when adopting new technology. This same theoretical basis can be used to examine the acceptance and compliance behavior of users who are to adopt new organizational policy. In the context of information assurance, the TRA model will capture the compliance of healthcare security and privacy policies (see Figure 2).

Our premise is that through one's external factors, beliefs, attitudes, and intentions, one will decide to either comply or not comply with security and privacy policies. We will examine the context of compliance with security and privacy policies within the TRA, because the user of the policy will, in essence, determine the success of such policies and compliance with HIPAA. The user can be defined as the physician, clerical staff member, IT staff, administrative staff, nursing staff, and so forth.

COMPONENTS OF THE FRAMEWORK

Organizational Culture

The main components of organizational culture which relate to information assurance compliance are management commitment and its enforcement of IA Policies. The extent to which management seems committed and is willing to enforce IA policies will affect how strongly one believes in management commitment and its enforcement of those policies.

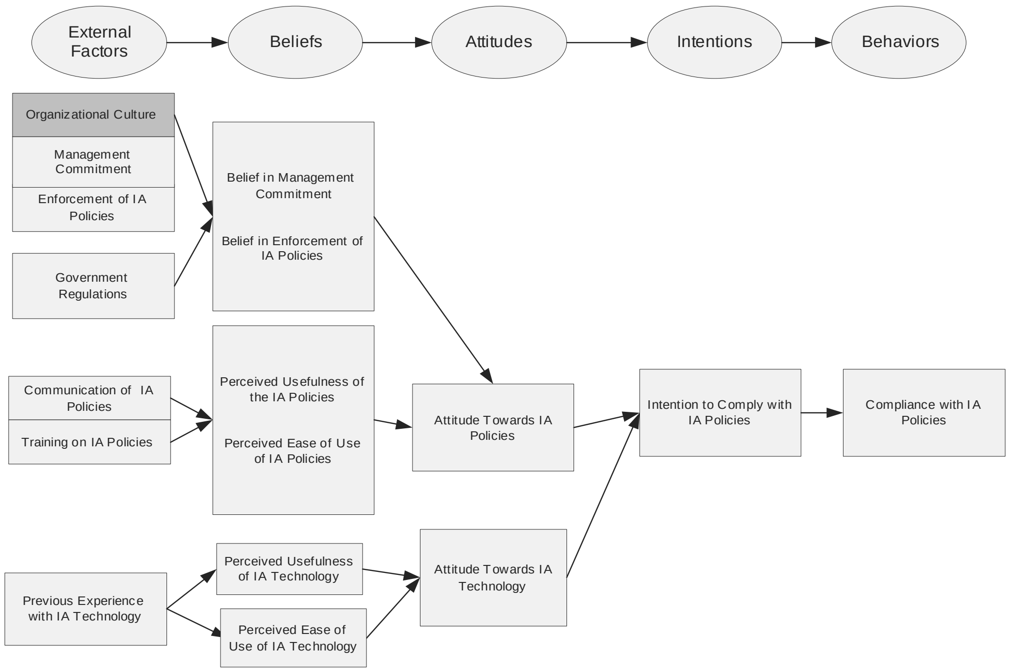
Government Regulations

Government regulations such as HIPAA and mandates such as EMRs will affect how strongly the management of healthcare organizations is committed to communicating and enforcing information assurance policies.

Training and Communication

Amoako-Gyampah and Salam (2004) found in their research on ERP implementation that Training and Communication influence the beliefs on attitude and its mediation effects on the intention to use an ERP system. Accordingly, training on IA Policies and the communication of those policies should also have an influence on the

Figure 4. A conceptual framework for applying the theory of reasoned action and the technology acceptance model in e-healthcare information assurance



perceived ease of use and perceived usefulness of such policies. If employees feel that training is helpful in information assurance and compliance, employers may be urged to offer further training, even in the maintenance phase, so that information assurance can be increased as much as possible. If the policies are communicated often, and in various ways, especially in how they are important and relate ensuring security and privacy of patient information, employees should be more apt to find it beneficial to use and comply with the policies.

Previous Experience with IA Technology

The TAM model is appropriate and supported in demonstrating the influence between a person's prior experience with a technology and how useful and easy that technology will be to use. Much of the research has shown that this perception of usefulness and ease of use enables one to develop attitudes towards that technology in order to determine whether they will use the technology (Davis, 1989; Venkatesh & Davis, 2000). In this case, the attitude towards technology will affect one's behavior and choice to comply with the policies. If the implementation of information assurance and security policies involves technological barriers that decrease functionality of the technology or are too difficult to manage, the user is less likely to use the technology, therefore, less likely to comply with the policy.

This framework captures how factors such as organizational culture, government regulations, training, policy, and previous experience lead to belief, attitude, and intention formation regarding IA policy compliance behavior. Belief systems such as management commitment, enforcement of policy, the ease of use and perceived usefulness of IA policy and technology, shape one's attitude towards IA technology and policy. This attitude development then affects one's intention to comply with IA policies and, therefore, actual compliance behavior.

FUTURE TRENDS IN E-HEALTHCARE

The environment for healthcare is changing rapidly, and the processes involved will transform how patient care is conducted. Along with the implementation of EMRs, the ultimate goal is the EHR (electronic health record), which will include the consolidation of a patient's lifetime health records from each healthcare provider. This EHR is a virtual roadmap that guides the physician in the journey to medical decision-making. Having this information at hand, even available via handheld devices, makes critical information available to physicians to share data while geographically dispersed. This is the dawn of telemedicine, which is already gaining popularity in New Zealand and Australia. Smart card technology is emerging as a new technology which can be applied to enable patients to carry their own virtual medical charts in their pockets (Raghupathi & Tan, 2002).

There are additional legal and financial issues to be handled in the future. The legal aspects of implementing telemedicine and EMRs need clarification. Who owns the EMR? Who will be responsible for attempts at unauthorized access? Will the patient be allowed to view and edit their EMRs—for example, add their living will or organ donor preferences (Pyper, Amery, Watson, Thomas, & Crook, 2001)? If a vendor is utilized to store healthcare data for an organization, what happens to the data if the company goes out of business (Songini & Dash, 2000)? Will insurance companies consider telemedicine consults for payment?

These legal, financial, and technical issues will need to be resolved so that the transformation of healthcare processes are successful and patient care is administered in the most appropriate manner possible, regardless of how, where, and what information is utilized to make the most informed medical decisions.

CONCLUSION

The contribution of this research is to propose a theoretically sound framework in which behavior and technology are analyzed in compliance and information assurance in e-healthcare. This will enable future research in moving the field forward in healthcare information assurance research and practice. The next step in this research is to examine how healthcare organizations can measure the success of security and privacy policy compliance to achieve an information assurance threshold. The importance of understanding compliance within this context will be critical to determining the success of security and privacy policies.

REFERENCES

Amoako-Gyampah, K., & Salam, A. F. (2004). An extension of the technology acceptance model in an ERP implementation environment. *Information and Management, 41,* 731-745.

Baldwin, F. (2000, May) In Web we trust? *Healthcare Informatics Online*, 56-64. Retrieved February 2005, from http://www.healthcare-informatics.com/issues/200/05_00/cover.htm

Davis, F. (1989, September). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340.

Department of Health and Human Services, Office of the Secretary. (2000, December 28). Security and Electronic Signature Standards: Proposed Rule (45 CFR Part 142). *Federal Register, 63*(155), 43242-43280.

Department of Health and Human Services, Office of the Secretary. (2000, December 28). Standards for privacy of identifiable health information: Final rule (45 CFR Parts 160 and 164). *Federal Register, Part II, 65*(250), 82462-82829.

Fariborz, F., Navathe, S., Sharp, G., & Enslow, P. (2003, September) Managing vulnerabilities of information systems to security incidents. *Proceedings of the Fifth International Conference on Electronic Commerce* (pp. 348-354).

Gue, D. G. (2002, March) *The step child of HIPAA compliance: Culture change*. Retrieved from www.hipaadvisory. com

Hagland, M. (2004, May) Reality EMRs. *Healthcare Informatics Online*, 34-40. Retrieved February 2005, from http://www.healthcare-informatics.com/issues/2004/05_04/cover.htm

Hu, P.J-.H., Chau, P., & Liu, S. O. (2002) Adoption of telemedicine technology by healthcare organizations: An exploratory study. *Journal of Organizational Computing and Electronic Commerce, 12*(3), 197-221.

Hu, P.J-.H., Chau, P., Sheng, O., & Tam, K. Y. (1999) Examining the technology acceptance model using physician acceptance of telemedicine technology. *Journal of Management Information Systems, 16*(2), 91-112

Maconachy, W. V., Schou, C., Ragsdale, D., & Welch, D. (2001, June). A model for information assurance: An integrated approach. *Proceedings of the 2001 IEEE Workshop on Information Assurance and Security, 11*(15), 306-310.

Maheu, M. (2000, February) *Telehealth: delivering behavorial telehealth via the Internet.* Retrieved April 2005, from http://telehealth.net/articles/deliver.html

Markus, T. (2004, February). Post-implementation assessment. *Healthcare Informatics Online*, 88-90. Retrieved February 2005, from http://www.healthcare-informatics.com/issues/2004/02_04/hipaa.htm

Marietti, C. (1998, May). Will the real CPR/EMR/EHR please stand up? *Healthcare Informatics Online*, 76-80. Retrieved February 2005, from www.healthcare-informatics.com/issues/1998/05_98/cover.htm

Mercuri, R. (2004, July) The HIPAA-potamus in healthcare data security. *Communications of the ACM*, *47*(7), 25-28.

Patel, K., & Rushefsky, M. (2002). *Healthcare policy in an age of new technologies.* Armonk, NY: ME Sharpe.

Pyper, C., Amery, J., Watson, M., Thomas, B., & Crook, C. (2001, October 13) Trust me! I'm the patient. *Oxford Health Authority: Electronic record development and implementation programme* (Version Final 1.0). Retrieved February 2005, from http://www.nhsia.nhs.uk/erdip/pages/demnstrator/bury/bury_(16).pdf

Raghupath, W. (1997, August). Health care information systems. *Communications of the ACM, 40*(8), 81-82.

Raghupathi, W., & Tan, J. (2002, December). Strategic IT applications in health care. *Communications of the ACM, 45*(12), 56-61.

Rindfleisch, T. (1997, August). Privacy, information technology, and health care. *Communications of the ACM, 40*(8), 93-100.

Salam, A. F., Iyer, Palvia, P., & Singh, R. (2005, February). Trust, technology usage and exchange relationship in e-commerce: A comprehensive framework. *Communications of the ACM*, *48*(2), 72-77.

Songini, M., & Dash, J. (2000, December 20). *Hospital hack points to need for standards.* Retrieved February 2005, from http://archives.cnn.com/2000/TECH/computing/12/20/health.care.security.idg/index.html

Thompson, T., & Brailer, D. (2004, July 21). The decade of health information technology—delivering consumer-centric and info-rich healthcare: Framework for strategic action. *Department of Health and Human Services, Office for the National Coordinator for Health Information Technology, and U.S. Federal Governmentment*, Washington, DC. Retrieved February 2005, from http://www.hhs.gov/healthit/documents/hitframework.pdf

Tulu, B., Hilton, B., & Horan, T. (2003). Physicians' acceptance of Web-based medical assessment systems: Findings from a national survey. *Ninth Americas Conference on Information Systems* (pp. 939-946).

Turn, R. (1999, November). Security and privacy requirements in computing. *Proceedings of 1986 ACM Fall Joint Computer Conference.*

Venkatesh, V., & Davis, F. (2000, February). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, *46*(2), 186-204.

KEY TERMS

Confidentiality: Not releasing or disclosing information to unauthorized individuals, entities, or processes (Fariborz, Navathe, Sharp, & Enslow, 2003).

EMR: Electronic medical record; also known as EHR (electronic health record), CMR (computerized medical record) or CPR (computer-based patient record). There are subtle differences between these; for example, EHR is often used specifically for ambulatory care records (Marietti, 1998).

HIPAA: Health Insurance Portability and Accountability Act (1996) requires the secretary of health and human services to publicize standards for the electronic exchange, privacy, and security of health information.

- **Privacy Rule:** addresses the need for healthcare organizations to allow patients to choose how their information is disclosed
- **Security Rule:** addresses the security of electronic transmission of medical information

Information Assurance: Information operations that protect and defend information and systems by ensuring their availability, integrity, authentication, confidentiality, and nonrepudiation (Maconachy, Schou, Ragsdale, & Welch, 2001).

NIST: National Institute of Standards and Technology

- NIST was founded in 1901 and is a non-regulatory federal agency within the U.S. Commerce Department's Technology Administration. NIST's mission is to develop and promote measurement, standards, and technology to enhance productivity, facilitate trade, and improve the quality of life. (see www.nist.gov)

Privacy: The rights of individuals regarding collection, storage, processing and use in decision making about personal information about themselves (Turn, 1999).

Security: A collection of policies, procedures, and safeguards that help maintain integrity and availability of information systems and control access to their contents (Rindfleisch, 1997).

Telemedicine: "When physicians use electronic communication and information technologies to provide or support clinical care from a distance" (Patel & Rushefshy, 2002, p. 328).

This work was previously published in Encyclopedia of E-Commerce, E-Government, and Mobile Commerce, edited by M. Khosrow-Pour, pp. 600-607, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 6.6

Information Security Management in Picture Archiving and Communication Systems for the Healthcare Industry

Carrison KS Tong
Pamela Youde Nethersole Eastern Hospital & Tseung Kwan O Hospital, Hong Kong

Eric TT Wong
The Hong Kong Polytechnic University, Hong Kong

INTRODUCTION

Like other information systems in banking and commercial companies, information security is also an important issue in the healthcare industry. It is a common problem to have security incidences in an information system. Such security incidences include physical attacks, viruses, intrusions, and hacking. For instance, in the U.S.A., more than 10 million security incidences occurred in the year of 2003. The total loss was over $2 billion. In the healthcare industry, damages caused by security incidences could not be measured only by monetary cost. The trouble with inaccurate information in healthcare systems is that it is possible that someone might believe it and do something that might damage the patient. In a security event in which an unauthorized modification to the drug regime system at Arrowe Park Hospital proved to be a deliberate modification, the perpetrator received a jail sentence under the Computer Misuse Act of 1990. In another security event (The Institute of Physics and Engineering in Medicine, 2003), six patients received severe overdoses of radiation while being treated for cancer on a computerized medical linear accelerator between June 1985 and January 1987. Owing to the misuse of untested software in the control, the patients received radiation doses of about 25,000 rads while the normal therapeutic dose is 200 rads. Some of the patients reported immediate symptoms of burning and electric shock. Two died shortly afterward and others suffered scarring and permanent disability.

BS7799 is an information-security-management standard developed by the British Standards Institution (BSI) for an information-security-management system (ISMS). The first part of BS7799, which is the code of practice for information security, was later adopted by the International Organization for Standardization (ISO) as ISO17799. The second part of BS7799 states the specification for ISMS. The picture-archiving and -communication system (PACS; Huang, 2004) is a clinical information system tailored for the management of radiological and other medical images for patient care in hospitals and clinics. It was the first time in the world to implement both standards to a clinical information system for the improvement of data security.

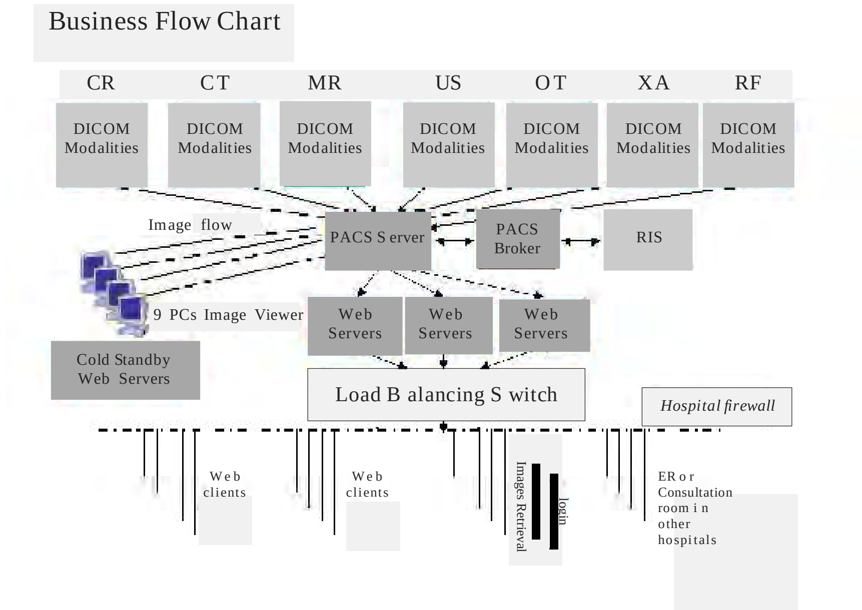
BACKGROUND

Information security is the prevention of, and recovery from, unauthorized or undesirable destruction, modification, disclosure, or use of information and information resources, whether accidental or intentional. A more proactive definition is the preservation of the confidentiality, integrity, and availability (CIA) of information and information resources. Confidentiality means that the information should only be disclosed to a selected group, either because of its sensitivity or its technical nature. Information integrity is defined as the assurance that the information used in making business decisions is created and maintained with appropriate controls to ensure that the information is correct, auditable, and reproducible. As far as information availability is concerned, information is said to be available when employees who are authorized access, and whose jobs require access, to the information can do so in a cost-effective manner that does not jeopardize the value of the information. Also, information must be consistently available to conduct business smoothly. Business-continuity planning (BCP) includes provisions for assuring the availability of the key resources (information, people, physical assets, tools, etc.) necessary to support the business function.

The origin of ISO17799/BS7799 goes back to the days of the UK Department of Trade and Industry's (DTI) Commercial Computer Security Centre (CCSC). Founded in May 1987, the CCSC had two major tasks. The first was to help vendors of IT security products by establishing a set of internationally recognised security-evaluation criteria and an associated evaluation and certification scheme. This ultimately gave rise to the information technology security-evaluation criteria (ITSEC) and the establishment of the UK ITSEC scheme. The second task was to help users by producing a code of good security practices and resulted in the *Users Code of Practice* that was published in 1989. This was further developed by the National Computing Centre (NCC) and later a consortium of users, primarily drawn from British industry, to ensure that the code was both meaningful and practical from a user's point of view. The final result was first published as the British Standards guidance document PD 0003, *A Code of Practice for Information Security Management*, and following a period of further public consultation, it was recast as British Standard BS7799: 1995. A second part, BS7799-2: 1998, was added in February 1998. Following an extensive revision and public consultation period in 1997, the first revision of the standard, BS7799: 1999, was published in April 1999. Part 1 of the standard was proposed as an ISO standard via the "fast track" mechanism in October 1999, and then published with minor amendments as ISO/IEC 17799: 2000 on December 1, 2000. BS7799-2: 2002 was officially launched on September 5, 2002.

PACS is a filmless (Dreyer, Mehta, & Thrall, 2001) and computerized method of communicating and storing medical image data such as computed radiographic (CR), digital radiographic (DR), computed tomographic (CT), ultrasound (US), fluoroscopic (RF), magnetic resonance (MRI), and other special X-ray (XA) images. A

Figure 1. Business flowchart of the Tseung Kwan O Hospital picture-archiving and -communication system



PACS consists of image and data acquisition and storage, and display stations integrated by various digital networks. Full PACS handles images from various modalities. Small-scale systems that handle images from a single modality (usually connected to a single acquisition device) are sometimes called *mini-PACS*.

The medical images are stored in an independent format. The most common format for image storage is DICOM (Digital Imaging and Communications in Medicine), developed by the American College of Radiology and the National Electrical Manufacturers' Association.

Tseung Kwan O Hospital (TKOH) is a newly built general acute hospital (built in 1999) with 458 in-patient beds and 140 day beds. The hospital is composed of several clinical departments including medicine; surgery; paediatrics and adolescent medicine; eye, ear, nose, and throat; accident and emergency, and radiology. A PACS was built in its radiology department in 1999. The PACS was connected with the CR, CT, US, RF, DSA (Digital Subtraction Angiogram), and MRI system in the hospital. The hospital has become filmless since a major upgrade of the PACS in 2003.

An ISO17799/BS7799 ISMS was implemented in the TKOH PACS in 2003. During the implementation, a PACS security forum was established with the active participation of radiologists, radiographers, medical physicists, technicians, clinicians, and employees from the information technology department (ITD). After a BS7799 audit conducted in the beginning of 2004, the TKOH PACS was the world's first system with the ISMS certification.

In this article, the practical experience of the ISO17799/BS7799 implementation and the quality-improvement process of such a clinical information system will be explained.

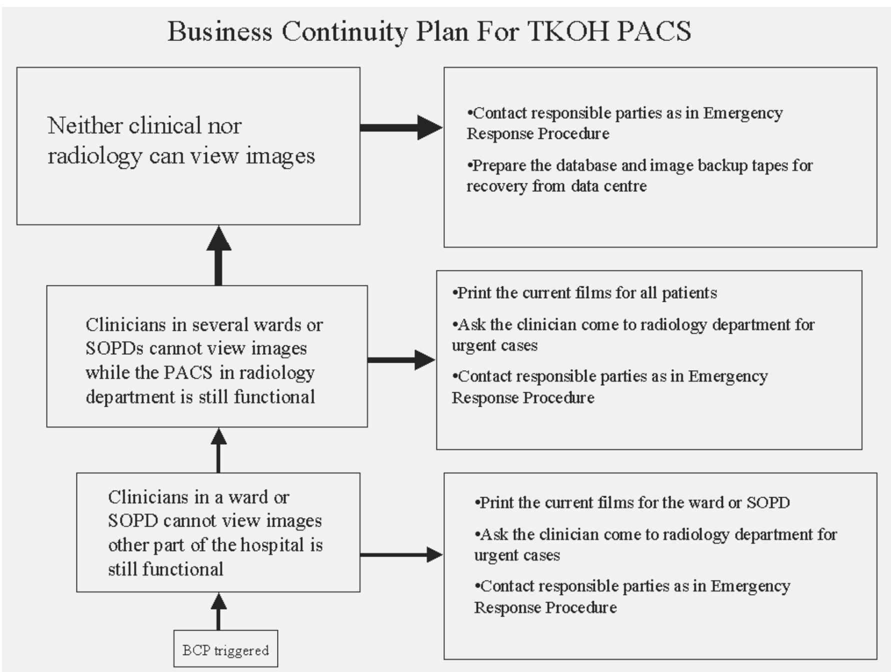
MAIN FOCUS OF THE ARTICLE

In TKOH, the PACS serves the whole hospital including all clinical departments. The implementation of ISO17799 and BS7799 was started with

Table 1. Result of BIA

ess No.	Location			Person		Level	bility	of Impor tance
1	PACS broker	Hardware failure	Patient demographic-data retrieval	Radiographers, ITD	Manual input of patient demographic data	1	1	1
2	PACS servers	Hardware failure	Image receiving	PACS team	PACS cannot receive new images	2	1	2
3	SAN	Hardware failure	Image online storage	PACS team	No online image available in PACS. User still can view the images in the Web servers.	2	1	2
4	PACS servers	Hardware failure	Image verification	PACS team	Image data maybe different from what is in the RIS	1	1	2
5	Image viewers	Hardware failure	Image reporting	PACS team, radiographers	Radiologists cannot view images in the PACS server for advanced image processing and reporting. However, they can still see the images in the Web servers.	2	1	2
6	Jukebox	Hardware failure	Image archiving to MOD jukebox	PACS team	Long-term archiving of the images. There is a risk of lost images in the SAN.	2	2	4
7	Tape library	Hardware failure	Image archiving to tape library	PACS team	Another copy of long-term archiving. There is a risk of lost images in the SAN.	2	1	2
8	Jukebox	Hardware failure	Image prefetching from MOD jukebox	Radiologists, radiographers	Users cannot see the previous images. They cannot compare the present study with the previous.	2	2	4
9	Tape library	Hardware failure	Image prefetching from tape library	Radiologists, radiographers	Users cannot see the previous images. They cannot compare the present study with the previous.	2	1	2
10	Web servers	Hardware failure	Image distribution to clinicians	Clinicians, radiographers	The clinician cannot make a diagnosis without the images.	3	2	6
11	Cisco switches	Hardware failure	Image distribution through Cisco switches	Clinicians, radiographers	The clinician cannot make a diagnosis without the images.	3	1	3
12	Load-balancing switch	Hardware failure	Web-server load balancing	PACS team	The clinician cannot make a diagnosis without the images.	2	1	2
13	RAS server and Cisco router	System malfunction	Remote maintenance	PACS team	Vendor cannot do maintenance remotely.	1	1	1

Figure 2. Business-continuity plan for the TKOH PACS



the establishment of an ISMS for the PACS at the beginning of 2003. For effective implementation of ISO17799 and BS7799 in general, four steps will be required.

1. Define the scope of the ISMS in the PACS.
2. Make a risk analysis of the PACS.
3. Created plans as needed to ensure that the necessary improvements are implemented to move the PACS as a whole forward toward the BS7799 objective.
4. Consider other methods of simplifying the above and achieving compliance with minimum effect.

Implementation of BS7799 Controls in the TKOH PACS Security Forum

A PACS security forum was established for the effective management of all PACS-related security issues in the hospital. The members of the forum were the hospital chief executive, radiologist, clinician, radiographers, medical physicists, technicians, and representatives from the information technology department. One of the major functions of the PACS security forum was to make the security policies for the management of the PACS (Peltier, 2001a). Regular review of the effectiveness of the management was also required.

Business-Continuity Plan

BCP (Calder & Watkins, 2003) is a plan that consists of a set of activities aimed at reducing the likelihood and limiting the impact of disaster events on critical business processes. By the practice of BCP, the impact and downtime of the hospital's PACS system operation due to some change or failure in the company operation procedure is reduced. BCP is used to make sure that the critical part of the PACS system operation is not affected by a critical failure or disaster. The design of this BCP is based on the assumption that the largest disaster is a complete breakdown of the PACS room in the radiology department of TKOH. The wards, the specialist out-patient department (SOPD), and the imaging modalities should still all be functional.

During the design of a BCP, a business-impact analysis (BIA) of the PACS was studied. The BIA was a study of the vulnerabilities of the business flow of the PACS, and it is shown in the following business flowchart.

In the above flowchart, image data were acquired by the CR, DR, CT, US, RF, MRI, XA, and other (OT) imaging modalities such as a film digitizer. The acquired image data were centrally archived to the PACS server, which connected to a PACS broker for the verification of patient demographic data with the information from the Radiology Information System (RIS). In the PACS server, a storage-area network (SAN), a magneto-optical disk (MOD) jukebox, and a tape library were installed for short-term, long-term, and backup storage. The updated or verified image was redirected to the Web server cluster (Menasce & Almeida, 2001) for image distribution to the entire hospital including the emergency room (ER) and consultation room. The load-balancing switch was used for nonstop service of image distribution to the clinicians. A cluster of Cisco switches was installed and configured for automatic fail-over and firewall purposes. The switches connecting between the PACS network and hospital network were maintained by the information technology department (*A Practical Guide to IT Security for Everyone Working in Hospital Authority*, 2004; *Security Operations Handbook*, 2004). A remote-access server was connected to the PACS for the remote service of the vendor.

Business-Impact Analysis

In the BIA (Peltier, 2001b), according to the PACS operation procedure, all potential risks and impacts were identified. The responsibilities

Table 2.

Step	Recovering Subprocess	Responsible Person	Process Location
1	Image distribution to clinicians	PACS team, contractor	Web servers
2	Image distribution through Cisco switches	PACS team, contractor	Cisco switches
3	Image online storage	PACS team, contractor	PACS servers, SAN
4	Image reporting	PACS team, contractor	Image viewers
5	Image prefetching from MOD jukebox	PACS team, contractor	MOD jukebox
6	Image prefetching from tape library	PACS team, contractor	Tape library
7	Image receiving	PACS team, radiographers, contractor	PACS servers
8	Image verification	PACS team, radiographers, contractor	PACS servers
9	Web-server load balancing	PACS team, contractor	Load-balancing switch
10	Image archiving to MOD jukebox	PACS team, contractor	Jukebox
11	Image archiving to tape library	PACS team, contractor	Tape library
12	Patient demographic-data retrieval	Radiographers, ITD	PACS broker
13	Remote maintenance	PACS team	RAS server and Cisco router

Table 3.

DRP Level Triggered	Scope	Recovery Time
1	Clinicians in a ward or the SOPD cannot view images while other parts of the hospital are still functional.	Half day for the recovering of subprocess no. 10
2	Clinicians in several wards or the SOPDs cannot view images while the PACS in the radiology department is still functional.	One day for the recovering of subprocess nos. 10 and 11
3	Neither the clinical department nor radiology can view images.	One week for the recovering of subprocess nos. 1 to 13

of relevant teams or personnel were identified according to the business flow of the PACS. The critical risk(s), which may affect the business operation of the PACS, could be determined by performing a risk evaluation of the potential impact. One of the methods in the BIA was to consider the contribution of the possibility of risk occurrence for prioritization purposes. The result of the BIA is shown in the following table.

In table 1, the responsible person for each business subprocess was identified to be PACS team, radiologists, radiographers, clinicians, or the information technology department. The most critical subprocess in the TKOH PACS was associated with the Web servers. Once the critical subprocess was identified, the BCP could be designed for the system as shown in the following figure. A responsible person for the BCP was also assigned.

Disaster-Recovery Plan

Disaster-recovery planning (DRP; Toigo, 1996), as defined here, is the recovery of a system from a specific unplanned domain of disaster events such as natural disasters, or the complete destruction of the system. Following is the DRP for the TKOH PACS, which was also designed based on the result of the above BIA.

Table 4.

Process Flow	Operation	Remark
Document Creation	Manuals, procedures, and work instruction should be written by the PACS team. Records should be kept in the general office.	If documents/manuals cover different departments, we should consider liaisons between different departments' roles.
Document Approval	Manuals should be approved by the chief of service (COS). Procedures and work instruction should be approved by the PACS manager. Records should be stored in the PACS room or general office.	Manual changes should be approved by the PACS manager.
Document Release	1. The distribution of manuals and procedures is controlled by the PACS manager. 2. The requirements from the customers and contracts related to information security of the PACS should be approved by the COS and released by the PACS manager.	Documents/manuals related to PACS should be signed by the PACS manager before distribution. The manual distributed should have a document number. Each personnel/department should update the document-control list regularly.
Document Revision	Manuals and documents should be amended by the document owner/department. If other personnel/departments are involved in the change, they should seek the approval from the owner/responsible departments.	Note the change and where the change is (e.g., which paragraph) on the first page. The original document/manual should be chopped or destroyed.
Document Check	For general manuals from an outsourcing party (e.g., Afga) or other department, if they are applicable for PACS operation, they should be approved and adopted for PACS operation.	For this kind of manual, if it has not been revised for 1 year, it should be reviewed.
Document Obsolescence	Obsolete documents should be collected by the PACS manager. There should be one copy (soft copy or hard copy) kept by the PACS.	Each personnel/department should keep the previous updated version of the document for future review. The other obsolete copy should be destroyed.
Document Execution	It should be guaranteed that the operator or other related PACS engineer should get the right document in the right version.	During operation, no document should be copied, duplicated, or distributed without appropriate approval.

Recovery Time for the DRP

During disaster recovery, timing was also important both for the staff and the manager. The recovery times of some critical subprocesses are listed as in the following table.

Backup Plan

Backup copies of important PACS system files, patient information, essential system information, and software should be made and tested regularly.

Security and Security-Awareness Training

Training (education concerning the vulnerabilities of the health information in an entity's possession and ways to ensure the protection of that information) includes all of the following implementation features.

i. Awareness training for all personnel, including management personnel (in security awareness, including, but not limited to, password maintenance, incident reporting, and viruses and other forms of malicious software)
ii. Periodic security reminders (employees, agents, and contractors are made aware of security concerns on an ongoing basis)
iii. User education concerning virus protection (training relative to user awareness of the potential harm that can be caused by a virus, how to prevent the introduction of a virus to a computer system, and what to do if a virus is detected)
iv. User education in the importance of monitoring log-in success or failure and how to report discrepancies (training in the user's responsibility to ensure the security of healthcare information)
v. User education in password management (type of user training in the rules to be followed in creating and changing passwords and the need to keep them confidential)

Documentation and Documentation Control

Documentation and documentation control serve as a control on the document and data drafting, approval, distribution, amendment, obsolescence, and so forth to make sure all documents and data are secure and valid.

Standard and Legal Compliance

The purpose of standard and legal compliance (Hong Kong Personal Data Privacy Ordinance, 1995) was to avoid breaches of any criminal and civil law; statutory, regulatory, or contractual obligations; and any security requirements. Furthermore, the equipment compliance of the DICOM standard can improve the compatibility and upgradability of the system. Eventually, it can save costs and maintain data integrity.

Quality of PACS

In a filmless hospital, the PACS is a mission-critical system for lifesaving purposes. The quality of the PACS was an important issue. One method to measure the quality of a PACS was measuring the completeness of the system in terms of data confidentiality, integrity, and availability. A third-party audit such as the ISO17799/BS7799 certification audit could serve as written proof of the quality of a PACS.

FUTURE TRENDS

Based on the experience in BS7799 implementation, the authors were of the view that more and more hospitals would consider similar healthcare applications of BS7799 to other safe-critical equipment and installations in Hong Kong.

CONCLUSION

ISO17799/BS7799 covers not only the confidentiality of the system, but also the integrity and availability of data. Practically, the latter is more important for the PACS. Furthermore, both standards can help to improve not only the security, but also the quality of a PACS because, to ensure the continuation of the certification, a security forum has to be established and needs to meet regularly to review and improve on existing processes.

REFERENCES

British Standards Institution. (2000). *Information technology: Code of practice for information security management* (BS ISO/IEC 17799: 2000 [BS 7799-1:2000]). UK: British Standards Institution.

British Standards Institution. (2002). *Information security management systems: Specification with guidance for use* (BS 7799-2 2002). UK: British Standards Institution.

Calder, A., & Watkins, S. (2003). *IT governance: A manager's guide to data security and BS 7799/ ISO 17799.* London: Kogan Page.

Dreyer, K. J., Mehta, A., & Thrall, J. H. (2001). *PACS: A guide to the digital revolution*. New York: Springer-Verlag.

Hong Kong Personal Data Privacy Ordinance. (1995). Hong Kong, China: Hong Kong Government.

Huang, H. K. (2004). *PACS and imaging informatics: Basic principles and applications*. Hoboken, NJ: Wiley-Liss.

The Institute of Physics and Engineering in Medicine. (2003). *Guidance notes on the recommendations for professional practice in health, informatics and computing.* UK and Institute of Physics and Engineering in Medicine.

Menasce, D. A., & Almeida, V. A. F. (2001). *Capacity planning for Web services: Metrics, models, and methods.* Upper Saddle River, NJ: Prentice Hall.

Peltier, T. R. (2001a). *Information security policies, procedures, and standards: Guidelines for effective information security management.* Boca Raton, FL: CRC Press.

Peltier, T. R. (2001b). *Information security risk analysis.* Boca Raton, FL: Auerbach Publishing.

A practical guide to IT security for everyone working in hospital authority. (2004). Hong Kong, China: Hong Kong Hospital Authority IT Department.

Security operations handbook. (2004). Hong Kong, China: Hong Kong Hospital Authority IT Department.

Toigo, J. W. (1996). *Disaster recovery planning for computer and communication resources.* John Wiley & Sons.

Toigo, J. W. (2003). *The holy grail of network storage management.* Prentice Hall.

KEY TERMS

Availability: Prevention of unauthorized withholding of information or resources.

Business-Continuity Planning: The objective of business-continuity planning is to counteract interruptions to business activities and critical business processes from the effects of major failures or disasters.

Confidentiality: Prevention of unauthorized disclosure of information.

Controls: These are the countermeasures for vulnerabilities.

Digital Imaging and Communications in Medicine (DICOM): Digital Imaging and Communications in Medicine is a medical image standard developed by the American College of Radiology and the National Electrical Manufacturers' Association.

Information-Security-Management System (ISMS): An information-security-management system is part of the overall management system, based on a business risk approach, to develop, implement, achieve, review, and maintain information security. The management system includes organizational structure, policies, the planning of activities, responsibilities, practices, procedures, processes, and resources.

Integrity: Prevention of unauthorized modification of information.

Picture-Archiving and -Communication System (PACS): A picture-archiving and -communication system is a system used for managing, storing, and retrieving medical image data.

Statement of Applicability: Statement of applicability describes the control objectives and controls that are relevant and applicable to the organization's ISMS scope based on the results and conclusions of the risk assessment and treatment process.

Threats: These are things that can go wrong or that can attack the system. Examples might include fire or fraud. Threats are ever present for every system.

Vulnerabilities: These make a system more prone to attack by a threat, or make an attack more likely to have some success or impact. For example, for fire, a vulnerability would be the presence of inflammable materials (e.g., paper).

This work was previously published in Encyclopedia of Multimedia Technology and Networking, edited by M. Pagani, pp. 396-403, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 6.7
Modelling Context-Aware Security for Electronic Health Records

Pravin Shetty
Monash University, Australia

Seng Loke
La Trobe University, Australia

INTRODUCTION

The Internet has proven to be the most convenient and demanding facility for various types of businesses and transactions for the past few years. In recent years, business information systems have expanded into networks, encompassing partners, suppliers, and customers. There has been a global availability (Anderson, 2001; BSI Global, 2003) of resources over the Internet to satisfy different needs in various fields. The availability factor has called for various security challenges in fields where information is very valuable and not meant for all. Potential threats to information and system security come from a variety of sources. These threats may result in violations to confidentiality, interruptions in information integrity, and possible disruption in the delivery of services. So it is essential to manage the flow of information over the network with the required level of security. There are many security technologies and models that have been introduced which are capable of realizing the functions and objectives of information system security.

This article first gives a brief overview of what we term basic security policies of an integrated security model. Then it suggests context-based security policies for a health organization scenario using contextual graphs augmented with details about specific security actions, which relate to the security policies enumerated in the integrated security model.

The plan of the article is as follows. We first overview the three concepts in detail and briefly describe the concept of contextual (meta-policy) graphs. We then develop a context-based security meta-policy for securing patient records based on the security policies overviewed and discuss related work, before concluding the paper.

BASIC SECURITY POLICIES

Mobile ambients were first proposed by Cardelli and Gordon (1998a, 1998b) and then further extended by Bugliesi, Castagna, and Crafa (2004); and Braghin et al. (2002) were very efficient in modeling multilevel security issues. These three notions are very effective in modeling a foolproof security solution in a computing scenario by stating various security steps to be taken in the corresponding scenario. On this basis we have five cases that form the basic security policies in this article which we note can be concisely and precisely modeled using the mobile ambients formalism, though we omit such details of the formalism here and only describe the policies in plain language. The article uses them in appropriate scenarios depending on the context. Thus, the combined use of these five policies and a contextual graph representing the contexts of use of these policies provides a context-based security solution for pervasive environments. This section briefly describes the five policies using ambient (representing a boundary of security restrictions) notions.

Policy 1: Authenticate Returning Mobile Agent

When a privileged process (agent or person) leaves the parent ambient (e.g., a host institution) to execute some external independent activities, it relinquishes its local privileges and authority within its bounding parent ambient and ambient community. It exits the parent and might later return to the parent ambient. At this point an *authentication mechanism* is needed to check the authenticity of the returning original process. Cardelli and Gordon (1998a, 1998b, 1999) suggest that these high-level privileges must not be automatically restored to the returning agents/ processes without first verifying their identity. This is to preserve the security and integrity of the ambient as well as the services and resources contained within it.

Policy 2: Firewall Access

If any agent/process has to enter an ambient, it has to know the name of the ambient and also possess the capability to enter it. The functionality of firewall is achieved with the help of restriction primitives and with the help of anonymity of the ambient name. Thus without knowing the ambient name, no process or agent can exit or enter the parent ambient. This helps in achieving protection of the resources from unwanted agents. The ambient name could be interpreted as a secret password.

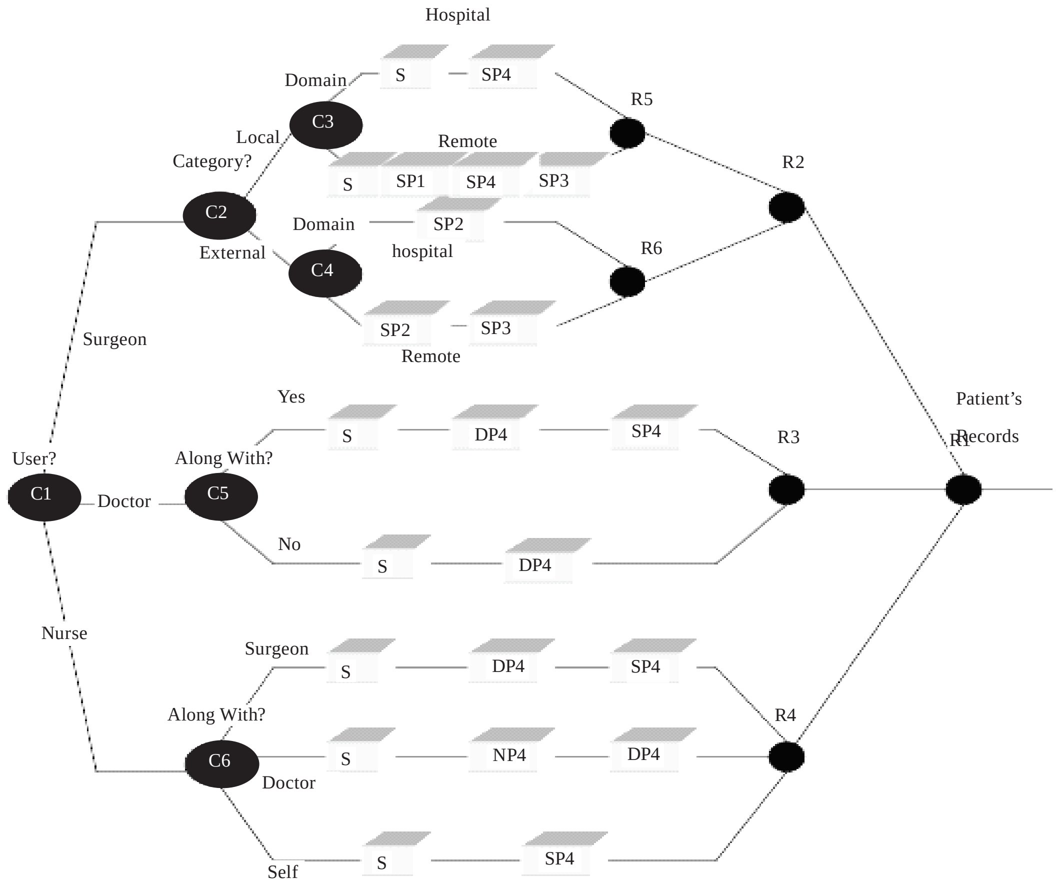
Policy 3: Encryption Using Shared Keys to Secure the Data While Communicating

Cardelli and Gordon (1998a, 1998b, 1999) also put forth the encryption primitives to communicate between two ambients or between an ambient and a remote agent. These primitives helped in maintaining the *confidentiality* of the message or data. Consider a Plaintext message *M*. The encryption of the plaintext message is done with the help of the encryption key *k*. A name can represent a shared key, as long as it is kept secret and shared only by certain parties. A shared key can be reused multiple times, for example, to encrypt a stream of messages. A message encrypted under a key *k* can be represented as a folder that contains the message and whose label is *k* (Cardelli & Gordon, 1998a, 1998b).

Policy 4: Security Across Multiple Levels

In general, an enclosed ambient environment would typically contain numerous subambients as well as active processes, agents, and information

Figure 1. Context graphs showing the various possible scenarios and security actions taken



resources. These groups of subambients within an ambient may be arbitrarily nested and organized in a hierarchical structure. Ambients and processes at the higher level of the nested structure are responsible for managing resources that are more vital and important than those at a lower level. In such multilevel environments, it is necessary to restrict the access to the flow of information depending upon the need and the security levels. Information can only flow from lower levels of security to higher levels and not conversely. A policy for this assigns levels to users and restricts information flow among the users.

Policy 5: Movement of Data and Entities Through Different Communities

The multilevel security policy mandatory access control security in the boxed ambients provided restricted access to information based on the various security levels in the hierarchical levels. The access is defined by the level at which the agents are which are predetermined based on their needs. But Braghin et al. (2002) were of the view that the implementation of mandatory access control security is complex, as agents and processes may move from one security level to another. The agents themselves may be confidential or may be carrying secure/confidential information. Thus there is no way of ensuring the agents will not be illegally attacked, accessed, or executed by untrustworthy entities at the lower security levels. The *security boundary* concept put forth by Braghin et al. (2002) guarantees absence of information leakage.

According to this concept, every high-level data or process should be encapsulated in a boundary ambient. A boundary ambient can be opened only when it is nested into another pre-specified boundary ambient. A policy for this states that the protected information cannot be read without being contained within some safety boundary (e.g., physically, an item cannot be viewed in the absence of a bodyguard).

CONTEXTUAL (META-POLICY) GRAPHS

Contextual meta-policy graphs are derived from contextual graphs (Braghin et al., 2002; Bugliesi et al., 2001a, 2001b). We replace the security actions in contextual graphs by security policies, which in turn, represent the security actions accordingly. By virtue of this embedding of policies (such as the five mentioned above) into a contextual graph, the graph becomes a meta-policy construct. The contextual meta-policy graphs are very general and can be used to depict security architecture in any scenario. The use of such graphs is to provide a high-level picture of the security framework, thereby avoiding lower details (security actions), which makes the overall architecture less cumbersome. The security actions are triggered according to the policies, which are predefined and programmed. The five security policies depicted in the integrated (in that such policies complement one another and have holistic coverage over security situations) security model above forms the basis of the graphs presented in this article. The contextual meta-policy graphs use various combinations of these five policies to define the various access paths to the patient's records. The next section presents the overall security architecture using such graphs.

SECURING ELECTRONIC HEALTH RECORDS USING CONTEXTUAL META-POLICY GRAPHS

The integrated security model discussed above is used to implement context-based security. The following model can be used in any context-aware situation. This article will explain the use of the model for securing health records. The various access possibilities depending on the context are described with the help of the contextual meta-policy graphs. This section is divided into two subsections. The first gives the details of the security policies used along with the low-level details (what security actions are triggered in each policy). The second subsection gives the actual approach for security purposes, along with an explanation of the individual access possibilities.

POLICY DEFINITION

This section outlines the various types of users associated with the electronic health records. It also specifies the various types of contextual information to be used for implementing the security policies. Finally, the various security actions to be taken depending upon the information received are also stated.

Roles: Surgeon, Doctor, Nurse.

Contextual Information Considered: Role, Location, Place

The five policies, collectively labeled the 'integrated security model', can be effectively used for implementing security in a hospital scenario for accessing a patient's records. The five policies can be described in context for the hospital scenario as follows.

The Secret code *S* in the diagram can be considered as a secret name of the hospital network, which has to be known by each and every local person who wants to access the patient's detail. This secret code is similar to the secret name of the ambient in the case of ambient terms. It is not known to foreign entities. The details about the individual usernames and passwords are stored in the systems database along with the corresponding roles.

Policy 1: Authentication

This security policy defines the way in which a valid user can access the hospital network once s/he is out of it. This type of authentication is required for the surgeon and the doctor who have to access the network and hence the patient's health records. When the surgeon goes out of the hospital network, temporarily s/he is given a secret password, which helps her/him to authenticate her/himself when s/he wants to access the network again. Apart from the local surgeon and the local doctor, no one else can access the records from outside the hospital. This security policy triggers the action such as asking the user for the secret password, which will help her/him to authenticate her/himself.

Policy 2: Foreign Agent Authentication

This security policy defines the access method for a foreign entity such as a surgeon from some other medical institution in case of emergency. The security infrastructure provides such foreign agents, which are required in case of an emergency with some special combination of passwords analogous to that of the foreign agents concept in mobile ambients. This security policy triggers actions such as asking the foreign agent for the three sets of passwords given to her/him. The first will help her/him to validate her/himself. In case the access is remote, then it is a password; otherwise, it is done by retina scan or any other biometrics. The second password is used to allow the external surgeon inside the network. The third password will help her/him to access the patient's records in the mode as per her/his role.

Policy 3: Encryption of Data

This policy secures the data from falling into destructive hands by using various encryption techniques. This security policy is required when the user is accessing the network from a remote place. For example, there might be a case where a local surgeon might have to access the patient record from a remote place and do some modifications according to the present condition of the patient. This transfer of data should be safe and confidential and not intercepted by a malicious intruder. Thus, encryption is required which is provided by this policy.

Policy 4: Security Levels

This security policy talks about role-based access control. In the hospital, the three main users considered in this article are the surgeon, the doctor, and the nurse. They have access to the patient's records in different modes and according to the need of their positions. For example the surgeon has the access to the records in all three modes (i.e., read/write/update). A doctor's access is limited to two modes (i.e., read/write), whereas the nurse's access is restricted to just one mode (i.e., read). This type of security hierarchy is analogous to the security level structure policy of the multilevel access model. This helps control access to the valuable resources depending on the role of the user in the medical institution.

Policy 5: Third-Party Authentication

This security policy discusses authentication from the third mediating party in the communication between two entities from different medical institutions. This type of authentication is required to make sure that the entities involved in the communication are authorized. This security policy plays an important role when an outside surgeon needs access to the medical record or communicates with the local network of the medical institution from a remote place. The third party should be a reliable party, giving the authorization of any information passed between the two communicating entities.

The five security policies defined above are based on the concepts of mobile ambients defined in the former sections and depict the various security approaches taken depending on the context. The next subsection elaborates on the various contexts in a hospital scenario and the appropriate security policies to be taken to have secured access to the patient's records using a context meta graph.

Security Paths Defined in a Contextual Meta-Policy Graph

The above model gives the overall security approach used in a hospital environment to access electronic health records of the patients. The behavior of the security infrastructure according to various contexts is described as follows:

1. **User→ Type→ Domain :: Surgeon→Local→Hospital:-** ***Policy 4.*** If the user is a local surgeon and s/he wants to access the patient's records from the hospital, then s/he has to follow security policy 4. When s/he tries to access the records, the system will ask for her/his username and password, which will depict her/his role in the institution. Being a surgeon, s/he enjoys the highest level of rights. S/he can access the records in all the three modes (read/write/update). For access from within the hospital, authentication can be provided by various biometrics.
2. **User→ Type→ Domain :: Surgeon→Local→Remote:-** ***Policy 1 + Policy 4 + Policy 3.*** When the local surgeon, say John, wants to access the information from outside the hospital, which can be in the case of an emergency, then a combination of three security policies is followed. First he has to get back into the hospital network by providing his secret key/password according to policy 1, which will validate him as a local user. Then with the help of his username and password, which is associated with his role (policy 4), he can access the records. As he is in a remote place, care should be taken that the data is not visible to the outside world. This is achieved by encrypting the channel according to policy 3. Thus, using the combination of three policies, the local surgeon is provided access to the records from a remote place.
3. **User→Type→Domain :: Surgeon→External→Hospital:-** ***Policy 2.*** In some emergency cases, it becomes necessary to invite an external surgeon to the local institution. For such cases, the system stores all the information of such emergency persons. Policy 2 defines the access for such a surgeon using predetermined passwords. The surgeon is provided with three sets of passwords. The first will help him/her to validate him/herself. The second password is used to allow the external surgeon into the network. The third password will help him/her access the patient's records in the mode according to his/her role. Biometrics instruments can provide the required authentication when he/she is accessing from the hospital. If the external surgeon is accessing the records from the hospital, then he/she can use his/her third password to get access according to his/her

role. The first and the second passwords are not required in this case.

4. **User→Type→Domain :: Surgeon→External→Remote:- *Policy 2 + Policy 3.*** When the external surgeon has to access the records from a remote place, then the access takes place as determined by the combination of two security policies. Policy 2 is as defined as above. In such a case, he/she must have all three sets of passwords with him/her in order to get into the hospital network and then access the patient's records. Policy 3 is used for encryption of the information to provide safe and confidential communication.

   The surgeon, whether local or external, does not need to be with any other staff, as he/she enjoys the maximum access rights. He/she has to follow the appropriate security procedures depending upon his/her category (i.e., local or external) and his/her location of access.

5. **User→Along with :: Doctor→ Along with Surgeon:- *Policy 4 (for doctor) + Policy 4 (for surgeon).*** When a local doctor wants to access the records, then s/he can do so by using a combination of policy 4 applied to two roles. If s/he is with the surgeon, then s/he can access the patient's record in a full mode (read/write/edit) with the same access privileges as the surgeon. But for this, the surgeon has to first specify her/his username and password according to policy 4 so that her/his role is specified. If the surgeon is not physically present in the hospital, then the doctor cannot access the records in full mode. Also, a doctor is not allowed access to the records from a remote place.

6. **User→Along with :: Doctor→ Alone:- *Policy 4 (for doctor).*** If a surgeon is not with the local doctor, then the doctor accesses the patient's information in the restricted mode (read/write). S/he can access the information according to policy 4. S/he has to present her/his username and password so that her/his role will be represented in the system. An important point in this security architecture is that the doctor can never access the patient's records from a remote place.

7. **User→Along with :: Nurse→ Along with Surgeon:- *Policy 4 (for nurse) + Policy 4 (for surgeon).*** A nurse, say Jane, is at the lowest security level in the hospital hierarchy. If she is with the surgeon, she is allowed to view the records in full mode as the surgeon has. For that, the surgeon must present his/her role to the system first, and then the nurse can access the information according to her role as in security policy 4. Remote access is not allowed in this case.

8. **User→Along with :: Nurse→ Along with Doctor:- *Policy 4 (for nurse) + Policy 4 (for doctor).*** When the nurse, Jane, is with a doctor in the hospital, then she can have access to the patient's information according to the doctor's privileges. The doctor first provides his/her username and password, and then the nurse can provide her role as per security policy 4. Remote access is also not allowed in this case.

9. **User→Along with :: Nurse→ Alone:- *Policy 4 (for nurse).*** When nurse Jane is alone, she is only allowed one mode (read). She can read the information but cannot delete or modify it. She can access the information using her username and password according to security policy 4. But she is not allowed to access the records from a remote place.

RELATED WORK

There has been some work on security policies in the field of electronic health records systems in past years. Reid, Cheong, Henricksen, and Smith (2003) presented a model that uses role-based access control to restrict the access to the health records on a need-to-know basis. The prototype described maintains databases consisting of explicit 'allow' and explicit 'denial' lists. The proposed model also permits allow and deny policies to successively qualify each other in a role hierarchy supporting inheritance. Thus, the access control framework exhibits a great flexibility and efficiency in the range of access policies that it can support.

Mostéfaoui and Brézillon (2004) put forth the concept of contextual graph for modeling security in context-aware environments. They present a new model for policy specification based on the new approach. The security policy based on such an approach depends on the contextual information of the user and the environment. Contextual graphs have proved to be very effective in modeling a complex situation. Mostéfaoui and Brézillon (2004) also mention how contextual graphs are used to model security in a context-aware environments. In their paper they gave an example of how context-based security is used in a hospital scenario, but it does not employ our meta-policy scheme in the way we do above.

CONCLUSION

Due to the ubiquitous nature of the today's computing world, security is of utmost important. The traditional static authentication techniques are no longer valid and justified. This situation is due to the lack of consideration for context in existing security systems. Context-based security helps the security policy to, in effect, adapt to the new "threats" as they come. It aims at providing flexible security models for distributed infrastructures, where the user and application environments are continually changing. In this article, we have presented an approach that helps with context-based security in a medical scenario. The type and nature of the authentications that are demanded by the security policy depend on the information that is collected from the environment. Further, the contextual graph approach helps to add/modify secure paths based on the newly detected contexts that need to be utilized for fine-grain security. The model presented is a generalized model that can be used in any context-aware environment or enterprise, from the office to factories.

REFERENCES

Anderson, R. (2001). *Security engineering*. New York: John Wiley & Sons.

Braghin, C., Cortesi, A., & Focardi, R. (2002a). Security boundaries in mobile ambients. *Computer Languages, 28(1),101-127.*

Braghin, C., Cortesi, A., Focardi, A., & van Bakel, S. (2002b). *Boundary inference for enforcing security policies in mobile ambients*. Retrieved from http://www.informatics.sussex.ac.uk/users/vs/myths/reports/papers/boundaries-tcs02.pdf

BSI Global. (2003). *Information security*. Retrieved from *h*ttp://www.bsi-global.com/Information+Security/Overview/Why.xalter

Bugliesi, M., Castagna, G., & Crafa, S. (2001a). Boxed ambients. *Proceedings of TACS* (pp. 38-63). Berlin: Springer-Verlag (LNCS 2215).

Bugliesi, M., Castagna, G., & Crafa, S. (2001b). *Reasoning about security in mobile ambients* (pp. 102-120). Berlin: Springer-Verlag (LNCS 2154).

Bugliesi, M., Castagna, G., & Crafa, S. (2004). Access control for mobile agents: The calculus of boxed ambients. *ACM Transactions on Programming Languages and Systems, 26*(1), 57-124.

Cardelli, L. (1999). *Abstraction for mobile ambients*. Retrieved from http://research.microsoft.com/Users/luca/Papers/Abstractions%20for%20Mobile%20Computation.A4.pdf

Cardelli, L., & Gordon, A.D. (1998a). Mobile ambients. *Proceedings of FOSSACS* (pp. 140-155). Berlin: Springer-Verlag (LNCS 1378).

Cardelli, L. and Gordon, A. D. (1998b). *Mobile ambients.* Retrieved from http://www.cis.upenn.edu/~lee/98cis640/Lectures/fm3.ppt#24

Cardelli, L., & Gordon, A.D. (2004). *Mobile ambients.* Retrieved from http://classes.cec.wustl.edu/~cs673/1

Covington, M.J., Fogla, P., Zhan, Z., & Ahamad, M. (2002). *A context-aware security architecture for emerging applications.* Retrieved from http://www.acsac.org/2002/papers/71.pdf

Mostéfaoui, G., & Brézillon, P. (2004a). Modeling context based security with contextual graphs. *Proceedings of the 2nd IEEE Annual Conference on Pervasive Computing and Communications Workshops.*

Mostéfaoui, G., & Brézillon, P. (2004b). Context-based security policies: A new modeling approach. *Proceedings of the 2nd IEEE Annual Conference on Pervasive Computing and Communications Workshops.*

Reid, J., Cheong, I., Henricksen, M., & Smith, J. (2003). *A novel use of RBAC to protect privacy in distributed health care information systems* (pp. 403-415). Information Security Research Center, Queensland University of Technology, Australia.

Wikipedia. (n.d.). *Ubiquitous computing.* Retrieved from http://en.wikipedia.org/wiki/Ubiquitous_computing

KEY TERMS

Contextual Graph: Graph whose edges represent the values that context information take, and have three types of nodes: branching and recombination nodes, and nodes representing security actions. A path through the graph represents a security action taken in response to particular context information.

Mobile Ambients: A process calculi that emphasizes the notion of boundaries and how processes with such boundaries interact.

Pervasive Computing: Integrates computation into the environment, rather

than having computers which are distinct objects. Other terms for ubiquitous computing include ubiquitous computing, calm technology, things that think, and everyware (Wikipedia, n.d.).

Security Action: Action taken to secure a resource, from authentication to encryption to other informational and physical measures (e.g., putting a man on guard).

Security Policy: A description of security actions to take under different circumstances. Such policies are typically specified as rules in a formal language.

This work was previously published in Encyclopedia of Information Ethics and Security, edited by M. Quigley, pp. 463-469, copyright 2007 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 6.8
Outsourcing in the Healthcare Industry:
Information Technology, Intellectual Property, and Allied Aspects

Amar Gupta
University of Arizona, USA

Raj K. Goyal
Harvard Medical School, USA & VA Boston Health Care System, USA

Keith A. Joiner
University of Arizona, USA

Sanjay Saini
Harvard Medical School, USA & Massachusetts General Hospital, USA

ABSTRACT

The healthcare industry is being impacted by advances in information technology in four major ways: first, a broad spectrum of tasks that were previously done manually can now be performed by computers; second, some tasks can be outsourced to other countries using inexpensive communications technology; third, longitudinal and societal healthcare data can now be analyzed in acceptable periods of time; and fourth, the best medical expertise can sometimes be made available without the need to transport the patient to the doctor or vice versa. The healthcare industry will increasingly use a portfolio approach comprised of three closely-coordinated components seamlessly interwoven together: healthcare tasks performed by humans on-site; healthcare tasks performed by humans off-site, including tasks performed in other countries; and healthcare tasks performed by computers without direct human involvement. Finally, this paper deals with intellectual property and legal aspects related to the three-pronged healthcare services paradigm.

INTRODUCTION

Advances in computing and communications technologies are dramatically altering the healthcare landscape around the world in a number of ways such as:

- enabling detailed analysis of healthcare data to elicit underlying trends and interrelationships;
- facilitating storage, transmission, integration, and retrieval of healthcare records;
- enabling healthcare professionals to render assistance to patients separated by significant geographic distance from each other;
- monitoring the safety of medical procedures and pharmaceutical drugs; and
- bringing the latest healthcare information to the attention of healthcare professionals and others.

In this article, we take five operational scenarios, one from each of the five illustrative categories delineated above. In each operational scenario, at least one of the co-authors of this article played a significant role and therefore possesses first-hand knowledge of that healthcare application. The operational scenario is analyzed, post-facto, from the viewpoint of diagnosing what subset of tasks can be handled by evolving information technologies without significant human intervention, what subset needs to be performed on-site by humans, both now and in the foreseeable future, and what subset can be potentially performed by humans located at a significant distance from the patient.

Based on the above analysis, we postulate that the future healthcare industry is unlikely to adopt a mono-operational scenario in which all the tasks occur on-site (as happened in the past), off-site, or by machines alone. Instead, the healthcare industry will gradually adopt an operational model in which there is a seamless and symbiotic combination of all three modes of operation.

After examining the future healthcare industry model from multiple perspectives, we conclude that we need a new approach to intellectual property in order to adequately safeguard the interests of the relevant constituencies. Based on the forces that will motivate the change, we further assert that healthcare organizations that are unwilling to adapt and embrace the evolving three-faceted work paradigm will be at a competitive disadvantage to their peers. National, state, and local medical regulatory agencies will need to respond to market pressures in order to support the long-term interests of both medical professionals and patients in their respective jurisdictions.

COMPREHENSIVE ANALYSIS OF HEALTHCARE DATA

One out of eight women in the United States will develop breast cancer during her lifetime. Early detection is a woman's best defense against breast cancer, which is 97% curable when detected and treated at an early stage. Mammography is the gold standard for screening for breast cancer. With the trend towards people living longer lives and taking proactive measures on their health, the demand for mammography is increasing at a significant pace. Unfortunately, 10-20% of the cancers currently detectable by a screening mammogram are missed by the human radiologist, allowing the disease another year to progress. In addition, there is a high degree of liability on radiologists due to missed diagnoses. To mitigate this problem, some radiology screening centers employ two radiologists to read each case. This approach involves significant cost to support an additional radiologist, reduces the number of total mammograms that can be performed within a center, and is problematic due to the shrinking numbers of radiologists in the field of mammography, especially in the United States.

Based on the latest information available on the FDA Web site (October 2006), there were 8,832 FDA-certified mammography centers and 13,511 accredited units in the United States. In the year 2006, there were 34.6 million mammograms performed in the United States alone, which translates into a total market of $5 billion (at $150 per mammogram). Globally, there are over 200 million mammograms per year, which translates to a market running into several billions of dollars per annum. As the population of women over 40 increases and the awareness of proactive health measures, such as mammograms, gets enhanced, the number of mammograms increases every year. However, a decrease in mammogram centers and the number of radiologists in this field are negatively correlated with the demand.

The area of mammography and the aspect of errors in diagnosis (both false positives and false negatives) have been studied in detail by many researchers (Berlin, 2005; Ghate et al., 2005; Gilbert et al.; 2006; Khoo, Taylor, & Given-Wilson, 2005; Sickles, Wolverton, & Dee, 2002; Skaane, Kshirsagar, Stapleton, Young, & Castellin, 2006).

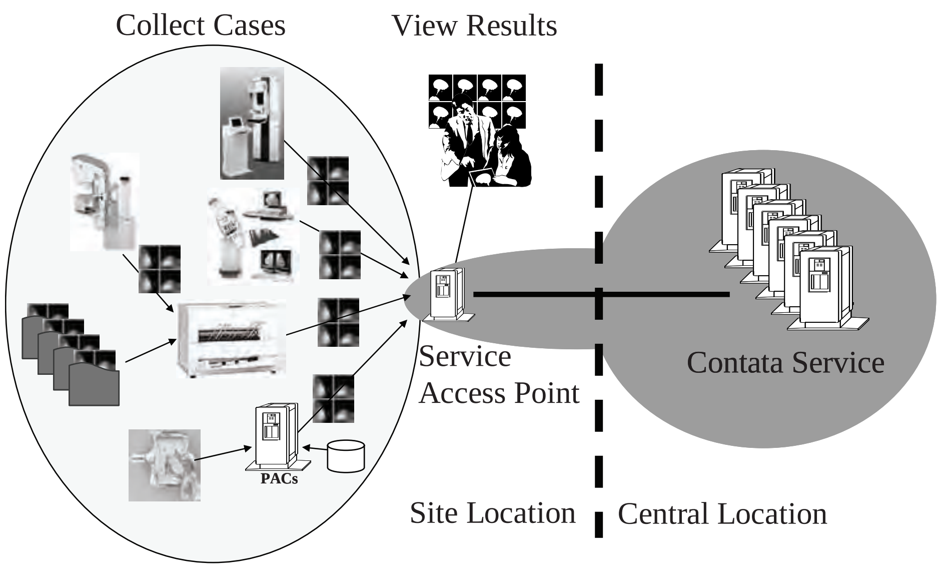
The use of computer-aided detection (CAD) techniques in mammography can mitigate the growing shortage of radiologists, as well as reduce or eliminate many of the instances of missed diagnosis. One of the authors of this article and several of his colleagues have developed new computer-based algorithms that allow a rapid analysis leading to the marking of cancerous and pre-cancerous regions, thereby providing a decision-support diagnostic facility to the radiologist. Using a CAD-based approach in conjunction with a human radiologist allows for the second reading of a mammogram, with the human radiologist actively involved in the process and making the final determination in each case. The advantages of the proposed approach are:

- The capital investment of using a CAD service is significantly less, when compared to that of employing a second radiologist.
- No additional hardware or space requirements are required.
- Current and previous cases can be made available to the radiologist online for necessary comparison.
- There is a minimal footprint in the workspace, allowing multiple radiologists to objectively view and analyze mammograms.
- Results and information can be made available anywhere via the Internet.
- Improvements to the algorithm and core technology can be readily disseminated.
- The approach is consistent with the trend towards teleradiology, allowing radiologists to perform analysis from anywhere, anytime.
- The proposed approach reduces the incidence of second visits and the level of patient anxiety by providing expert (specialty) second opinions when needed in a timely manner through a teleradiology model.

The proposed approach was developed in 2001 and 2002 (Gupta, Norman, Mehta, & Benghiat, 2002), and its technology component is described in Norman and Gupta (2002), which advocated that instead of taking the mammogram image and its interpretation by the radiologist at the same location, it would be more advantageous to use a geographically decentralized strategy that utilized the following principles:

- taking of the mammogram image by a technician at the location where the patient was available;
- transmitting the image to a central location where the image was analyzed by advanced data mining techniques and compared with other images;
- interpretation by a qualified radiologist at the same location where the centralized computing facility is located or at a different place altogether; and
- making the image available, with appropriate privacy and security safeguards, to the patient when she visits other clinics and to appropriate medical personnel with the patient's consent.

Figure 1. The integrated architecture for the use of on-site medical personnel, off-site medical personnel, and computer-based techniques for mammography and other medical applications



Based on unique and patentable distributed computing technology, a peer-to-peer model was implemented; it could act as a pure application service provider (ASP), a pure client application, or any combination in between. The proposed solution could:

- minimize or eliminate the costs of a "second reader";
- provide scalability for large, medium, and small centers (national and international);
- support the development of distributed teleradiology systems;
- decrease the liability factor for false readings; and
- improve equity of access to radiology services by employing only technicians on-site and performing both human and CAD readings off-site.

The proposed technology envisaged a long-term vision of "Image Anywhere." where there is a network of mammography screening centers in shopping malls across the country. A woman could step into a mammography center as easily as walking into a drug store. While the screening was performed at that site, the radiologist could be located hundreds of miles away looking at images from several mammography centers (via teleradiology) and could provide an opinion back to the concerned location within a few minutes of the screening time.

Finally, the proposed architecture and the technology could be readily adopted for use with all other types of medical images, as depicted in Figure 1. After establishing a track record with mammography CAD, the plan envisaged expansion into auxiliary applications (first medical, then non-medical). These include dentistry, CT colonography, and bone cancer detection.

The above concept was unanimously voted to be the first-place winner at the Big Red Venture Fund Innovation Contest, organized on an annual basis by Cornell University (2002). To the best of the authors' knowledge, this is the pioneer instance where the three-pronged strategy of using resident medical resources, off-site medical resources, and advanced computational techniques was explicitly delineated; this was done in the context of optimizing the productivity of radiologists, enhancing access to mammography centers by women, improving the quality of interpretation of mammograms, reducing the incidence of errors (both false positives and false negatives), and reducing the costs incurred in performing mammograms.

MANAGEMENT OF HEALTHCARE RECORDS

As healthcare costs continue to rise, researchers are exploring new options for enhancing the process of sharing medical data across disparate information systems, both within and across hospitals and other healthcare facilities. This has the potential to reduce costs by billions of dollars each year, estimated at $77.8 billion for the United States alone (Walker et al., 2005), and concurrently improve the quality of healthcare rendered to patients. Each of the entities in the current generation of hospital systems was built to function on an individual basis, with each island of information governed by its own idiosyncratic data model (Shortliffe, 1998). In the U.S., only regional interoperability has been implemented, so far, on an experimental basis (Halamka et al., 2005). The absence of a larger-scale interoperable system presents other problems too. For example, in the case of a plane crash or other catastrophic situation, the manual approach to accessing large numbers of patient records is very weak (Teich, Wagner, Mackenzie, & Schafer, 2002). There are several other facts that further strengthen the requirement for interoperable healthcare systems, such as:

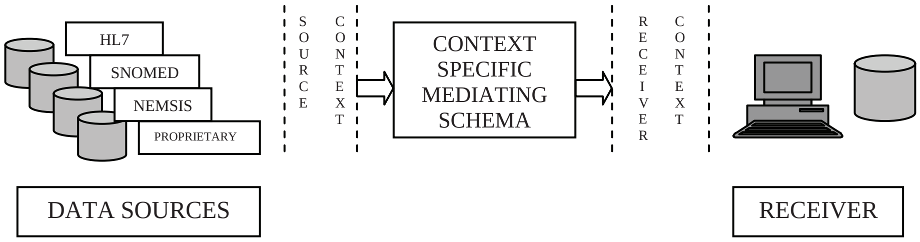
1. There is a large number of cases where institutions have split or merged, and existing data are physically distributed.
2. There are several 'mobile' individuals who frequently use medical facilities in different states or even countries.
3. The aged population contribute significantly to overall healthcare costs and very often have limited mobility.

This increases the requirement of access to healthcare records from a previous residence or transmittal of medical data to remote facilities for diagnosis rather than to move the patients themselves.

The challenges inherent in transforming disparate islands of data into an archipelago of integrated information have been highlighted by Gupta (1988) and others (Reddy, Prasad, Reddy, & Gupta, 1994; Arellano & Weber, 1998; Arts, Keizer, & Scheffer, 2002). The constituent systems differ in terms of data types, data definitions, data structures, data hierarchies, data categorizations, and underlying assumptions that are not expressly denoted in the concerned information systems. Imagine that you receive an electronic medical record of a new patient. The weight of the person as shown in the corresponding database field is 75. If you are a medical professional in the United States, your immediate reaction is that the person is extremely underweight, based on the assumption that weight is specified in pounds, whereas it is actually in kilograms. The U.S. is among the handful of countries that still use the traditional British system for most measurements; another county in this category are in Africa. Britain moved to the metric system a few decades ago. Ironically, the healthcare arena is the only one where the metric system has been adopted in the U.S. for measurement of mass and volume. However, many other types of differences continue to remain in terms of underlying assumptions of data types and other parameters; such assumptions are not revealed by looking at the data alone.

The above example highlights that healthcare data must often be converted from one form into another to facilitate communication, either at the

Figure 2. The use of the mediation approach to reduce the effort involved in integration data from heterogeneous information systems (adapted from Sarnikar & Gupta, 2007)



source or at the destination. The process of conversion of data, manually or by computer, involves significant costs and is prone to the loss of information (Barthell, Coonan, Pollock, & Cochrane, 2004; Shapiro et al., 2006). Since many healthcare applications require access to each other's data, and neither the source nor the destination is willing to do the required conversions, the only possible solution is to transform the data en route from the source to the destination.

The format and other details related to the data at the source are maintained in the source schema. The target schema contains the same types of underlying information for the destination. As the number of potential sources and potential targets increase, the number of likely transformations increases in a non-linear fashion. Haas, Miller, Niswonger, Roth, and Wimmers (1997), Milo and Zohar (1998), Abiteboul, Cluet, and Milo (2002), and Shaker, Mork, Barclay, and Tarczy-Hornoch (2002) advocated the use of middleware and the use of one common schema that incorporates data elements from multiple client schemas. The complexity of this common schema increases with the number of sources and targets, thereby restricting the use of this approach in large, diverse healthcare applications.

Wiederhold and Genesereth (1997) highlighted the fact that a single mediating schema in a large domain such as healthcare is not feasible; they advocated that the implementation of domain-specific data standards and mediators across heterogeneous medical information systems results in a cost decrease for mediating the transfer of data. Despite these findings, there is minimal standardization in the creation of healthcare information systems in the United States and in many other countries. This means, for example, that if there are "m" ambulance systems and "n" emergency department systems that could potentially need the medical data from any of the "m" ambulance systems, one could be dealing with as many as "m multiplied by n" transformations of data. Given the complexity of dealing with such a large number of possibilities, the process of data transformation is clumsy, time consuming, and costly (Gupta, 1998).

One approach to address the above problem of non-linearity is to create a mediating schema as the framework for reconciling heterogeneous information systems within and across hospitals and other healthcare facilities. Using the approach described in several papers, including Reddy and Gupta (1995), the number of necessary transformations can be reduced from "m x n" to "m + n". This requires that the mediating schema contain a core set of context-specific patient care-related information and a comprehensive methodology for specifying the ontology (vocabulary) of the relevant healthcare domain. The lattice-based context interchange approach, described by Reddy and Gupta (1995), allows evolutions of the semantics of data in the source or target schemas to be managed in a more effective manner as compared to traditional approaches. Differences in ontology always exist in large information systems (Wiederhold & Genesereth, 1997). Within a single hospital, each healthcare unit may collect, store, and process its own set of data, based on its own specialty and needs. The application of the mediating schema approach requires the careful study of the different ontologies, and the formulation of rules to transform data from one particular ontology to another.

Sarnikar and Gupta (2007) describe the problems and challenges involved in creating a mediating schema that would transfer data between the heterogeneous databases of a particular pre-hospital and ED system in the Boston area, and in extending the prototype system to cater to the idiosyncrasies of additional medical systems (see Figure 2). Based on the assessment of the size and complexity of the source and destination databases, as well as the associated data dictionaries, it was concluded that the mappings between single data elements needed to be done manually. This was based on the realization that the available software approaches to automate the mapping process between databases were frequently unreliable and required human intervention to analyze and to make corrections to the mappings generated. The manual creation of mappings could, however, lead to a set of development rules that could facilitate the creation of additional schemas in the future.

These tasks involved extensive consultation between members of the development team and experts from several concerned medical specialties, especially for "analysis of what information most succinctly and completely composed the patient care record across the source and target hospital information systems." The analysis of the data elements involved acquiring deep understanding of the meaning of each element in the source schema and its corresponding element in the target schema, as well as the associated differences in cardinality.

Sarnikar and Gupta (2007) presented a performance analysis of the above system in terms of information loss during the automated data translation process, relative to the coverage (amount of fields populated in the target schema using information from the source schema). Due to the presence of certain type conflicts and missing-data conflicts, a small amount of information was lost (<15%) in the presence of significant coverage values (>80%). Further, this was eliminated through the use of appropriate converters and filters (Gaynor, Gupta, Rawn, & Moulton, 2008; Gupta Martin, Avanavadi, & Sarnikar, 2007).

Industries and applications that are still at an evolutionary stage are frequently characterized by the existence of no standards or by the existence of too many standards, none of which carry broad acceptance. In the case of the healthcare arena, multiple standards currently exist, including HL7, EDIFACT, X12, ASTM, NCPDP, DICOM, and XDT (Dudeck, 1998). Further, as a reviewer of this article pointed out, the need for the transmission of medical data may be acute, sub-acute, or delayed. The electronic transmission of acute data (pre-hospital to emergency department) usually occurs in the case of healthcare providers who can use a common data set for their medical records. Sub-acute transmission could apply in situations involving hospitals and providers who use similar data elements, but not necessarily a similar platform. Delayed transmission, such as transmission of data across a national boundary, will need a standardized healthcare language, but not common data elements or platforms.

Based on the above discussion, we find that the use of cutting-edge computer and communications technology alone cannot accomplish the goal. Instead, large amounts of computational power and human expertise are needed to create the bridges across legacy healthcare information systems at this stage. Hospitals and other healthcare facilities in the U.S. have traditionally used medical and information technology personnel in the U.S. to undertake initial efforts in this area. Unfortunately, a vast majority of work remains to be done, and this cannot be accomplished, in terms of both time and costs, by the personnel available in the U.S. As such, one needs to explore non-conventional solutions that involve the use of resources from abroad.

Until the eighties and the nineties, large U.S. companies belonging to other sectors of the economy focused on doing all of their information technology work in the U.S. (Gupta, Seshasai, Mukherji, & Ganguly, 2007b). This was the traditional model, and the concerned management and technical personnel were satisfied with the pace of progress. Options for doing such work abroad, or by persons recruited from abroad for temporary work in the U.S., were frequently discarded on the basis that:

1. Only the persons currently associated with the work were familiar with the intricacies involved, thereby implying that such peculiarities were too complex or too confidential to be shared with others.
2. Organizational procedures or governmental regulations did not permit data and process information to be transmitted abroad or shown to foreign nationals.
3. The concerned work was too important for the success of the company, and the option of saving costs on this particular application was miniscule in comparison.
4. The work could not be performed by persons working in Asia because the difference in time zones made it impossible for them to interact with domain experts in the U.S.

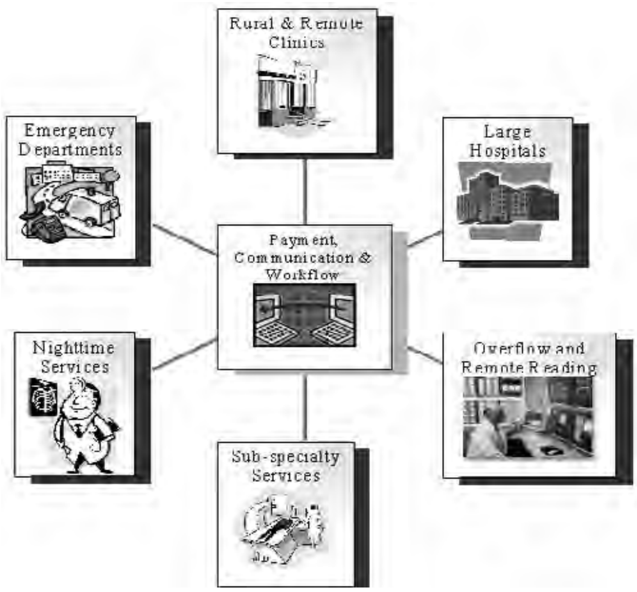
The above situation was dramatically altered by the Y2K dilemma. Management and technical personnel could no longer plead for six more months to complete a project. The date, December 31, 1999, was a very hard deadline, and the conversion of information systems had to be completed by then: no exceptions, indeed. This inflexible scenario forced the companies to become flexible. They permitted parts of the conversion work to be done abroad, as well as to employ persons with foreign qualifications and experience. The conversion from national currencies to the common currency, euro, had a similar impact in Europe. The unqualified success in both cases forced companies to depart from their arrogance in maintaining status quo.

A similar compelling need currently exists in the case of management of healthcare records. Conventional database systems are clumsy, costly, and time consuming. The integration of information in such systems, as well as the gradual incorporation of newer concepts, requires the healthcare industry to seriously consider the use of the hybrid model, involving the use of human resources in the U.S., human resources abroad, and state-of-the-art information technology. The success of the banking industry, the insurance industry, and several other industries too relies heavily on ensuring the privacy and the security of data belonging to their customers. Yes, glitches do happen from time to time. In 2004, information on 20 million credit card holders was compromised at a credit card processing facility in Tucson, Arizona. This particular facility was subsequently acquired by another company. But this unauthorized disclosure was not used as the basis to cease this type of application altogether. Instead, it should serve as the motivator for human experts, both in the U.S. and abroad, to develop more robust applications that can be used by more industries.

REMOTE DIAGNOSIS

Remote diagnosis is both an outsourcing and an insourcing phenomenon. For many years, medical records and medical images from patients and healthcare practitioners in Latin American countries, as well as from other countries in the world, have been reviewed by doctors at Massachusetts

Figure 3. A service delivery model for teleradiology (adapted from Kalyanpur, Latif, Saini, & Sarnikar, 2007)



General Hospital in Boston and by a number of other leading hospitals in the U.S. The Arizona Telemedicine Program (ATP) is a pioneer in terms of medical professionals located in Tucson, Arizona, providing expert advice to patients located elsewhere in the state, in neighboring states, and in other countries. ATP provides teleradiology, telepathology, and teleoncology services to patients in hospitals, in prisons, and in other settings (Weinstein et al., 2007). A significant number of the patients, who seek remote diagnostic advice, are from Navajo Nation and from other Native American nations in Arizona and neighboring states of the U.S.

On the outsourcing side, medical personnel in other countries are looking at medical records of patients in the U.S. The specialty that has witnessed the most attention is teleradiology (Pollack, 2003). Teleradiology involves the electronic transmission of radiological images, such as X-rays, computed tomograms, and magnetic resonance images, across geographical locations via telephone lines, satellite connections, and wide area networks. It enables a single radiologist to serve multiple hospitals concurrently, even ones in other continents. Further, it enables the image to be interpreted by an alert physician working a day shift rather than a radiologist who has been up all night (Weinger & Ancoli-Israel, 2002; Firth-Cozens & Cording, 2004).

Telemedicine is the delivery of healthcare services in situations where the physician and the patient are not at the same geographic location; it is a broad term and includes teleradiology, telepathology, and teleoncology. Telepathology involves the use of video microscopy at the patient's location and a pathologist's workstation at the physician's location. Teleoncology refers to the use of remote technology to address different aspects of cancer care.

For the purposes of studying "remote diagnosis," we will focus primarily on teleradiology for four reasons. First, the area of teleradiology has attracted wide attention in the media (e.g., Pollack, 2003). Second, this article focuses on outsourcing rather than insourcing, and the other specialties are characterized more by insourcing than by outsourcing. Third, within outsourcing, we are focusing on offshoring issues and are looking at applications that transcend national boundaries, not just local or state boundaries. Fourth, the teleradiology scenario allows an objective assessment of the potential risks and opportunities for individual radiologists in the U.S., as well as for the broader medical community in the U.S.

The growth in teleradiology is being driven by four major forces. First, there is a significant shortage of radiologists, because of a significant number of radiologists retiring from practice and training programs not keeping pace with growing demand (Sunshine, Maynard, Paros, & Forman, 2004; Bhargavan & Sunshine, 2002). Second, the aging population and the advent of newer imaging technologies are leading to annual increases in imaging volumes; for example, a 13% increase in the utilization of radiological imaging was observed among Medicare beneficiaries (Maitino, Levin, Parker, Rao, & Sunshine, 2003). Third, the increased use of imaging technologies in trauma situations has led to a corresponding need for round-the-clock radiological services in hospital emergency rooms (Spigos, Freedy, & Mueller, 1996). Fourth, changing regulations and guidelines have contributed to the need; as an example, the Health Care Financing Administration (HCFA) requires that overnight coverage be provided by certified radiologists, rather than by residents or trainees, in order to be billable (HCFA Medicare Program, 1995).

Typically, the radiology group outsources its night calls to a teleradiology provider and pays the latter for preparing the preliminary report. The insurer is billed for the final report that is prepared by the radiology group the following morning. The service delivery model depicted in Figure 3 includes mechanisms for communications, workflow, and payments.

Teleradiology offers several advantages:

1. A single professional can support multiple hospitals concurrently via teleradiology links to a central reading facility (often the radiologist's home).
2. Remote locations with radiological scanning, but no on-site radiologist, can be supported leading to improved patient care (Franken et al., 1995; Lee et al., 1998).
3. The productivity of the radiologist can be enhanced by bringing the images to the radiologist, rather than vice versa, thereby eliminating commuting time and delays.
4. The work can be optimally assigned among multiple radiologists in large hospitals.
5. Greater availability of subspecialty consultations results in better patient care (Kangarloo et al., 2000; Franken et al., 1997; Sickles et al., 2002).
6. Residents (junior doctors) covering night shifts in academic hospitals can use teleradiological services to ensure correct diagnosis and to seek confirmation.
7. The increasing disparity in the patient-to-radiologist ratio, especially during off-peak hours, can be effectively addressed by offshoring teleradiology services.

The increasing availability of technologies that replace invasive screening procedures (virtual colonoscopy replacing actual colonoscopy) means enormous increases in patient volume. Further, defensive medicine increasingly employs tests to guard against missing even the most unlikely diagnosis. The use of head CT in the ED under circumstances when the chance of an abnormal reading is extremely small is one example. It remains commonplace to get a head CT done if the patient has expected meningitis before doing a lumbar puncture, even though the literature indicates this is typically unnecessary. While this example of defensive medicine is not gratifying to mention, it represents the current reality. It also provides another reason why the workload in the radiology department has increased and will continue to increase further.

Teleradiology providers must conform to the guidelines of the Health Insurance Portability & Accountability Act (HIPAA), and are required to implement adequate privacy and security practices, as Protected Health Information (PHI) and Electronic Protected Health Information (EPHI) are transmitted over public networks on a regular basis. HIPAA requires that "covered entities execute contracts that consist of specific provisions for protection, use and disclosure of health information" (Hilger, 2004). The Privacy Rule deals with all forms of patients' protected health information, whether electronic, written, or oral, while the security rule covers only protected health information that is in electronic form, including EPHI that is created, received, maintained, or transmitted. The security rule does not prescribe any specific technologies; being technology neutral, it allows the HIPAA-covered entities to choose solutions based on their specific requirements. Technical safeguard standards include stringent guidelines for Access Control, Audit Controls, Data Integrity, Person or Entity Authentication, and Transmission Security.

The biggest hurdle to the rapid deployment of teleradiology services is the credentialing process. The U.S. requires statewide licensing requirements and board certifications; as such, a teleradiologist based in Australia must be registered to practice in all the relevant states in the U.S., so to look at radiological images from hospitals in these states, as well as pay appropriate fees to these states on an annual basis. Canada does certification at the national level for radiologists. And in Europe, some of the members of the European Union allow still greater flexibility. The current U.S. regulatory and credentialing structure was designed for a physical presence of medical professionals and needs to be adapted for the evolving technologies and procedures. Recently, some states have modified credentialing laws to allow out-of-state radiologists to perform remote diagnosis. Further, several federal and military healthcare organizations in the U.S. have licensure laws that enable them to render services independent of state and national boundaries. However, such privileges have not been extended to the private sector.

Other obstacles to the growth of teleradiology are:

1. limited availability of reliable Internet connections, especially in remote locations;
2. limited availability of trained technicians;
3. traditional billing and reimbursement procedures that vary by country and state, such as Medicare not paying for services rendered from abroad;
4. variations by nations and states, as well as underlying ambiguity, in medical malpractice liability laws (Gantt, 1999); and
5. the need for incorporating new encryption methods while transmitting image data (Cao, Huang, & Zhou, 2003).

Other related aspects of radiology have been discussed by several researchers (Bradley, 2004; Graschew, Roelofs, Rakowsky, & Schlag, 2006; Hayward & Mitchell, 2000; Jacobson & Selvin, 2000; Kalyanpur, Weinberg, Neklesa, Brink, & Forman, 2003; Kalyanpur, Neklesa, Pham, Forman, & Brink, 2004; Levy & Yu, 2006; Maitino et al., 2003; Mun, Tohme, Platenbery, & Choi, 2005; Takahashi, 2006; Weinger & Ancoli-Israel, 2002).

Over time, the concept of using both onshore and offshore radiologists will grow in terms of overall numbers of radiological images analyzed, as well as in terms of the breadth of the cases studied. Further, we expect the trends in both insourcing and outsourcing to continue. Parts of some medical diagnostic and allied applications will be performed abroad because of lower costs, quicker response, and load balancing. Conversely, more patients from abroad will seek professional advice from medical experts in the U.S. Overall, it appears appropriate to gradually lift barriers that currently impede outsourcing and insourcing activities; most of these barriers are in the former category.

MONITORING AND ENHANCEMENT OF SAFETY

Timely information on adverse drug effects can save lives and reduce healthcare costs. Previous studies show that more than two million adverse drug reactions occur yearly and are responsible for an estimated 100,000 deaths (Lazarou, Pomeranz, & Corey, 1998; Gurwitz, Field, Avorn et al., 2000; Fontanarosa, Rennie, & DeAngelis, 2004). Some systems exist for identifying drugs with serious adverse effects, but they have had limited success (Kopec, Kabir, Reinharth, Rothschild, & Castiglione, 2003; Ray & Stein, 2006). Between 1997 and 2005, the MedWatch system in the U.S. identified 15 drugs with toxic side effects, taking an average of 5.9 years for the identification phase and the subsequent drug withdrawal phase. In general, after the introduction of a drug into the market, the process of eliciting and analyzing information from patients is weak, especially when problems arise during extended use of the drug (U.S. Department of Health and Human Services, 1999; Brewer & Colditz, 1999; Okie, 2005). The problem with Vioxx, for example, was only uncovered during controlled clinical trials (Bombardier, Laine, Reicin et al., 2000). And there is still no system, either in existence or under discussion, for addressing this need at a global level.

In order to address these issues, the Institute of Medicine (2006) presented a report on "The Future of Drug Safety" and made the following recommendations:

1. increase the FDA's authority to ensure sponsor compliance with standards and regulatory requirements, especially those related to packaging and distribution;
2. establish separate performance goals for safety, in addition to existing goals for speed of approval;
3. ensure proper communication of the drug approval/testing status to consumers and medical practitioners through effective package indicators and advertisements; and
4. improve the facilities, available resources, and organization structure of the FDA.

Towards these goals, a prototype Community Pharmacy Safety Network (CPSN) was developed. Pertinent raw data are spread over the computer systems of multiple organizations including: (1) the one who performed the original drug development work, (2) the one that produced the drug, (3) the one that conducted the clinical trials, (4) the FDA in the U.S. and equivalent government agencies in other countries, and (5) the one that prescribed the medication. In addition, pharmacies contain information on the buyer of the drug, on what date, and in how much quantity. The problems involved in integrating these types of information from diverse sources were discussed in a different context earlier in this article; additional details are available in recent papers by Kalyanpur, Parsia, and Hendler (2005), Corcho and Gomez-Perez (2005), and Cristani and Cuel (2005). The problems become even more complex in the present case because of the need to access information from multiple countries and cultures.

The prototype system, including the key modules, is depicted in Figure 4. Detailed information on the system architecture and allied issues are available in Gupta, Crk, Sarnikar, and Karunakaran (2007a) and Gupta, Woosley, Crk, and Sarnikar (2007c). With proper infrastructure and incentives, pharmacists and pharmacy technicians would be designated agents for collecting raw information on the patient's medication history, including the adverse reactions experienced by the patient.

Development of the prototype system revealed several issues and opportunities. First, some drugs are given as samples by physicians to patients, and significant effort would be involved in incorporating such information into the overall system. Second, individual patients buy drugs from multiple pharmacies; in view of the current guidelines for privacy of patient records, it is very difficult to link records concerning the same patient from different pharmacies. Third, the procedures differ very significantly across countries. Fourth, analysis of the information requires the use of sophisticated data mining technology, with the help of (human) domain experts and (human) data mining experts. Such experts are already pressed for time in the U.S.; they are expensive too. It therefore seems appropriate to explore if this type of work could be done abroad. For example, one could envisage the creation of a global center of excellence for this particular field in Mexico, close to the U.S. border, so that one could benefit from the less expensive rates in Mexico in conjunction with the feasibility of experts from the U.S. visiting such a center on a frequent basis.

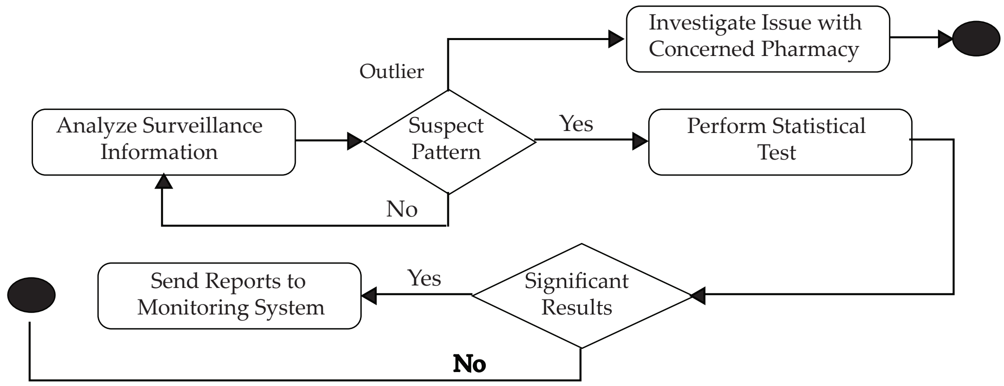
Gopal (2007) and her associates have employed an entirely different approach to assimilate information from patients, especially related to the side effects of alternative medicines. Their approach focuses on the mining of information from chat groups and other online repositories of information voluntarily provided by the patients, such as message boards, blogs, and listservs. They utilize proprietary search and aggregation techniques to distill raw data into structured information that answers critical questions. Their approach, again, utilizes a combination of computer power and human expertise.

DISSEMINATION OF THE LATEST HEALTHCARE INFORMATION

The ideas presented in this section are currently geared more towards medical education and medical research, rather than the provision of improved patient care. The current version of the prototype system enables users to search for current and previous literature, how-to articles, and other educational items. In the future, the enhanced Web site will assist healthcare providers by providing immediate lookups for symptoms and diagnosis. This would not only aid practitioners by giving them access to the latest medication data available, but also reduce treatment times. Further, based in part on a suggestion from a reviewer of this article, access to a human expert will be provided for the purpose of providing additional advice and support on specific cases.

Medical information has been maintained in books, journals, and specialty magazines. Now, a growing number of people turn to the Internet to retrieve healthcare-related information, and they do so from a variety of sources, most of which are run by commercial entities. The next area of growth will be sites that focus on specific fields of medicine, contain data culled from scholarly publications, and are operated by eminent specialists in the field. One such site is being developed for the field of gastrointestinal motility; it builds upon the concept of existing healthcare information sites with the intention of serving the diverse needs of laypeople, medical students, and experts in the area. The site, called Gastrointestinal Motility Online, leverages the strengths of online textbooks, which have a high degree of organization, in conjunction with the strengths of journal collections, which are more comprehensive, to produce a knowledge base that can be easily updated, is comprehensive, and can provide accurate and high-quality information to users. In addition to implementing existing Web technologies such as Wiki- and Amazon-style commenting options, Gastrointestinal Motility Online uses automatic methods to collect information from various heterogeneous data sources to create coherent, cogent, and current information for the diverse base of users.

Figure 4. Surveillance process for medical drugs (adapted from Gupta et al., 2007a)

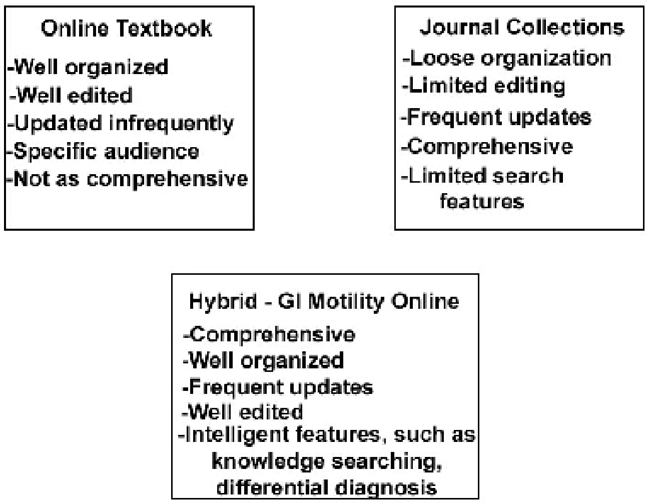


Gastrointestinal motility is a very small part of the vast field of medicine. Books such as *Harrison's Principles of Internal Medicine* serve as the main source of information in the field of medicine including gastrointestinal motility (Kasper et al., 2005). Most of these books now also have electronic versions. More recently, electronic texts sites such as Wiki, eMed, WebMD (http://www.webmd.com/), and UpToDate (http://www.uptodate.com/) have evolved and are rapidly gaining popularity. Because these works are designed to provide a broad overview of the field of medicine, they only provide a very superficial treatment to a topic such as gastrointestinal motility. More detailed information about gastrointestinal motility disorders may be found in medical journals. Many journals that were previously in paper format now come out both in paper and electronic formats. Searchable electronic archives, such as PubMed (http://www.pubmedcentral.nih.gov/), now place a plethora of information into the hands of researchers and physicians. However, such searches are very time consuming and inconvenient. Sites like AccessMedicine (http://www.accessmedicine.com/) serve as search engines that attempt to place the most suitable information on a medical topic in a user's hand.

As shown in Figure 5, Gastrointestinal Motility Online is a hybrid between standard textbooks and review articles (Au & Gupta, forthcoming). It seeks to centralize and present the information in a scalable manner that is customized to the user's information requirements. The user base includes: laypersons, patients, medical students, biomedical scientists, physiologists, pathologists, pharmacologists, biomedical students, researchers, pharmaceutical personnel, house staff, specialty fellows, internists, surgeons, and gastroenterologists.

Creation of the gastrointestinal motility knowledge repository started with calls to key gastrointestinal experts inviting them to submit a chapter, in electronic form, for inclusion in this knowledge repository. The titles and themes of these chapters were determined through discussions involving the concerned authors and the editors for this project (Dr. Goyal and Dr. Shaker). The inputs from the contributing authors were reviewed by the editors and by others. Under the aegis of an unrestricted grant from Novartis Corporation, the two editors worked closely with the staff of Nature Publishing Group on a number of tasks that ultimately led to the creation of the Web site, http://gimotilityonline.com. This site may serve as a model for user-driven, scalable, and interactive medical information at a single site for other specific medical topics. These sites

Figure 5. Gastrointestinal motility online is a hybrid of online textbooks and journals



can then be interlinked to cover the broad field of medicine. Since GI Motility Online is a hybrid that includes textbook-type material as well as authoritative up-to-date reviews and interactive features, it poses many new problems and opportunities. Currently, all updates must be initiated manually. Automated content generation or extraction from other publications is not feasible, due to the stringent need to maintain relevance and quality. This applies to addition of new material, editing of existing material, and deletion of parts of existing material as new test results become available.

In the next part of the endeavor, the goal is to enable machine-assisted updating of the material in the gastrointestinal knowledge repository. Given the large number of articles published weekly and the difficulty in ascertaining relevance and quality, a number of automated tools will be used to optimize the updating process. One technique that will assist in the maintenance and updating of the site is presented in Sarnikar, Zhao, and Gupta (2005). This method selects articles ranked by relevance using a combination of both rule-based and content-based methods, using the following principles:

1. Profiles are modeled in the form of rules.
2. The purpose of rule-based profile use is to identify a subset of documents of interest to a user.
3. Each role has a set of predefined rules associated with it.
4. Rules specify knowledge sources to access (for example, nursing journals for nurses).
5. Rules can specify knowledge depth and knowledge breadth.
6. Rules can specify semantic types of primary importance to roles.

Profiles are used in the gastrointestinal motility context to separate information into categories—for example, new clinical findings vs. basic science. Articles could be assigned a category and a weight, given categorization rules based on Unified Medical Language System (UMLS) synonym lists and the categories sign or symptom, diagnostic procedure, therapeutic or preventive procedure, and disease or syndrome semantic types. These tools can form the basis for an RSS XML news feed or to efficiently assemble relevant articles for use by the editors or Web site administrators. While these tools will aid the editor, there is no replacement for the role of humans in selecting and classifying information.

Ontologies and semantic networks are necessary prerequisites to the development of and classification of information repositories. Ontologies serve many purposes including to: reuse and share domain knowledge, establish classification schemes and structure, and make assumptions explicit. They also allow analysis of information and complement the stricter terminology that is used in straightforward text searches, with or without synonyms. Examples of ontologies in use today include the National Library of Medicine's Medical Subject Heading (MeSH), disease-specific terminologies such as the National Cancer Institute's PDQ vocabulary, drug terminologies such as the National Drug Data File, and medical sociality vocabularies such as the Classification of Nursing Diagnoses and the Current Dental Terminology. In Gastrointestinal Motility Online, the ontological hierarchy will be used to make distinctions between parts of the gastrointestinal tract and the different sections of the stomach and the esophagus (Kumar, 2005).

One of the keys to developing an automated system is the set of ontologies presented in the UMLS semantic network that relies on the concepts built into the UMLS concept hierarchy. An overview of the UMLS is available at http://www.nlm.nih.gov/research/umls/presentations/2004-medinfo_tut.pdf.

THE UNIFIED MEDICAL LANGUAGE SYSTEM: WHAT IS IT AND HOW TO USE IT?

UMLS is an aggregate of more than 134 source vocabularies, including the classifications from such lists as ICD-10 and DSM: IIIR-IV. It represents a hierarchy of medical phrases that can be used to classify articles and textbook entries.

The system described in Sharma (2005) uses techniques of Natural Language Processing (NLP) to construct a semantic understanding that goes beyond text searching. Using the ATIMED (Automated Integration of Text Documents in the Medical Domain) system, the content and order of phrases are related lexically using a concept called Word-Net. Word-Net operates on the verbs, subjects, and objects of the sentences, comparing sets and subsets of subject-verb-object collections in order to determine relatedness to a desired topic. Sharma (2005) uses the following two sentences as examples: Dysphagia is a disease and defined as a sensation of sticking or obstruction of the passage of food. Dysphagia is related to obstruction of passage of food. Since both sentences use similar objects and subjects, and use the verb "is," the sentences would be deemed similar. However, the phrase "Dysphagia relates to obstruction of passage of food" would not result in potential match because the action verb is not similar (Sharma, 2005).

Developing a schema to accurately represent journal abstracts and determine the relevance of those abstracts is one method of exchanging contexts. Innovation in this domain allows Gastrointestinal Motility Online to maintain updated, consistent, quality references without requiring an editor to read all journal articles published immediately. Knowledge-mining tools are being developed to utilize this information as it becomes available to add fast, relevant access and other utility to the information repository. Advanced technologies to aid in the conversion and integration of articles and research into the mainstream science are being integrated and look to impact the breadth and speed of knowledge-base upgrades. These activities involve the use of medical and computer professionals, both in the U.S. and abroad.

The base site is hosted by Nature Publishing Group. As this site was being developed, it was found that commercial tools were available to handle the production of electronic journals and static textbook efforts like AccessMedicine (http://www.accessmedicine.com/) and WebMD. The gastrointestinal knowledge repository falls somewhere in between these two cases; accordingly, few off-the-shelf tools and algorithms were available for immediate use. As such, a significant fraction of the necessary material had to be generated and refined through experimentation.

An interesting addition to the above system is the addition of a "Gastrointestinal Guru," based in part on a suggestion from one of the reviewers. In specific cases where a person needs access to a human expert, the system will facilitate access to such experts who can provide support on an instantaneous basis. The situation is somewhat similar to the one we experience when we try to make a travel reservation online, experience difficulty, and feel relieved when we are able to connect to a human being, either by phone or online. However, the level of expertise and the degree of structure with respect to the desired knowledge are vastly different in the two cases. This difference is highlighted in Figure 6 (later in this article) and its accompanying text.

MULTI-PRONGED APPROACH

While the five scenarios discussed above are drawn from different aspects of medical practice, they have several aspects in common. In all cases, the advent of new information technologies is making a major impact on how the particular task or medical specialty is performed (Siau, 2003; Wachter, 2006). Further, in all cases, automation applies only to part of the effort. Human beings still need to be involved, though to varying degrees in the five examples considered. In some cases, a significant part of the work needs to be performed in very close proximity to the patient, whereas in others, the concept of remote tasks can be applied to a large extent. Finally, in some cases, the off-site work is medical in nature, whereas in others it is largely non-medical in character. While similar approaches have been followed in the healthcare industry to some extent in the past, the relevance and practical importance of this model in today's scenarios is more significant. This is happening for several reasons including the following:

- The technology that is available now did not exist earlier. A growing number of previously underserved, remote locations are increasingly able to access medical services through the Internet; and PACS systems are improving rapidly; data and image transfer methodologies are becoming less expensive, more effective, and more error-free. This is resulting in a corresponding need for embarking on international telemedicine endeavors, instead of regional collaborations.
- There is a growing awareness of the advantages of having interoperable health information systems, especially in addressing mass casualty situations, where fast and timely diagnosis, via telemedicine, attains paramount importance.
- The increasing incidence of persons moving across national borders for work or other reasons, as well as the continued trend towards globalization, are making national boundaries lose their traditionally strong importance. As such, one needs the ability to ensure the safety of citizens when they are traveling in foreign countries and taking drugs that were prescribed for them by doctors in the U.S., and to support allied functions. For example, the creation of new global drug efficacy monitoring systems could enable Americans with the same ethnicity or ancestral homeland to be grouped by predictable responses to drugs

based on findings by researchers in their native countries.

We consider in the following paragraphs the significance of the diverse examples highlighted throughout this article.

In the case of mammography, a technician needs to attend to the patient in order to take the mammogram; the doctor can be off-site, either in the same country or a different one; and the computer-based data mining algorithms will be executed on an off-site basis too. This is perhaps the earliest example in healthcare where the concept of doing part of the work on-site, part off-site, with the computer providing active decision-support capability, was mooted as the most accurate and the most cost-effective way.

In the case of integration of medical records in heterogeneous systems, either for emergency needs or for routine needs, computer-based techniques can be of significant help, but human experts are needed at the initial stage as well as on a continuing basis. Integration of major systems in other fields, such as logistics and manufacturing, is being increasingly done on an offshoring basis. Banks and financial institutions located in the U.S., Switzerland, and other countries were very reluctant in the eighties and nineties to let such work be performed abroad; by the late nineties, many of them accepted the offshoring concept, after they were satisfied that the evolving security and privacy protocols were adequate for their needs. The same is expected to occur for the integration of heterogeneous hospital information systems; we will consider the legal and regulatory aspects later in this article. In addition to experts in information technology, medical personnel from multiple subspecialties will need to be involved, both on-site and off-site, in order to integrate the concerned systems based on the specific characteristics of the concerned medical institutions.

In the current versions of teleradiology, the technician can be in the U.S. or another developed country, and the radiologist can be in India or another less expensive country; in addition, there is a second radiologist in the same state (or nation) as the patient who issues and signs off on the final report. Technology is used almost entirely for transmitting the image from the developed country to the less expensive one. Over time, technology could be used to partially analyze the images, to compare the image to an earlier one involving the same part of the body of the same patient, and even to compare the image with images of the same part of the body of other patients, in order to make a diagnosis and to evaluate how the symptoms may change over time.

In the case of monitoring for adverse drug effects, the MedWatch system needs to be augmented. Some of the new tasks need to be performed by pharmacists in the U.S. Other tasks, such as reconciliation of duplicate records, need to be performed using inexpensive manpower, wherever available, with support from computer-based techniques. Further, individuals in the U.S. are increasingly obtaining less expensive equivalent drugs from Canada and other countries. Such drugs must also be taken into account while designing any comprehensive system for monitoring adverse drug effects. Ultimately, this system will need to be implemented across multiple systems, as a harbinger of a global system. A growing percentage of clinical trials are now being conducted in India and other countries because of lower costs, availability of drug-naïve persons (individuals who had not taken other drugs in the past for addressing the same disease), and access to persons with different heritages. The ability to conduct these clinical trials at lower costs increases the probability that a drug company would decide to take a potential new drug from the lab to the clinical trial stage. This can also reduce delay in the launch of the new drug. However, this also implies transforming local systems for monitoring adverse drug effects into international ones. In such a case, both medical personnel and other personnel need to work from multiple countries.

In the case of dissemination of healthcare information, the initial endeavor focused on getting reputed experts from multiple countries to contribute material for inclusion in the evolving knowledge repository. The idea was to gradually support automated updates to such material on a continuing basis. So, if new results from a trusted clinical study became available, such results should be incorporated into the appropriate chapters. While part of this work can be done using computer-based techniques, the experience of the development team is that high quality and accuracy will only be accomplished if the suggested edits and updates were reviewed and approved by human experts. The use of domain experts and editorial personnel located in less expensive countries is more appealing, as it makes the overall endeavor more viable (see Table 1).

While improvements in Internet communication have made the multi-pronged approach described above possible, the lack of standardization in messaging protocols is a roadblock to the creation of a global healthcare model. We now describe how standards can be applied effectively to each of the five operational scenarios discussed, to reduce costs and improve clinical outcomes:

1. In the area of CAD mammography and teleradiology, the communication of radiological images and other data clearly depends on the standards used for creating, maintaining, and exchanging medical images (such as PACS and DICOM).
2. The problem on data exchange and integration from heterogeneous data sources can be solved by the effective use of standards such as HL7. Though the current scenario does not easily allow the adoption of a single data standard among all hospital information systems nationwide, using a standard message development framework as a mediating schema could eliminate some of the problems and make the system more portable and scalable.
3. The area of drug monitoring and post-market surveillance could also benefit from the use of standards. Using standards to create and store medication and adverse effects reports and experimental results sent from various participating agencies would create a much richer database allowing for better analysis and quicker response times.
4. Finally, while considering the issue of dissemination of medical information, the use of standards (not medical standards, though) for storing and displaying the information to the end users can increase the productivity and usefulness of search portals.

24-HOUR KNOWLEDGE FACTORY

Earlier in this article, we discussed the scenario of seeking immediate assistance from an expert in a particular specialty; the geographic location of that specialist was totally irrelevant. The same is true of situations related to some tasks related to mammography and radiology. Finally, in the other two examples of the integration of medical information systems and the creation of global drug monitoring systems, it would be appropriate for the concerned pieces of work to proceed on a continuous basis, around the clock. In all such scenarios, the paradigm of a 24-Hour Knowledge Factory bears relevance.

The University of Arizona has signed a three-party collaborative agreement with the Wroclaw University of Technology in Poland and the University of Technology located in Sydney, Australia. Under the aegis of this agreement, researchers at the University of Arizona can work from 9:00 a.m. to 5:00 p.m., Arizona time. At around 5:00 p.m., Arizona time, the research-in-progress can be transferred to fellow scientists at the University of Technology in Australia, who can work from around 9:00 a.m. to 5:00 p.m., Sydney time. At the end of the "research shift" in Sydney, the

Table 1.

Scenario	Tasks by Medical Personnel		Tasks by Non-Medical Personnel		Tasks by Computer-Assisted Techniques
	On-Site	Off-Site	On-Site	Off-Site	
CAD Mammography	X	X			X
Integration of Heterogeneous Healthcare Data Sources	X	X	X	X	X
Teleradiology	X	X			X
Monitoring of Adverse Drug Effects	X	X		X	X
Dissemination of Healthcare Information	X	X		X	X

professional work can be transferred to the Wroclaw University of Technology in Poland where researchers can conduct incremental activities over the next period of approximately eight hours, and can then transfer the evolving endeavor to the first set of researchers in the U.S. This process is akin to the passing of a baton in a relay race, with the notable difference that the baton is returned back to each participant exactly 16 hours after that participant transfers it to a colleague located on a different continent. We believe that this model will be gradually adopted by the healthcare industry, and further analysis of the historical and structural aspects will determine what subset of healthcare tasks can benefit most from the adoption of this evolving paradigm.

In general, the 24-Hour Knowledge Factory paradigm is appropriate for situations where the healthcare endeavor can be broken down into components, the underlying knowledge can be digitized, different individuals can potentially work on such components with minimal support from their peers, and the work-in-progress can be transferred at minimal cost from one collaborating center to another.

INTELLECTUAL PROPERTY AND LEGAL ISSUES

The performance of medical tasks in a collaborative fashion, on a regular basis, by individuals located across state and national boundaries, raises new issues. Who owns the intellectual property such as patents on new medical or drug inventions? Who can be sued for medical malpractice, and under which set of laws and regulations? How should the charges for medical services be apportioned? What are the corresponding avenues for seeking reimbursements from insurance companies? And what are the mechanisms for seeking redress if and when it becomes necessary? Besides these, there are also other related social and policy-related concerns such as quality control, and assurance and intensity of workflow across boundaries.

In the case of the United States, the FDA plays the dominant role at the national level on issues related to drugs. However, medical professional credentialing and registration are done almost entirely at the state level. As mentioned in the section on radiology, the radiologist can render an initial opinion from outside the particular state or country, but the final opinion is still issued by another radiologist who resides within the particular state and is licensed to practice there. The use of two radiologists, though providing quicker action, increases overall costs.

In the case of patents and intellectual property, there is a common feeling among patent holders around the world that others are exploiting your work. In the U.S., there is a feeling that companies in foreign countries are exploiting U.S. inventions and patents without authorization and payment of royalties. On the other side, there are people in India and China that feel the same way; there are instances of patents issued by the U.S. patent office on items of indigenous nature that have existed for thousands of years. This is somewhat akin to the situation where different states in the U.S. would render conflicting decisions on the same case, creating confusion. For example, until the 1970s, child custody cases were handled entirely at the state level. So in a case of divorce involving two parents residing in two different states, the first state might well give the custody of the child to the father (who resided in that state), and the second state would likely give custody of the same child to the mother (who resided in the latter state). Finally, in 1992 the Uniform Interstate Family Support Act (UIFSA) was drafted., According to this act, states have the power to reach beyond their borders for the establishment and enforcement of support orders. A similar type of action is now warranted in the healthcare domain.

Let us analyze the issue based on how laws, regulations, and norms have evolved over history. Three thousand years ago, all rules were at the village level. The village was the basis of the economy. If a person did something undesirable, the person could be ostracized from using the village well. Without water, the person could not survive. Therefore, the person had to plead with his or her peers. This was one of the mechanisms, then prevalent, to enforce conformity with the norms and mores of that era. As time progressed, the size of the geographic unit increased. In England, one saw the advent of the concept of the manor, typically a collection of a dozen villages, that functioned as a unit for economic and security purposes. The manor was replaced by still larger entities in the form of principalities, which were ultimately replaced by nation states. The lawmaking and enforcement evolved too, sometimes with overlapping provisions. For example, a person residing in Tucson today may be governed by up to four sets of regulations, of the City of Tucson, Pima County, the State of Arizona, and the United States of America, respectively. (In some areas, such as intellectual property, there may be a supranational layer as well, that is, the WTO's Trade Related Intellectual Property Agreement, which although not applied directly in the United States affects U.S. intellectual property law). Being governed by laws of the U.S. does not imply having to go to Washington, DC, to seek redress; benches of U.S. federal courts exist in most large and medium-sized cities in the United States.

Similarly, in the case of healthcare, the ultimate solution may be an international regulatory system that maintains offices in large cities around the world. This organization could deal with issues related to performing healthcare work across national boundaries. This could include credentialing, registration, medical malpractice, medical accounting, and reimbursement. Such an international regulatory system could be operated under the aegis of the World Health Organization.

Issues of trade and intellectual property are currently coordinated at the international level by the World Trade Organization (WTO) and the World Intellectual Property Organization, respectively. These organizations could serve as the nucleus for establishing streamlined mechanisms that would enable better coordination of emerging types of practices, in healthcare and in other disciplines, perhaps under the aegis of the WTO's General Agreement on Trade in Services. The Agreement on Trade Related Aspects of Intellectual Property is another mechanism of IP protection. Healthcare services will increasingly transcend national boundaries as efforts are made to perform them with speed, efficiency, and in the most cost-effective manner.

CONCLUSION

The traditional model of healthcare required medical personnel to be in immediate proximity to patients being attended to. This model will gradually be replaced, for a growing number of healthcare applications, into a three-faceted model that requires: some personnel to be on-site, other personnel to be off-site, and the use of evolving technologies to render support in a manner that is beyond the capabilities of the best medical personnel available anywhere in the world. Computers can look at millions of images of mammograms in very short periods of time to locate ones that match the key characteristics of the one currently in the clinic; such power is clearly beyond the capability of a single doctor or even groups of doctors. Off-site personnel can be located in the same state (such as physicians and surgeons of the Arizona Telemedicine program assisting doctors in clinics in Navajo Nation and in other Native American nations), in a different state or a different country. If the support is being provided from a different country, it could be in the same time zone (to provide good overlap) or in a different time zone (to provide complementary advice, especially advice from specialists, during the night in the patient's country). The time difference was initially perceived to be a hindrance; today, it is considered an asset, as it enables better usage of medical and other personnel in both countries. Initially, outsourcing will be embraced using the notion of two collaborating groups that are 10-12 hours apart in terms of time. Gradually, the model of three collaborating centers will be embraced. The use of this 24-Hour Knowledge Factory paradigm allows three centers located in three countries to continue work on a round-the-clock basis, with all the tasks being performed primarily during the day in the respective countries. New international systems must evolve to address the intellectual property, legal, accounting, and other issues related to the various forms of outsourcing.

Medicine is geared to assist mankind as a whole. The offshoring of medical services will benefit developed countries because it can lower overall costs, provide quicker response, and facilitate load balancing. Such offshoring will be advantageous to developing nations because it can widen the range of available medical expertise and enhance the knowledge of healthcare professionals in developing countries. At the same time, one must be conscious of the fact that there is a shortage of medical professionals both in developed and developing countries, and the diversion of such resources to address the needs of foreign patients can potentially aggravate the shortage in their respective home countries. These issues will be partially resolved by market forces. Over time, we will witness more cooperative endeavors involving on-site and off-site activities in the healthcare arena.

ACKNOWLEDGMENT

The authors acknowledge, with sincere thanks, valuable information, comments, suggestions, and assistance provided by many persons, including Shiu-chung Au, Harvey Meislin, Elizabeth Krupinski, Kurt Denninghoff, Rick McNeely, Richard Martin, Surendra Sarnikar, Yan An, Shawna Sando, David Branson Smith, Kit Cheong, Georgina Apresa, and Zak Campbell. Their insights have contributed to enhance the breadth and depth of this article.

REFERENCES

Abiteboul, S., Cluet, S., & Milo, T. (2002). Correspondence and translation for heterogeneous data. *Theoretical Computer Science, 275,* 179-213.

Arellano, M.G., & Weber, G.I. (1998). Issues in identification and linkage of patient records across an integrated delivery system. *Journal of Healthcare Information Management, 12*(3), 43-52.

Arts, D.G.T., Keizer, N.F.D., & Scheffer, G.J. (2002). Defining and improving data quality in medical registries: A literature review, case study, and generic framework. *Journal of the American Medical Informatics Association, 9,* 600-611.

Au, S.C. and Gupta, A. (forthcoming). Gastrointestinal motility on-line educational endeavor. *International Journal of Healthcare Information Systems and Informatics. 3*(1).

Barthell, E.N., Coonan, K., Pollock, D., & Cochrane, D. (2004). Disparate systems, disparate data: Integration, interfaces, and standards in emergency medicine information technology. *Academic Emergency Medicine, 11,* 1142-1148.

Berlin, L. (2005). Errors of omission. *American Journal of Roentgenol*ogy, *185,* 1416-1421.

Bhargavan, M., & Sunshine, J. (2002). Workload of radiologists in the United States in 1998-1999 and trends since 1995-96. *American Journal of Roentgenology, 179,* 123-1128.

Bombardier, C., Laine, L., Reicin, A. et al. (2000). Comparison of upper gastrointestinal toxicity of Rofecoxib and Naproxen in patients with rheumatoid arthritis. *New England Journal of Medicine, 343,* 1520-1528.

Bradley, W.G. (2004). Offshore teleradiology. *Journal of the American College of Radiology,* (6), 244-248.

Cornell University. (2002). *Big red venture fund innovation contest.* Retrieved from http://www.news.cornell.edu/chronicle/02/5.2.02/VentureFund.html

Brewer, T., & Colditz, G.A. (1999). Postmarketing surveillance and adverse drug reactions: Current perspectives and future needs. *Journal of the American Medical Association, 281,* 824-829.

Cao, F., Huang, H.K., & Zhou, X.Q. (2003). Medical image security in a HIPAA mandated PACS environment. *Computerized Medical Imaging and Graphics, 27*(3), 185-196.

Corcho, O., & Gomez-Perez, A., (2005). A layered model for building ontology translation systems. *International Journal of Semantic Web and Information Systems, 1*(2), 22-48.

Cristani, M., & Cuel, R. (2005). A survey on ontology creation methodologies. *International Journal of Semantic Web and Information Systems, 1*(2), 49-69.

Dudeck, J. (1998). Aspects of implementing and harmonizing healthcare communication standards. *International Journal of Medical Informatics, 48,* 163-171.

Firth-Cozens, J., & Cording, H. (2004). What matters more in patient care? Giving doctors shorter hours of work or a good night's sleep? *Quality and Safety in Healthcare, 13,* 165-166.

Fontanarosa, P.B., Rennie, D., & DeAngelis, G.D. (2004). Postmarketing surveillance: Lack of vigilance, lack of trust. *Journal of the American Medical Association, 292,* 2647-2650.

Franken, E.A., Berbaum, K.S., Brandser, E.A., D'Alessandro, M.P., Schweiger, G.D., & Smith, W.L. (1997). Pediatric radiology at a rural hospital: Value of teleradiology and subspecialty consultation. *American Journal of Roentgenology, 168,* 1349-1352.

Franken, E.A., Berbaum, K.S., Smith, W.L., Chang, P.J., Owen, D.A., & Bergus, G.R. (1995).

Teleradiology for rural hospitals: Analysis of a field study. *Journal of Telemedicine and Telecare, 1*(4), 202-208.

Gantt, A.H. (1999). Telemedicine and healthcare e-commerce. *Maryland Bar Journal, 32*(6).

Gaynor, M., Myung, D., Gupta, A., Rawn, J. and Mouton, S. (forthcoming). Interoperability of Medical Applications and Devices. In *Proceedings of the Hawaii International Conference on Systems Sciences.* Hawaii, Jan 2008.

Ghate, S.V., Soo, M.S., Baker, J.A., Walsh, R., Gimenez, E.I., & Rosen, E.L. (2005). Comparison of recall and cancer detection rates for immediate versus batch interpretation of screening mammograms. *Radiology, 235,* 31-35.

Gilbert, F.J., Astley, S.M., McGee, M.A., Gillan, M.G.C., Boggis, C.R.M., Griffiths, P.M., & Duffy, S.W. (2006). Single reading with computer-aided detection and double reading of screening mammograms in the United Kingdom National Breast Screening Program. *Radiology, 241*(1), 47-53.

Gopal, S. (2007). Retrieved from http://chatmine.com/howitworks/index.htm

Graschew, G., Roelofs, T.A., Rakowsky, S., & Schlag, P.M. (2006). E-health and telemedicine: Digital medicine in the virtual hospital of the future. *International Journal of Computer Assisted Radiology and Surgery, 1,* 119-135.

Gupta, A. (1988) *Integration of information systems: Bridging heterogeneous databases*. IEEE Press.

Gupta, A., Crk, I., Sarnikar, S., & Karunakaran, B. (2007a). The drug effectiveness reporting and monitoring system: Discussion and prototype development. *International Journal of Technology Management,* (forthcoming).

Gupta, A., Norman, D., Mehta, V., & Benghiat, G. (2002). *Contata Health Inc. business plan*. Contata Health.

Gupta, A., Martin, R., Avanavadi, P. and Sarnikar, S. (2007). Information systems and healthcare XVII: A HL7v3-based mediating schema approach to data transfer between heterogeneous health care systems. *Communications of AIS, 19*(28).

Gupta, A., Seshasai, S., Mukherji, S., & Ganguly, A. (2007b). Offshoring: The transition from economic drivers toward strategic global partnership and 24-Hour Knowledge Factory. *Journal of Electronic Commerce in Organizations, 5*(2), 1-23.

Gupta, A., Woosley, W., Crk, I., & Sarnikar, S. (2007c). An information technology architecture for drug effectiveness reporting and post-marketing surveillance. *International Journal of Healthcare Information Systems and Informatics,* 2(3), 65-80.

Gurwitz, J.H., Field, T.S., Avorn, J. et al. (2000). Incidence and preventability of adverse drug events in nursing homes. *American Journal of Medicine, 109,* 87-94.

Haas, L., Miller, R., Niswonger, B., Roth, M., & Wimmers E. (1997). Transforming heterogeneous data with database middleware: Beyond integration. *IEEE Data Engineering Bulletin, 22*(1), 31-36.

Halamka, J., Aranow, M., Ascenzo, C., Bates, D., Debor, G., Glaser, J. et al. (2005). Healthcare IT collaboration in Massachusetts: The experience of creating regional connectivity. *Journal of American Medical Information Association, 12,* 596-601.

Hayward, T., & Mitchell, J. (2000). The cost-effectiveness of teleradiology at the Women's and Children's Hospital in Adelaide. *Journal of Telemedicine and Telecare, 6*(1), 23-25.

HCFA Medicare Program. (1995). *Revisions to payment policies and adjustments to the relative value units under the physical fee schedule* *for the calendar year 1996*. Federal Regulation 60: 63124.

Hilger, D. (2004). The HIPAA privacy rule and HR/benefits outsourcing: Does the business associate label belong on your recordkeeper? *Benefits Quarterly, 20*(1), 44.

Hristidis, V., Clarke, P.J., Prabakar, N., Deng, Y., White, J.A., & Burke, R.P. (2006, November 11). A flexible approach for electronic medical records exchange. *Proceedings of HIKM'06,* Arlington, VA.

Howe, G.R. (1998). Use of computerized record linkage in cohort studies. *Epidemiologic Reviews*, *20,* 112-121.

Institute of Medicine. (2006, September). *The future of drug safety—promoting and protecting the health of the public. The Institute of Medicine's 2006 report.* Author.

Jacobson, P., & Selvin, E. (2000). Licensing telemedicine: The need for a national system. *Telemedicine Journal and E-Health, 6*(4), 429-439.

Kalyanpur, A., Latif, F., Saini, S., & Sarnikar, S. (2007). Inter-organizational e-commerce in healthcare services: The case of global teleradiology. *Journal of Electronic Commerce in Organizations, 5*(2), 47-56.

Kalyanpur, A., Neklesa, V.P., Pham, D.T., Forman, H.P., & Brink, J.A. (2004). Implementation of an international teleradiology staffing model. *Radiology, 232,* 415-419.

Kalyanpur, A., Parsia, B., & Hendler, J. (2005). A tool for working with Web ontologies. *International Journal of Semantic Web and Information Systems, 1*(1), 36-49.

Kalyanpur, A., Weinberg, J., Neklesa, V., Brink, J.A., & Forman, H.P. (2003) Emergency radiology coverage: Technical and clinical feasibility of an international teleradiology model. *Emergency Radiology, 10,* 115-118.

Kangarloo, H., Valdez, J.A., Yao, L., Chen, S., Curran, J., Goldman, D. et al. (2000). Improving the quality of care through routine teleradiology consultation. *Academic Radiology, 7*(3), 149-155.

Kasper, D.L., Fauci, A.S., Longo, D.L., Braunwald, E., Hauser, S.L., Jameson, J.L. eds. (2005). *Harrison's principles of internal medicine* (16th ed.). New York: McGraw-Hill.

Khoo, L.A.L., Taylor, P., & Given-Wilson, R.M. (2005). Computer-aided detection in the United Kingdom National Breast Screening Programme: Prospective study. *Radiology, 237,* 444-449.

Kopec, D., Kabir, M.H., Reinharth, D., Rothschild, O., & Castiglione, J.A. (2003). Human errors in medical practice: Systematic classification and reduction with automated information systems. *Journal of Medical Systems, 27*(4), 297-313.

Lazarou, J., Pomeranz, B., & Corey, P. (1998). Incidence of adverse drug reactions in hospitalized patients. *Journal of the American Medical Association, 279,* 1200-1205.

Lee, J.K., Renner, J.B., Saunders, B.F., Stamford, P.P., Bickford, T.R., Johnston, R.E. et al. (1998). Effect of real-time teleradiology on the practice of the emergency department physician in a rural setting: Initial experience. *Academic Radiology, 5*(8), 533-538.

Levy, F., & Yu, K.H. (2006). Offshoring radiology services to India. Manuscript submitted for submission in *Journal of International Business Studies*: *Special Issue on Offshoring Administrative and Technical Work: Implication for Globalization, Corporate Strategies, and Organizational Designs.*

Maitino, A.J., Levin, D.C., Parker, L., Rao, V.M., & Sunshine, J.H. (2003). Practice patterns of radiologists and non-radiologists in utilization of noninvasive diagnostic imaging among the Medicare population 1993-1999. *Radiology, 228,* 795-801.

Milo, T., & Zohar, S. (1998). Using schema matching to simplify heterogeneous data translation. *Proceedings of the 24th International Conference on Very Large Data Bases* (pp. 122-133).

Mun, S.K., Tohme, W.G., Platenbery, R.C., & Choi, I. (2005). Teleradiology and emerging business models. *Journal of Telemedicine and Telecare, 11,* 271-275.

Norman, D., & Gupta, A. (2002). *Contata Health Inc. technology plan.* Contata Health.

Okie, S. (2005). Safety in numbers—monitoring risk in approved drugs. *New England Journal of Medicine, 352*(12), 1173-1176.

Pollack, A. (2003). Who's reading your X-ray? *New York Times,* (November 16).

Ray, W.A., & Stein, C.M. (2006). Reform of drug regulation—beyond an independent drug-safety board. *New England Journal of Medicine, 354*(2), 194-201.

Reddy, M., & Gupta, A. (1995). Context interchange: A lattice based approach. *Knowledge-Based Systems, 8*(1), 5-13.

Reddy, M.P., Prasad, B.E., Reddy, P.G., & Gupta, A. (1994). A methodology for integration of heterogeneous databases. *IEEE Transactions on Knowledge and Data Engineering, 6*(6), 920-933.

Sarnikar, S., & Gupta, A. (2007). A context-specific mediating schema approach for information exchange between heterogeneous hospital systems. *International Journal of Healthcare Technology and Management, 8*(3/4), 298-314.

Sarnikar, S., Zhao, J., & Gupta, A. (2005, August 11-14). Medical information filtering using content-based and rule-based profiles. *Proceedings of the AIS Americas Conference on Information Systems* (AMCIS 2005), Omaha, NE.

Shaker, R., Mork, P., Barclay, M., & Tarczy-Hornoch, P. (2002). A rule driven bi-directional translation system for remapping queries and result sets between a mediated schema and heterogeneous data sources. *Proceedings of the AMIA Symposium* (pp. 692-696), San Antonio, TX.

Shapiro, J.S. et al. (2006). Approaches to patient health information exchange and their impact on emergency medicine. *Annals of Emergency Medicine, 48*(4), 426-432.

Sharma, R. (2005). *Automatic integration of text documents in the medical domain.* Undergraduate Thesis, University of Arizona, USA.

Shortliffe E.H. (1998). The evolution of health-care records in the era of the Internet. *Medinfo,* 9(suppl.), 8-14

Siau, K. (2003). Health care informatics. *IEEE Transactions on Information Technology in Biomedicine, 7*(1), 1-7.

Sickles, E.A., Wolverton, D.E., & Dee, K.E. (2002). Performance parameters for screening and diagnostic mammography: Specialist and general radiologists. *Radiology, 224,* 861-869.

Skaane, P., Kshirsagar, A., Stapleton, S., Young, K., & Castellino, R.A. (2006). Effect of computer-aided detection on independent double reading of paired screen film and full-field digital screening mammograms. *American Journal of Rontgenology, 188,* 377-384.

Spigos, D., Freedy, L., & Mueller, C. (1996). 24 hour coverage by attending physicians, a new paradigm. *American Journal of Rontgenology, 167,* 1089-1090.

Sunshine, J., Maynard, C.D., Paros, J., & Forman, H.P. (2004). Update on the diagnostic radiologist shortage. *American Journal of Rontgenology, 182,* 301-305.

Takahashi, M. (2006). Contribution of teleradiology in Japanese medicine. *International Journal of Computer Assisted Radiology and Surgery*: *Special Session on Teleradiology: Debate on Teleradiology, 1,* 455-456.

Teich, J., Wagner, M., Mackenzie, C., & Schafer, K. (2002). The informatics response in disaster, terrorism, and war. *Journal of the American Medical Informatics Association*, *9*(2), 97-104

U.S. Department of Health and Human Services. (1999, May). *Managing the risk from medical product use, creating a risk management framework*. *Taskforce on Risk Management report*. Washington, DC: Author.

Wachter, R.M. (2006). The "dis-location" of U.S. medicine—the implications of medical outsourcing. *New England Journal of Medicine, 354,* 661-665.

Walker, J., Pan, E., Johnston, D., Adler-Milstein, J., Bates, D.W., & Middleton, B. (2005). The value of health care information exchange and interoperability. *Health Affairs Web Exclusive*, (January 19).

Weinger, M.B., & Ancoli-Israel, S. (2002). Sleep deprivation and clinical performance. *Journal of the American Medical Association, 287*(8), 955-957.

Weinstein, R.S., Lopez, A.M., Barker, G.P., Krupinski, E.A., Descour, M.R., Scott, K.M., Richter, L.C., Beinar, S.J., Holcomb, M.J., Bartels, P.H., McNeely, R.A., & Bhattacharyya, A.K. (2007). The innovative bundling of teleradiology, telepathology, and teleoncology services. *IBM Systems Journal*, *46*(1), 69-84.

Wiederhold, G., & Genesereth, M. (1997). The conceptual basis for mediation services. *IEEE Expert*, *12*(5), 38-47.

This work was previously published in Information Resources Management Journal, Vol. 21, Issue 1, edited by Mehdi Khosrow-Pour, pp. 1-26, copyright 2008 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 6.9
Strategic Maneuvering in Healthcare Technology Markets:
The Case of Emdeon Corporation

Kirill M. Yurov
University of Illinois at Chicago, USA

Yuliya V. Yurov
University of Illinois at Chicago, USA

Richard E. Potter
University of Illinois at Chicago, USA

ABSTRACT

Healthcare technology markets have been recently identified as potential investment targets. Having survived a major environmental shock, the dot. com bust, firms in the healthcare technology industry are presently experiencing an impressive revenue growth. In this study, we investigate the strategies of Emdeon Corporation, a healthcare technology firm whose e-business model provides clues for achieving a sustained revenue growth and profitability. We trace the current sustainability of Emdeon's e-business model to a related diversification strategy that the firm's upper management has pursued via mergers and acquisitions (M&As). We also address the motivation behind current restructuring of Emdeon's e-business model. We argue that maturation of diversified e-business models leads to the transformation of individual segments into distinct entities focusing on specific technology markets.

INTRODUCTION

A growing specialization of e-business models has offered a variety of new services that create business value for the healthcare industry (Parente, 2000; Payton, 2003; Singh, O'Donoghue & Soon, 2002). These diverse services can encompass online execution of medical transaction processing, health information retrieval, and/or online enrollment to health plans, to name a few examples. Recent publications in the trade press have reflected a growing sense of optimism on the part of investors in e-business firms that serve niche technology markets. In particular, healthcare technology markets have been praised as potential investment targets (*New York Times*, 2006; *Wall Street Journal*, 2006). Investors presently focus on funding start-up companies that provide access to health-related information as well as offer Internet-based capabilities to compare quality and outcomes of healthcare services.

The gigantic size of U.S. healthcare industry presents many opportunities for technology firms that have a potential to improve value chains. According to the National Coalition on Health Care, the total healthcare spending in the United States reached $1.9 trillion or about 16% of the gross domestic product in 2004. The fact that healthcare is a data-rich industry creates opportunities for technology firms to make health data exchanges more efficient and reliable. The other distinctive characteristic of healthcare is that it is strictly regulated. In this regard, the value proposition of technology firms is evident in their capability to decrease complexity of medical-transaction processing and reduce the number of data-related medical errors.

U.S. healthcare organizations first began to build proprietary information systems in the 1960s (Collen, 1991). At that time, hospital networks were most prolific adopters of information systems given that they possessed sufficient capital bases to do so. Only recently has it become cost-effective for physician practices to embrace medical software systems. Mass adoption of broadband Internet and decreasing costs of worldwide delivery of digital materials offer opportunities for technology firms to interconnect hospital information systems and connect to software systems of physician practices.

In addition to improved affordability of information systems and technological advancements, healthcare institutions could potentially benefit from incentives offered by the regulatory agencies to digitize the exchange of health-related information. In the mid-2000s, the U.S. government intensified efforts to raise the adoption rate of electronic health record (EHR) systems. This technology enables caregivers to collect and circulate digitized patient data across the network of authorized healthcare providers (Goldschmidt, 2005; Ford, Menachemi & Phillips, 2006). The U.S. government plans to link individual EHR systems in a centralized network, allowing access to patient data on a national basis (Office of the National Coordinator for Health Information Technology, 2005). The planned centralization of EHR systems would be a massive undertaking on the part all the stakeholders of the U.S. healthcare industry. The increased quantity of digitized patient data would fuel further demand for medical-transaction processing services. Such a scenario indicates greater revenue-growth opportunities for the healthcare technology industry. Firms that are capable of building EHR systems, processing digitized data, and facilitating health-related decision-making would benefit from proliferation and centralization of EHR systems.

In the light of these important developments in the U.S. healthcare industry, we investigate the strategies of Emdeon Corporation, a healthcare technology firm whose e-business model provides clues for achieving a sustained growth of revenues and earnings in the emerging healthcare technology industry. Whereas e-business models built on a single source of revenue are dominant in this industry, Emdeon Corporation relies on a variety of revenue streams to sustain a leading market position. This case study explores the evolution of Emdeon's e-business model from a strategic management perspective. We trace the current sustainability and profitability of Emdeon's e-business model to a related diversification strategy that the firm's upper management has pursued through mergers and acquisitions (M&As). We also address the motivation behind the current restructuring of Emdeon's e-business model. The time period for this investigation spans 1998 to 2005.

We will discuss the literature on healthcare e-business models in the second section. Research method and data sources will be addressed in the third and fourth sections, respectively. In the fifth section, we analyze the evolution of Emdeon's e-business model. This model went through a number of development phases, which were reflections of the changes in the firm's strategy. In the sixth section, we offer conclusions in the form of lessons learned from Emdeon's strategic maneuvering in the emerging healthcare technology markets. These lessons point to the factors that contributed to the firm's sustained profitability in the observed period.

LITERATURE ON HEALTHCARE E-BUSINESS MODELS

There have been a number of academic and practitioner-oriented accounts that shed light on e-business model types that have emerged in healthcare technology industry. Given that e-business in healthcare is a recent phenomenon, this literature is at the early stage of development. The authors writing on this phenomenon largely investigate how healthcare technology firms add value to medical processes. The literature on trans-industry e-business models, on the other hand, has been plentiful. Timmers (1999, p. 2) advanced his definition of business models that is applicable for electronic environments: "A business model is defined as the organization of product, service and information flows, and the sources of revenues and benefits for suppliers and customers." Given that the focus of this article is on healthcare e-business models, we will focus on the current state of research on healthcare e-business models. In addition, we will selectively address research on general e-business models that complements our discussion on healthcare e-business models.

Parente (2000) distinguished four categories of healthcare e-business models: e-commerce portals, e-commerce connectivity, business-to-business (B2B) e-commerce, and business-to-consumer (B2C) e-commerce. In the first category, the author recognized that healthcare portals primarily obtained their revenue through advertising fees. Parente (2000) also acknowledged that healthcare portals offered information retrieval capabilities to both providers and recipients of healthcare services. Portals originated their value proposition through provision of up-to-date and in-depth medical information. Payton (2003) analyzed the features of a number of healthcare Web portals to identify needed enhancements for information services targeting consumers of health plans.

The second category advanced by Parente (2000) in his taxonomy of healthcare e-business models was e-commerce connectivity. This model was primarily supported by revenues originating from transaction-processing activities. Provision of online accessibility to electronic medical records (EMR) and delivery of information on quality and outcomes of healthcare services were the other revenue sources for e-commerce connectivity model. The e-business model of Emdeon Corporation was cited as exemplary for the e-commerce connectivity category. Parente (2000) emphasized that Emdeon's primary value proposition was based on minimal investments in information infrastructure by users of transaction-processing systems. The author also noted that the major hurdle for Emdeon in terms of advancing this e-business model was a low rate of acceptance of new technologies by healthcare organizations. Parente (2000) explained Emdeon's strategy of aggressive acquisitions of healthcare technology firms as a reaction to this impediment. Abrams (2004) conducted a cross-sectional analysis of M&As in the healthcare technology industry and concluded that e-commerce connectivity firms were behind a rise in spending on M&As in 2004. In particular, production of EMR solutions was an area posed for growth. Abrams (2004) noted that e-commerce connectivity firms benefited from an increasing interest, on the part of healthcare institutions, in adding Internet- and software-based components to their services.

Healthcare B2B and B2C e-commerce models function similarly to a general e-commerce model that facilitates efficient market exchanges of goods and services. Parente (2000) described a B2B e-commerce model that offered services unique to healthcare settings. Such B2B platform would facilitate procurement processes involving employers and health insurance companies. A firm deploying this e-business model would act as an agent for employers seeking competitive health plans for their employees. An agent firm would handle a variety of tasks, including assessment of alternative propositions and setting online accounts for individual employees. Singh et al. (2002) described benefits and shortcomings of B2B healthcare e-commerce models. In particular, they addressed cost savings that B2B models delivered to healthcare institutions which had procured medical supplies online. The authors also identified a number of key e-commerce firms specializing on sales of healthcare supplies. The value proposition of the healthcare B2C e-commerce model is based on facilitating online transactions between organizational sellers and individual customers. According to Parente (2000), this e-business model is represented in the healthcare industry by applications that support filling prescriptions online as well as remote management of health plans.

A number of research accounts that have focused on general e-business models have been influential in shaping our analysis of strategy development and implementation at Emdeon Corporation. In particular, taxonomies of Afuah and Tucci (2003) and Weill and Vitale (2001) have complemented findings of the literature assessed above by stressing importance of economic factors for successful exploitation of e-business models. The taxonomy of e-business models devised by Afuah and Tucci depicted various economic dimensions of e-business models: profit site, revenues, and pricing. Our study relies on the "traditional" approach when evaluating business value creation by emphasizing revenue generation and profitability of the firm. When discussing revenue sources, Afuah and Tucci noted that e-business models could rely on commissions, subscription and advertising fees in addition to production, markup and referral revenue bases. These insights are helpful in distinguishing the revenue bases for the e-business model of Emdeon Corporation.

While the distinctive feature of the e-business model taxonomy by Afuah and Tucci (2003) was revenue bases, Weill and Vitale (2001) illuminated several other dimensions of the e-business model. They recognized such dimensions as strategic objective, value proposition, critical success factors, and core competencies in addition to sources of revenue. Weill and Vitale identified these dimensions for each of their eight e-business models, including content provider, intermediary, virtual community, and value net integrator. The dimensions of core competencies and value proposition have also been useful for our analysis of strategy development and implementation at Emdeon Corporation.

Finally, the taxonomy of e-business models developed by Timmers (1999) shed light on the influence of electronic business environments on enterprise value-chain activities. Having placed a major emphasis on Internet instigated value chain modifications, the author distinguished such categories of e-business models as e-procurement, value-chain service provider, virtual business community, collaboration platform, and value-chain integrator in addition to six other categories. While the taxonomies of Weill and Vitale (2001) and Timmers offered a high-level view on various distinct e-business arrangements for value creation, we find it difficult to apply these taxonomies to the case of Emdeon Corporation. The taxonomy of Afuah and Tucci (2003) provided a clearer path for our discussion of strategy development and implementation at Emdeon due to the fact that we decided to focus on the dimensions of revenue generation and profitability when assessing the effectiveness of strategizing.

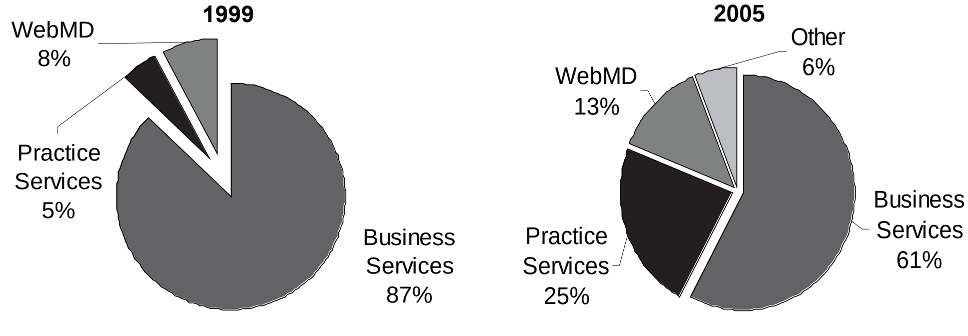
METHOD

This study primarily aims to explore competitive strategies in the emergent healthcare technology industry. We employ the case study approach to investigate the evolution of the e-business model of a healthcare technology firm. Benbasat, Goldstein, and Mead (1997) argued that the case study method could be used in information systems (IS) research for both exploration and explanation. Benbasat, Goldstein, and Mead (1997) noted that "a case approach is an appropriate way to research an area in which few previous studies have been carried out" (p. 370). The literature on e-business models in healthcare technology industry has not gained a critical mass yet. Healthcare technology markets have been undergoing a period of formation from the mid-1990s to the present day. Only recently, leaders have emerged in respective technology markets. We aim to trace the evolution of interactions between a leading firm in healthcare technology industry and pertinent client bases to identify lessons that we can learn from inception and implementation of a sustainable e-business model.

DATA

The data used in this research was primarily compiled from U.S. Securities and Exchange Commission (SEC) 10-K quarterly and annual reports provided by Hoover's Online. In particular, we collected the data on revenues and earnings in 1998-2005 as well as in the first quarter of 2006 for the individual segments of Emdeon's e-business model. SEC 10-K quarterly and annual reports from Hoover's Online also provided the data on the M&As that Emdeon completed in the period of study. The second source of research data was Standard & Poor's Market Insight database. We used analytical reports contained in this database to collect the data for the analysis of healthcare technology industry.

Figure 1. Emdeon's Annual Sales (Revenues) by Segments in 1999 and 2005



Note. From Securities and Exchange Commission 10-K annual and quarterly reports available from Hoover's Online database.

EVOLUTION OF EMDEON'S E-BUSINESS MODEL

Emdeon Corporation has evolved from a company that relied on a single source of revenue with 648 employees and $48.8 million of annual sales in 1998 to a leading multi-segment firm with 6,100 employees and $1,277 million of annual sales in 2005 (SEC Reports, 1998, 2005). Currently, the firm controls a 17 % share of the market for healthcare technology services (GICS Sub-Industry Profile, 2006). We argue that the key factor to achieved sustainability and market success has been a related diversification strategy that Emdeon carried out via M&As. This strategy has enabled the firm to build a sophisticated e-business model that relies on several robust sources of revenue.

Entry to Healthcare Technology Markets

Emdeon acquired electronic transaction-processing capabilities by purchasing ActaMed Corporation in 1998. In the next year, the firm made an IPO on the NASDAQ under the name of Healtheon. At that point, healthcare transaction processing services (Business Services according to Emdeon's terminology) constituted Healtheon's primary source of revenue, which accounted for 87% in 1999 (see Figure 1). The same year, Healtheon merged with WebMD Corporation aiming to expand into healthcare portal services (WebMD segment according to Emdeon's terminology). The addition of Medical Manager Corporation formed the other core competency: software development services (Practice Services according to Emdeon's terminology) targeting healthcare providers. Therefore, by pursuing a related diversification strategy, Emdeon laid a foundation for its e-business model that largely remains intact to the present day. Through the key acquisitions made in 1998-2000, Emdeon achieved the economy of scope effectively building a diversified e-business model that relied on multiple sources of revenue.

When Emdeon entered the market for medical portal services following the merger of Healtheon and WebMD Corporation in 1999, the portal services market was still in the experimentation phase. Elfenbein and Lerner (2003) outlined the two milestones for the early phase of development of portal services. The inception of portal services is traced back to 1994 when the users of the World Wide Web began to actively adopt the Internet browser. Portal services at that time attracted users through agglomeration of hyperlinks leading to the Web sites of interest, Web search capabilities, and proprietary content. In 1997, portal services entered a new phase of development, which denoted a boost in new content offerings such as online news and stock price information. Later, portal firms also expanded service offerings by adding such features as Internet-based auctions and electronic mail. According to Elfenbein and Lerner, Web portal firms primarily received revenue from two sources: fees for placing advertisements on portal Web pages and alliance agreements. In the portal services market, Emdeon chose to target a specific audience: providers and recipients of healthcare services.

In the 2000s, the Healtheon/WebMD portal has gained a noticeable position on the World Wide Web. A number of commentaries assessed the impact of this portal from a variety of perspectives. According to one commentary (Singh et al., 2002), the visibility level that Healtheon/WebMD portal achieved made it a much-discussed destination for online medical services (e.g., transcription). Another commentary (Damsgaard, 2002) pointed to the general popularity of Healtheon/WebMD portal. The author viewed this portal as successful, noting that the survival rate in the portal services market was very low. The model of portal management that Damsgaard advanced pointed to the importance of building a strong customer base and proactively responding to changes in the external environment.

Table 1. Emdeon's annual sales by segments in 1998-2005 (millions of dollars)

	2005	2004	2003	2002	2001	2000	1999	1998
Business Services	759	687	506	467	384	269	46	34
Practice Services	304	296	303	275	260	120	3	
WebMD	168	134	111	84	75	102	4	
Other	79	77	72					

Note. From Securities and Exchange Commission 10-K annual and quarterly reports available from Hoover's Online database.

Emdeon Corporation entered the medical software market in 2000 when the firm acquired Medical Manager Corporation, whose most established unit was a practice management system. Emdeon pursued opportunities in an emerging technology niche that served physician practices. It is important to note that the market for software development services in healthcare at that point had reached maturation for large institutions. U.S. hospital networks began to deploy information systems as early as 1960s. Initially, hospitals invested their own resources in the development of such systems. For the most part, affiliated academic research centers provided the required computer resources and professional expertise (Collen, 1991). In the 1970s and 1980s, the number of software development firms serving U.S. hospitals considerably increased (Michell & Singh, 1996). By the early 1990s, the market for hospital information systems reached maturity in terms of both cumulative sales amount and number of firms offering information systems solutions. According to Michell and Singh, 491 firms offered such solutions to financial and business office operations of U.S. hospitals in 1991. In addition, 218 firms brought IS solutions designed to manage patient records to the hospital software market in the same year.

The Medical Manager set of applications targeted a different niche by offering practice management software to physicians. Its integrated software solutions served major business aspects of a doctor's office and consisted of clinical, financial, and patient data management modules. Prior to the acquisition by Emdeon, Medical Manager had been installed at 25,000 businesses representing 80 medical specialties, according to the 1999 SEC 10-K annual report of Medical Manager Corporation. Acquired software development capabilities produced $120 million, which accounted for 24% of Emdeon's total revenue in 2000 (see Table 1).

Emdeon gained an additional revenue stream from an unrelated segment: production and distribution of plastic materials. These materials primarily target customers in the healthcare industry. This segment was inherited from Medical Manager Corporation acquired for its software development line of products. For the two years following the acquisition, Emdeon tried unsuccessfully to divest this unrelated segment. This fact presents evidence that the firm focused on a related diversification strategy upon the entry to healthcare technology markets.

Strengthening a Diversified E-Business Model

For the three e-business segments, Emdeon pursued two strategies to advance its market positioning: economy of scale and economy of scope. The former strategy dealt with customer base expansion while the latter focused on quickly adding new technologies and services. Emdeon Corporation skillfully mastered the art of expanding its customer base as well as scope of technologies and services via M&As. Table A1 summarizes the history of M&As completed by Emdeon for the software development segment in 2000-2005. In the period of 2001-2003, Emdeon acquired 38 small software development companies that provided technology services to physician practices for the total amount of $24.9 million. This allowed Emdeon to reach the level of critical presence in the market of physician technology services.

Table A2 illustrates the dynamics of acquisitions made by Emdeon for the segment of electronic medical-transaction processing services. The first series of additions were made in 1999-2000. Emdeon acquired firms that had closely related e-business models, including Envoy Corporation, Kinetra, LLC (a joint venture of EDS and Eli Lilly), and MEDE America Corporation. The next wave of acquisitions for this segment, performed in 2003-2004, signaled a change in strategy. Emdeon sought and acquired targets that would expand its scope of technologies and services. The firm added print-and-mail services through the acquisition of Advanced Business Fulfillment, Inc. The addition of Medifax-EDI, Inc. provided a new capability for transaction processing services: real-time eligibility verification. In 2004, Emdeon acquired Dakota Imaging, Inc. and ViPS, Inc. to add capabilities in the areas of fraud detection and predictive modeling. At the same time, these acquisitions expanded the customer base of Emdeon. The addition of ViPS, Inc. resulted in entering a technology market serving healthcare government agencies and large insurance companies. The acquisition of Claims Processing Systems, Inc. brought in the clientele consisting of dental practices.

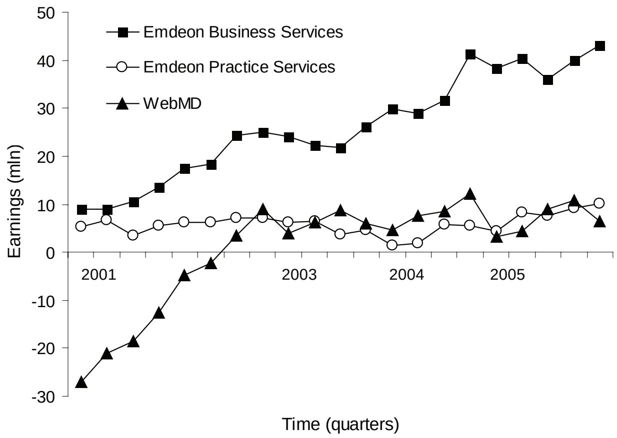
Emdeon faced a greater challenge for strengthening the portal services segment. This segment differed from the two other segments as it primarily relied on advertisement fees, and was a new e-business model at the time. Emdeon's portal services experienced losses for a number of years. The segment first became profitable in the third quarter of 2003 (see Figure 2). To this end, Emdeon made several acquisitions to provide enhancements to the portal's features as well as to diversify its customer base. Initially, portal services targeted online consumers of health information with such services as enrollment to health plans and health information retrieval. In 2001, Emdeon acquired Medscape, Inc. to penetrate a new customer base: providers of healthcare services. This development required addition of new portal features. Emdeon embarked on a series of acquisitions to offer additional Web site content: an online drug directory, online capabilities to compare costs and quality of healthcare providers, and online databases containing physician contact information and medical treatment guidelines. Table A3 displays the M&As that Emdeon pursued to carry out these changes.

Figures 2 and 3 display a steady growth of revenues and earnings for each segment. The M&As that Emdeon executed in 1998-2005 resulted in achieving sustainability and profitability of its e-business model.

Organizational Restructuring

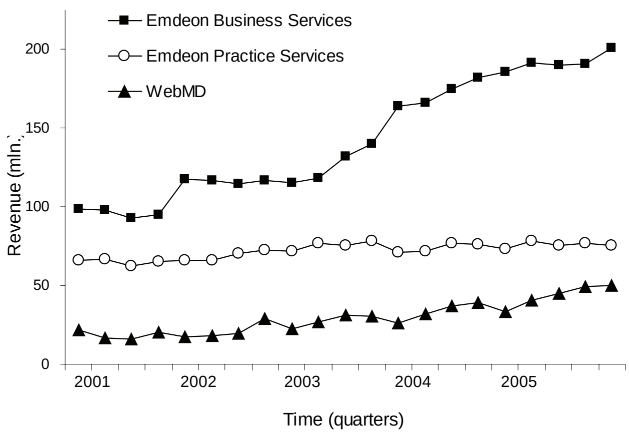
In the period of 2001-2005, Emdeon aggressively acquired companies that would fit the profiles of its three e-business segments: online medical-transaction processing, software development and portal services. Currently, the largest fraction of revenue for Emdeon comes from the transaction-processing segment. Development and sales of specialized software targeting healthcare providers make up the second largest source of revenue. The portal services segment generates the smallest amount of revenue. Changes in the breakdown of

Figure 2. Emdeon's quarterly earnings before interest by segments in 2001-2006



Note. From Securities and Exchange Commission 10-K annual and quarterly reports available from Hoover's Online database. Quarterly sales and earnings data by business segments prior to year 2001 are not available because Emdeon Corporation started to distinguish sales and earnings by business segments for quarterly reporting only in year 2001.

Figure 3. Emdeon's quarterly revenues by segments in 2001-2006



Note. From Securities and Exchange Commission 10-K annual and quarterly reports available from Hoover's Online database. Quarterly sales and earnings data by business segments prior to year 2001 are not available because Emdeon Corporation started to distinguish sales and earnings by business segments for quarterly reporting only in year 2001.

revenues by segment and distribution of revenues by segment over the observed period are displayed in Figure 1 and Figure 3 respectively.

Having built a sustainable e-business model, Emdeon embarked on a new strategy: organizational restructuring. In 2005, the firm spun off its portal services division under the name of WebMD Health Corporation. Realized profitability and a greater revenue growth rate in the recent years made this development possible. As evident in Figure 3, a revenue growth of the portal services segment enjoyed a healthy rate. The earnings of WebMD Web sites have been positive for a number of years as shown in Figure 2.

WebMD Health Corporation has been very active in pursuing M&As. The firm added Summex Corporation in July 2006 to supply information about wellness and health education programs, effectively expanding its array of services. In addition, WebMD made an announcement about its intention to purchase Medsite Inc., whose e-business model offered services in the areas of interactive medical education and physician recruitment. WebMD evidently continues the tradition of its parent, Emdeon, aiming to expand its market share and diversify its array of services through aggressive M&As.

CONCLUSION

The healthcare technology industry is dynamically evolving. Having survived a major environmental shock, the dot.com bust, some firms in this industry are experiencing an impressive revenue growth. The industry currently enjoys a 21.8% revenue growth based on the 12-month period compared with a 13.4% growth for the rest of the market[1]. The firm in our study, Emdeon Corporation, emerged as a market leader in the healthcare technology industry by being able to adjust quickly to evolving market conditions. The key to the firm's success was its e-business model, which relied on multiple sources of revenue.

To enter healthcare technology markets, Emdeon Corporation pursued a related diversification strategy. In the span of three years, Emdeon made key acquisitions that formed distinct revenue segments of its e-business model: electronic medical-transaction processing, software development and portal services. The first lesson learned from Emdeon's experience is that a related diversification strategy creates a competitive advantage upon the entry to emerging niche technology markets.

Emdeon implemented a related diversification strategy by the means of M&As to achieve the economies of scope and scale quickly. Although Emdeon has entered a number of strategic alliances, the firm invests more aggressively in acquiring companies. The second lesson learned from Emdeon's experience is that firms in emerging niche technology markets can use M&As to sustain revenue growth and to increase market power.

Emdeon's upper management made a major strategy shift by spinning off one of the segments, portal services. Having secured a great deal of market power for this segment, Emdeon currently focuses on strengthening market positioning of the other segments. This change in Emdeon's organizational structure points out that the firm's e-business model has reached a certain level of maturity. The lesson learned from Emdeon's restructuring strategy is that maturation of diversified e-business models in niche technology industries leads to the transformation of individual segments into distinct entities focusing on specific technology markets.

REFERENCES

Abrams, S.R. (2004). Healthcare IT: Hot M&A area. *Investment Dealers Digest*, *70*(30), 12-13.

Afuah, A., & Tucci, C. (2003). *Internet business models and strategies.* Boston: McGraw-Hill/ Irwin.

Benbasat, I., Goldstein, D.K., & Mead, M. (1987). The case research strategy in studies of information systems. *MIS Quarterly*, *11*(3), 369-386.

Collen, M.F. (1991). A brief historical overview of HIS evolution in the USA. *International Journal of Bio-Medical Computing, 29*, 169-189.

Damsgaard, J. (2002). Managing an internet portal. *Communications of the Association for Information Systems*, *9*, 408-420.

Elfenbein, D.W., & Lerner J. (2003). Ownership and control rights in internet portal alliances, 1995-1999. *RAND Journal of Economics*, *34*(2), 356-369.

Emdeon Corporation. (n.d.). *Competitive landscape*. Retrieved February 8, 2007, from Hoover's Online, http://www.hoovers.com

Ford, E.W., Menachemi, N., & Phillips, M.T. (2006). Predicting the adoption of electronic health records by physicians: When will health care be paperless? *Journal of the American Medical Informatics Association*, *13*(1), 106-112.

GICS Sub-Industry Profile (2006). *Health care technology*. Retrieved July 30, 2006, from Standard & Poor's Market Insight, http://emi.compustat.com

Goldschmidt, P. (2005). HIT and MIS: Evaluating the potential advantages and considering the risks associated with electronic health care Records. *Communications of the ACM*, *48*(10), 69-74.

Michell, W., & Singh, K. (1996). Survival of businesses using collaborative relationships to commercialize complex goods. *Strategic Management Journal*, *17*(3), 169-195.

Microsoft to offer software for health care industry. (2006, July 27). *New York Times*.

National Coalition on Health Care. (n.d.). *Facts about health care*. Retrieved February 8, 2007, from http://www.nchc.org/facts/cost.shtml

Office of the National Coordinator for Health Information Technology. (2005, June). *Summary of nationwide health information network (NHIN) request for information (RFI) responses*. Retrieved February 8, 2007, from http://www.hhs.gov/healthit/rfisummaryreport.pdf

Parente, S.T. (2000, November-December). Beyond the hype: A taxonomy of e-health business models. *Health Affairs*, pp. 89-102.

Payton, F.C. (2003). E-health models leading to business-to-employee commerce in the human resources function. *Journal of Organizational Computing and Electronic Commerce*, *13*(2), 147-161.

Prognosis positive: Web health venture. (2006, May 25). *Wall Street Journal*.

Securities and Exchange Commission Fillings. (1998-2006). *Securities and Exchange Commission 10-K and 10-Q Reports*. Retrieved February 8, 2007, from Hoover's Online via http://www.hoovers.com

Singh, G., O'Donoghue, J., & Soon, C.K. (2002). Telemedicine: Issues and implications. *Technology and Health Care*, *10*, 1-10.

Timmers, P. (1999). *Electronic commerce: Strategies and models for business-to-business trading*. Chichester, UK: John Wiley.

Weill, P., & Vitale, M.R. (2001). *Place to space: Migrating to e-business models*. Boston: Harvard Business School Press.

ENDNOTE

1. Public companies trading on the New York Stock Exchange, the American Stock Exchange, and the NASDAQ National Market. Data obtained from Hoovers Online via http://www.hoovers.com.

APPENDIX

Table A1. Emdeon's acquisitions made for its Practice Services Segment in 2000-2005

Company	Date	Price (in mln)	Profile
Medical Manager Corporation/ CareInsite Inc.	2000	2907	Medical Manager® practice management system.
10 small physician services companies	2001	8.2	Services for physician practices.
21 small physician services companies	2002	14	Services for physician practices.
7 small physician services companies	2003	2.3	Services for physician practices.
Conceptis Technologies, Inc.	2005	19.6	Medical education and promotion services aimed at physicians and other healthcare professionals with a strong online presence in the cardiology community.

Note. From Securities and Exchange Commission 10-K annual reports available from Hoover's Online database.

Table A2. Emdeon's mergers and acquisitions made for its Business Services Segment in 1998-2005

Company	Date	Price (in mln)	Profile
ActaMed Corporation	1998	n/a	Electronic data interchange via proprietary non-Internet network.
MEDEAmerica Corp.	1999	417	Automated transaction operations.
Kinetra LLC	2000	292	Electronic clinical transactions.
Envoy Corporation	2000	2,440	Electronic medical-transaction operations.
Advanced Business Fulfillment, Inc.	2003	113	Paid-claims communications services for third-party administrators and health insurers; print-and-mail capabilities.
Claims Processing Systems, Inc.	2003	5.6	Clearinghouse for dental practices.
Medifax-EDI, Inc.	2003	268	Real-time medical eligibility transaction services and other claims management solutions.
Dakota Imaging, Inc.	2004	39	Automated claims processing and business process outsourcing services; advanced data scrubbing.
ViPS, Inc.	2004	167	Information services to government and commercial healthcare payers, including provider performance measurement, fraud detection, disease management and predictive modeling.

Note. From Securities and Exchange Commission 10-K annual reports available from Hoover's Online database. All acquisitions were accounted for as purchases, except for ActaMed Corporation, which was accounted for as a pooling of interests.

Table A3. continued

Optate, Inc.	2003	4.1	Online healthcare benefit decision support tools.
The Little Blue Book	2003	13	Online database and reference book containing physician practice information.
RxList, LLC	2004	5.2	Online drug directory for consumers and healthcare professionals.
MedicineNet, Inc.	2004	17	Health information Web site for consumers.
HealthShare Technology, Inc.	2005	30	Online tools to compare cost and quality of hospitals for use by consumers, providers and health plans.
eMedicine.com, Inc.	2006	26	Medical reference information and clinical knowledge base for healthcare professionals in 59 medical specialties; a consumer site with articles written by physicians for patients.
Summex Corporation	2006	10	Online and offline health and wellness information and lifestyle education.

Note. From Securities and Exchange Commission 10-K annual reports available from Hoover's Online database.

This work was previously published in International Journal of E-Business Research, Vol. 3, Issue 3, edited by I. Lee, pp. 1-13, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 6.10
M-Health:
A New Paradigm for Mobilizing Healthcare Delivery

Nilmini Wickramasinghe
Illinois Institute of Technology, USA

Steve Goldberg
INET International Inc., Canada

ABSTRACT

Medical science has made revolutionary changes in the past decades. Contemporaneously, however, healthcare has made incremental changes at best. The growing discrepancy between the revolutionary changes in medicine and the minimal changes in healthcare processes is leading to inefficient and ineffective healthcare delivery and one if not the significant contributor to the exponentially increasing costs plaguing healthcare globally.
Healthcare organizations can respond to these challenges by focusing on three key solution strategies, namely, (1) access – caring for anyone, anytime, anywhere; (2) quality – offering world-class care and establishing integrated information repositories; and (3) value – providing effective and efficient healthcare delivery. These three components are interconnected such that they continually impact the other and all are necessary to meet the key challenges facing healthcare organizations today. The application of mobile commerce to healthcare, namely, m-health, appears to offer a way for healthcare delivery to revolutionize itself. This chapter serves to outline an example of adopting mobile commerce within the healthcare industry, namely, in the area of a wireless medical record. In particular, it discusses an appropriate, feasible mobile solution to enable hospitals operate effectively and efficiently in today's competitive and costly healthcare environment as well as meet all the necessary regulatory requirements. The lessons learnt from these case study data should be of interest to both practitioners and researchers since they will outline realistic and feasible solutions to enable hospitals to incorporate a wireless/m-commerce solution as well as highlighting key areas for further research in this important area of high-quality, effective, and efficient healthcare management.

INTRODUCTION

Currently the healthcare industry in the United States as well as globally is contending with relentless pressures to lower costs while maintaining and increasing the quality of service in a challenging environment (Pallarito, 1996, pp. 42–44; Wickramasinghe & Silvers, 2003, pp. 75–86). It is useful to think of the major challenges facing today's healthcare organizations in terms of the categories of demographics, technology, and finance. Demographic challenges are reflected by longer life expectancy and an aging population; technology challenges include incorporating advances that keep people younger and healthier; and finance challenges are exacerbated by the escalating costs of treating everyone with the latest technologies. Healthcare organizations can respond to these challenges by focusing on three key solution strategies, namely, (1) access – caring for anyone, anytime, anywhere; (2) quality – offering world-class care and establishing integrated information repositories; and (3) value – providing effective and efficient healthcare delivery. These three components are interconnected such that they continually impact the other and all are necessary to meet the key challenges facing healthcare organizations today.

In short then, the healthcare industry is finding itself in a state of turbulence and flux (Wickramasinghe & Mills, 2001, pp. 406–423). Such an environment, we believe, is definitely well suited for a paradigm shift with respect to healthcare delivery. Therefore, in this chapter we address the issue of wireless solutions for healthcare delivery and management.

First, we discuss the findings from INET's study on mobile Internet (wireless) technology in healthcare by Ontario Hospitals in Canada. We use these findings as a launching place to review a rigorous way to accelerate healthcare delivery improvements. Next, we outline some preliminary evidence for using a standardized mobile Internet (wireless) environment in healthcare. For example, INET International is advocating the use of a wireless healthcare portal to validate the possible reduction in IT infrastructure costs. A portal may reside on a wireless PDA device as single point of contact for clinicians to obtain immediate patient data (radiology reports, lab results, and clinical findings). This wireless portal may also improve patient care outcomes with access to the best available clinical evidence at the point of care. We shall also describe the current status of a standardized mobile Internet (wireless) environment in terms of technology requirements, security readiness, and IT management practices. In addition, we will also outline some of the key challenges that a hospital's IT department, medical units, administration, and clinicians will face regarding a wireless project and provide some reasonable solutions to these challenges. Finally, we shall outline the key steps necessary for a hospital to transition from proprietary information systems to a three-tier Web-based architecture.

INETS STUDY IN MOBILE INTERNET TECHNOLOGY IN HEALTHCARE

INET International Inc. delivers rigorous e-business acceleration projects in large corporate, government, and healthcare organizations. The organization focuses on custom Mobile, Internet, Intranet, Extranet (INET), and wireless solutions. It was founded by Steve Goldberg in 1998 and it leads a Wireless Technology Consortium (WTC) to collect evidence on the best way to use wireless technology to accelerate healthcare delivery improvements. These applications are designed to improve patient care, reduce costs, increase healthcare quality, and enhance teaching and research.

Over a period of 2 years INET has been conducting research that has been directed at how to apply mobile Internet wireless technologies' low-cost advantages to evolve a wireless healthcare portal. A portal is a single point of contact for healthcare providers and handheld technology applications (HTA) to access and process various data pertaining to patients such as: (1) patient-specific data (i.e., patient ID, radiology reports, lab results, clinical findings, and research data); (2) medical knowledge (primarily from evidence-based medicine training and journals); (3) clinical guidelines (i.e., association guidelines such as the Association of Radiologists clinical practice publications); and (4) reimbursement rules and data (i.e., Ontario Health Insurance Plan, known as OHIP). This research has shown that mobile/wireless solutions for healthcare can achieve four critical goals of (1) improving patient care, (2) reducing transaction costs, (3) increasing healthcare quality, and (4) enhancing teaching and research.

Improving Patient Care and Increasing Healthcare Quality

In the final report compiled by the Committee on the Quality of Healthcare in America (AIM, 2001), it was noted that improving patient care is integrally linked to providing high-quality healthcare. Furthermore, in order to achieve a high quality of heathcare the committee identified six key aims, namely, (1) healthcare should be safe – avoiding injuries to patients from the care that is intended to help them; (2) effective – providing services based on scientific knowledge to all who could benefit and refraining from providing services to those who will not benefit (i.e., avoiding under use and overuse); (3) patient centered – providing care that is respectful of and responsive to individual patient preferences, needs, and values, and ensuring that patient values guide all clinical decisions; (4) timely – reducing waiting and sometimes harmful delays for both those receiving care and those who give care; (5) efficient – avoiding waste; and (6) equitable – providing care that does not vary in quality based on personal characteristics.

Most of the poor quality connected with healthcare is related to a highly fragmented delivery system that lacks even rudimentary clinical information capabilities resulting in poorly designed care processes characterized by unnecessary duplication of services and long waiting times and delays (AIM, 2001). The development and application of sophisticated information systems is essential to address these quality issues and improve efficiency, yet healthcare delivery has been relatively untouched by the revolution of information technology that has transformed so many areas of business today (Wickramasinghe & Mills, 2001, pp. 406–423; Wickramasinghe & Silvers, 2003, pp. 75–86). This, then, certainly justifies the need for e-business solutions for healthcare delivery; however, from various applied research scenarios (Goldberg et al., 2002a, 2002b, 2002c, 2002d, 2002e), we can see why mobile solutions appear to be superior since they enable even better care for patients with the added advantage of significant cost savings. Physicians themselves are excited by the possibilities offered by mobile solutions as the following quotation discussing benefits to radiologists exemplifies:

"The demands on radiologists in today's fast-paced and cash-strapped hospitals are tremendous," says Dr. Brian Yemen, radiologist at Hamilton Health Sciences. "The possibility of transmitting radiology reports between hospitals and physicians on-demand is exciting. It will mean less time and travel for busy doctors, but more importantly, it will mean faster turnaround and quicker results for patients."

A mobile Internet wireless solution uses a personal digital assistant (PDA) device at the point of patient care. Such handheld technology applications are simple to use and require little, if any, training. Furthermore, they can significantly reduce change management costs which are usually dominated by technology education. Typically, the PDA devices and application require very little training unlike most PC-based e-health alternatives where even simple skills like typing prove to be challenging hurdles for physicians and other users to overcome. This means that not only do physicians have the information they need in a timely fashion but they also are not restricted by typing or other activities that are not directly going to benefit the patient's encounter. In Canada, most manufacturers and IT national resellers have found no need to set up educational practices to support over 1 million PDA users. Currently, Canadian retail stores provide the bulk of the customer service requirements, such as, product pricing, availability, promotions, and product warranty. Usually, the product warranty process replaces defective PDA devices with a new or refurbished unit within hours or the next business day with minimal inconvenience to users.

Cost Reductions: Wireless Technology May Reduce IT Infrastructure Costs by 84%

- Integral to the set up of such a mobile solution is the wireless healthcare portal. This portal can dramatically reduce IT infrastructure project costs. For example, the government of Ontario, Ministry of Health and Long-Term Care is planning to spend $150 million to develop a turn-key IT solution for 6,400 physicians. By using wireless technology it is possible to reduce this IT infrastructure cost to $24 million by leveraging hospital IT infrastructure investments. The preliminary cost estimate is $1,259/year/physician over 3 years. Table 1 outlines the cost estimates/ year/physician for each component.

Thus, the findings summarized above from the research conducted by INET demonstrate not only does the mobile/wireless solution support improving patient care, but it also is a cost-effective solution. Furthermore, not only can cost savings be enjoyed by leveraging off existing infrastructure, but the wireless solution also lends itself to rapid healthcare delivery improvements as well as the ability to deploy healthcare improvements within short time cycles as discussed below.

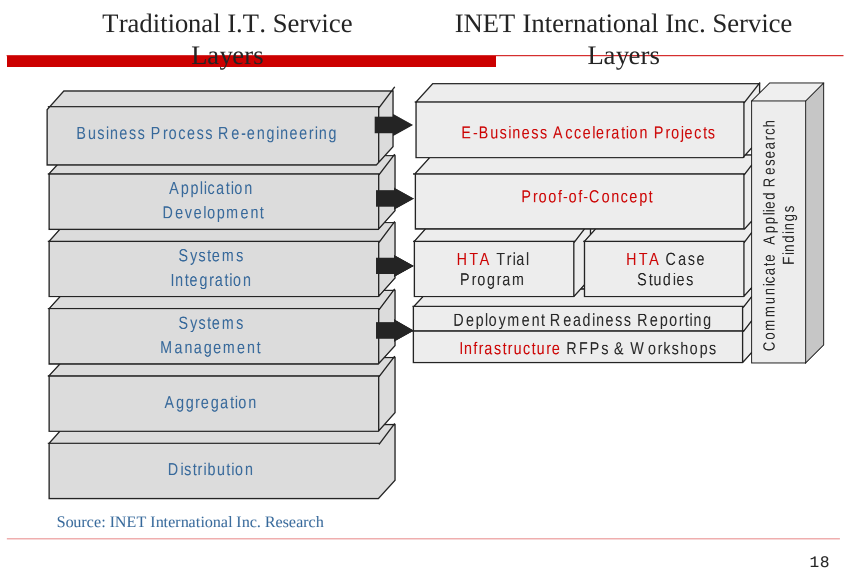
Achieve Rapid Healthcare Delivery Improvements

By adopting a mobile/wireless healthcare delivery solution, it is possible to achieve rapid healthcare delivery improvements, which impacts both the costs and the quality of healthcare delivery. This is achieved by using an e-business acceleration project which provides hospitals a way to achieve desired results within a standardized mobile Internet (wireless) environment as shown in Figure 1.

Table 1. Cost estimates

Component	Total Cost	Comment
Mobile Internet (wireless) Infrastructure Support	$60	Annual
Mobile Internet (Wireless) Infrastructure Upgrade	$100	Amortize Over 3 Years
PDA Device	$133	Amortized Over 3 Years
Cell Phone/Data Cable/Modem	$66	Amortized Over 3 Years
Wireless Communication Costs	$420	$35/Month – transmits 25 to 50 patient records/day
Physician Wireless Portal	$480	$40/Month – assume four handheld technology applications

Figure 1. E-business accelerated project



Integral to such an accelerated project is the ability to build on the existing infrastructure of the hospital. This then leads to what we call the three-tier Web-based architecture. In such an environment, tier 1 is essentially the presentation layer, which contains the Web browser but no patient data is stored within this layer, and thereby ensuring compliance with international security standards/policies like HIPAA. Tier 2 then provides the business logic, including but not limited to, lab, radiology, and clinical transcription applications; messaging of HL7, XML, DICOM, and other data protocols; and interface engines to a Hospital Information Systems (HIS,) Lab Information Systems (LIS), Radiology Information Systems (RIS), as well as external messaging systems such as Smart Systems for Health (an Ontario Healthcare IT infrastructure project). Finally there is the tier 3 architecture which consists of the back-end databases such as Oracle or Sybase.

Deploying Healthcare Improvements within Short Time Cycles

A mobile Internet (wireless) infrastructure may achieve rapid technology deployments to individual practices, small clinics, and healthcare networks at a very low cost. For example, for $35/month a physician can have wireless access to a Radiology Information System (RIS) to retrieve diagnostic imaging (DI) reports. With a minimum cost these RIS systems may easily and securely access reports from the hospital, multiple hospitals (using Smart Systems for Health) or interface into other diagnostic imaging (DI) clinics as a single point of contact. The objective is to improve patient care by getting DI results faster to patients (through referring physicians). Once the solution achieves approval for general release, it has the possibility of being deployed to 3,000 to 4,000 physicians within short time cycles so that, for example, clinicians using a mobile Internet (wireless) connection are able to hook into an existing hospital infrastructure as in the case of the Ontario hospitals where they use the Hamilton and Toronto downtown hospitals' IT infrastructures.

Teaching and Research

The provision of the best available clinical evidence at the point of care is playing a pivotal role in medical research and training (www.cochrane.org). Today, a physician-led research center (www.cebm.utoronto.ca/projects/index.htm) is conducting an applied research project on the impact on mobile Internet (wireless) technology to help clinicians with the following:

- Consistently translate high-quality clinical evidence into practice.
- Obtain the information they need from the right resources.
- Engage in a process to replace the need to read 17 articles per day.

The challenges of using mobile Internet technology have been documented by this center. This includes the following:

- What kinds of information are most useful to clinicians?
- What is the most effective way of querying evidence-based resources?
- How do we format answers?

INET International Inc. applied research on a standardized mobile Internet (wireless) environment may provide some unique solutions. The solutions may include the following:

- Clinicians can participate in a rigorous e-business acceleration project to narrow information requirements into a few useful requirements. For instance, radiologists have identified and confirmed the importance of referring physicians getting imaging reports and results to patients faster.
- By harnessing a hospital's information systems (HIS) and standardized mobile Internet (wireless) environment a physician can gain access to patient-specific data (i.e., imaging reports). This may provide a new opportunity to evaluate how physicians can use current patient data to determine the best ways to query evidence-based resources. Also, researchers may extend the use of patient-specific data to nurse practitioners, home care providers, and patients to help develop the most effective way of querying evidence-based resources within a consumer-centric model.
- Finally, INET International Inc. research on a standardized mobile Internet (wireless) environment may help physician-led research centers select and aggregate Internet search forms and Internet "answer" formats from multiple resources into a single wireless personal digital assistant (PDA) presentation. Additionally, researchers can make rapid changes to Internet "answer" formats to accommodate user feedback. They work within this standardized environment to control iterative prototyping for the delivery of small and frequent proof of concepts and randomized trials. The use of release management and quality assurance tools enable the scaling of successful trials into wireless production systems for tens of thousands of healthcare providers and patients within a very short time cycle.

To summarize, based on the findings from the various research endeavors conducted by INET Inc. (Goldberg et al., 2002a, 2002b, 2002c, 2002d, 2002e), using a standardized mobile Internet (wireless) environment appears to be the simplest way to access the best-available clinical evidence at the point of care for everyone, anywhere, anytime at a very low cost. The achievement of superior quality cost-effective healthcare is uppermost on the agendas of most countries' healthcare initiatives.

We believe this can only be realized by adopting some type of the wireless solution.

THE MOBILE VISION

Having described the key benefits of a mobile/wireless solution, let us now turn to focus on how to set about actualizing such a mobile vision. We illustrate the possibilities by discussing INET's strategies and approaches.

Research Goal and Approach

Simply stated, the research goal is to use a standardized mobile Internet (wireless) environment to improve patient outcomes with immediate access to patient data and provide the best-available clinical evidence at the point of care. To do this, INET International Inc. research starts with a 30-day e-business acceleration project in collaboration with many key actors in hospitals such as clinicians, medical units, administration, and IT departments. Together they follow a rigorous procedure that refocuses the traditional 1–5 year systems development cycle into concurrent, 30-day projects to accelerate healthcare delivery improvements. Figure 2 highlights the key success factors required in such an approach.

The completion of an e-business acceleration project delivers a scope document to develop a handheld technology application (HTA) proof of concept specific to the unique needs of the particular environment. The proof of concept is a virtual lab case scenario. A virtual lab operates within a mobile Internet (wireless) environment by working with hospitals and technology vendors. The final step is the collection of additional data with clinical HTA trials consisting of 2-week hospital evaluations.

E-Business Acceleration Project Outcomes

The first e-business acceleration project for healthcare was conducted for Hamilton Health Sciences (HHS) Diagnostic Imaging (DI) Department.

They followed a procedure to divide an HHS enterprise DI process into smaller mini-processes. This provided a new way for radiologist to enhance DI delivery in small manageable pieces and minimize risk to patients. Hence we can see that a critical step in the e-business accelerated project is often to reengineer existing processes to ensure they are as effective and efficient as possible.

During the first e-business acceleration project a DI sign-off report process was selected and a handheld technology application (HTA) proof of concept was developed.

The next step, HHS is planning an HTA trial to evaluate a low-cost, secure, and simple way for radiologists and referring physicians to access critical healthcare information at the point

Figure 2. Key success factors

Mobile Internet (wireless) Technology Key Success Factors For Health Care	
Physician-led	Widespread physician acceptance of handheld technology applications
Simple to Use	Personal Digital Assistance, Bluetooth, Java Cell Phones, Digital communicators
Low Cost	Convergence of wireless IP WAN and wireless IP LANs.

of patient care. The radiologist and referring physicians could retrieve and review a patient's DI reports using a wireless PDA device—getting diagnostic results faster to patients and thus enhancing quality as well as effective and efficient healthcare delivery.

A series of 2-week evaluations are planned for hospitals in Ontario and United States. This trial program is presented in Figure 3. Each hospital provides a different environment to evaluate the scalability of the DI sign-off report HTA. The aggregation of such 2-week evaluation outcomes provides hospitals evidence on key areas such as:

- Showing clearly how reduction in large-scale IT infrastructure project costs can occur.
- Demonstrating that the PDA is simple to use, requiring little or no training of clinicians.
- Clearly outlining how to achieve rapid healthcare delivery improvements.
- Demonstrating how to scale and deploy healthcare delivery improvement to clinicians within short time cycles.

Applied Research Challenges

Engaging in such a project, naturally, has several challenges. Initially technology vendors saw the immediate opportunities to engage in INET International Inc. Virtual Lab to work with Hamilton Health Sciences (HHS). The lab is a low-cost way for vendors to connect demonstration facilities and resources to conduct proof of concepts and clinical trials. However, growing the lab to 10 hospitals involves the formalization of a hospital procurement process that includes applied research, that is, a proof of concept and a trial program.

Traditionally IT operates within 1- to 5-year systems development life cycle to meet hundreds of end-user requirements. Whereas the e-business acceleration project process selects one or two requirements to deliver improvements in concurrent short time cycles. To expand the applied research of a rigorous approach to accelerate healthcare delivery improvements requires additional investment by hospitals, something that is clearly challenging.

Possibly the most significant impact of the virtual lab is the resolution of mobile Internet (wireless) technology challenges. The HHS applied research findings (lesson learned) provide a way to help wireless projects achieve the same results in a much shorter time cycle. HSS is an academic health sciences center whose primary role is healthcare delivery. It is one of the largest hospitals in Ontario, operating across four sites with approximately 8,000 employees and 1,000 physicians and thus the research findings from HHS are detailed and extensive and a useful resource to help any other wireless project in this area.

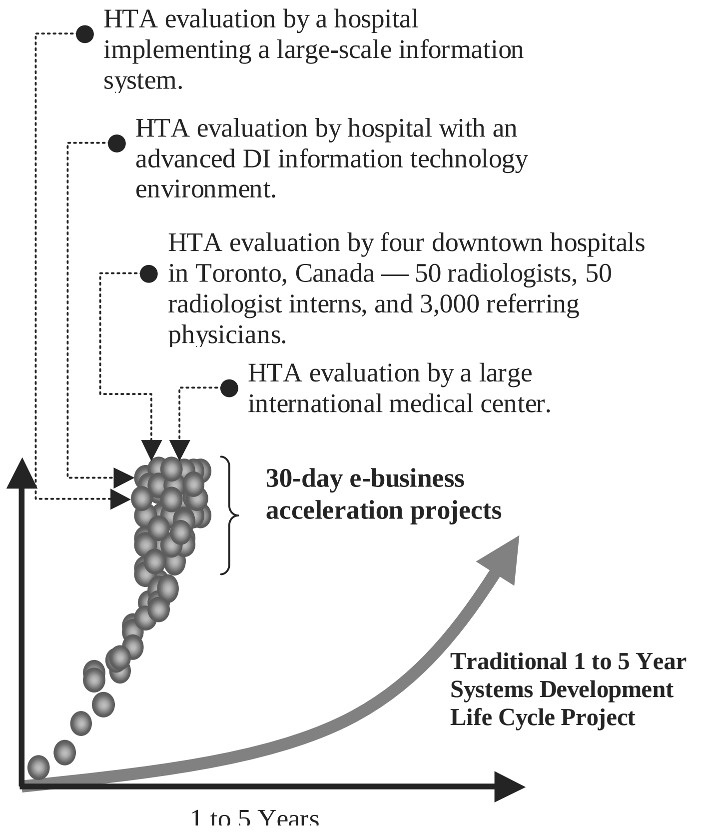
A Standardized Mobile Internet (Wireless) Environment

INET virtual lab also encounters unknown challenges when working with key actors in hospitals such as clinicians, IT departments, medical units, and administration. This is to be expected because each situation is unique and it typically happens during consensus building on HTA proof of concepts and clinical HTA trial programs. The virtual lab uses these challenges to identify, prioritize, and select the requirements for a standardized mobile Internet environment.

Ontario's Management Board Secretariat (MBS) is taking the lead in the use of a Three-Tier Web-Based Architecture, a standardized application environment to implement and scale applications for the Ontario Public Sector. Used as the starting point, this is the foundation to architect a standardized mobile Internet (wireless) environment for an Ontario hospital.

By design, the mobile Internet (wireless) environment is a very thin layer on top of a standardized application environment and represents

Figure 3. A wireless handheld technology applications (HTA) trial program



the extension of the tier 1 presentation layer. The purpose is to prevent time-consuming and costly tier 2 (business logic) and tier 3 (back-end database) development (see Figure 4 on this reading from the bottom up).

INET International Inc. started a wireless technology consortium in 2001 to collect evidence on the best way to define a standardized mobile Internet (wireless) environment. Every member works in collaboration to build a consensus on the IT management practices and IT professional services related to each component. The member's competency is validated through applied research on using mobile Internet (wireless) technology in collaboration with an Ontario hospital. The virtual lab is the basis of INET's applied research. INET uses case scenarios by running selected handheld technology application proof of concepts and trial evaluations. All unpublished data (Goldberg et al., 2002a, 2002b, 2002c, 2002d, 2002e) is available to INET International Inc. collaborations with business and physician-led research centers.

In 2001 the WTC members' roles and responsibilities are listed in Table 2, while Appendix 1 provides the necessary background information on all participating members.

The WTC acceptance process is currently under review. The process may include the following steps:

1. Identify vendor's IT management practice investments.
2. Map IT management practices and IT professional services to a mobile Internet (wireless) environment component(s).
3. INET assigns WTC responsibilities: WTC leaders, complete management qualifications, IT service transition plans and workshops. Build a consensus on a mobile Internet (wireless) technology certification program. Assist IT associations in training.
4. Each member meets or sets international management practice standards (based on ITIL Information Technology Infrastructure Library http://www.itil.co.uk/index.html). They participate in academic conference panels, present papers, and collaborate on applied research.
5. Participate in INET virtual lab proof of concepts and trial evaluations to validate a standardized mobile Internet (wireless) environment.
6. At any time INET International Inc. may reassign a WTC role, request a new member to apply to the WTC, or ask WTC to reengage the WTC acceptance process in step 1.

KEY CHALLENGES

The introduction of wireless into healthcare naturally brings with it many unique challenges which must also be addressed if success is to ensue in these initiatives. One of these challenges is concerned with security and privacy of highly sensitive patient data. In a healthcare setting, in the case of the mobile electronic patient record, typically the physician is accessing patient data from the hospital's Web services. The order information is transmitted is from the mobile device to a base wireless station, and from there, through the mobile communication infrastructure, to the wireless application gateway of the hospital. There are naturally many valid business reasons why all concerned parties—the patient,

Figure 4. A standardized mobile Internet (wireless) environment

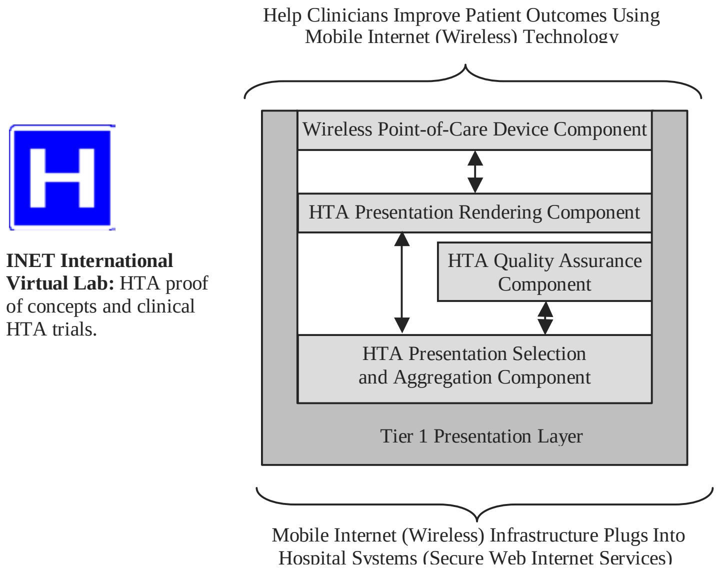


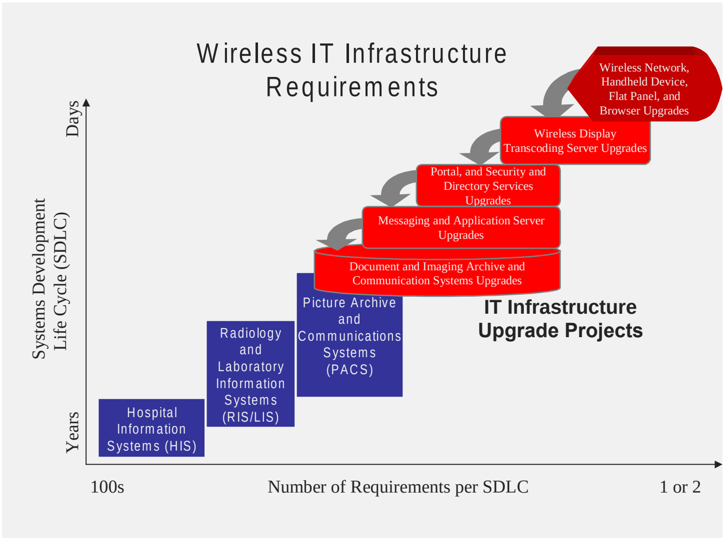
Table 2. Wireless Technology Consortium virtual lab roles and responsibilities

<table>
<tr><th rowspan="2">Wireless Technology Consortium Member</th><th colspan="4">Virtual Lab Roles and Responsibilities</th></tr>
<tr><th>Management Responsibilities</th><th>Provide IT Professional Services</th><th colspan="2">Validate Mobile Internet Technology</th></tr>
<tr><td>INET International Inc. [1]</td><td>Project management</td><td>Collaboration, consensus building, and coordination services to deliver a handheld technology application (HTA) within a wireless healthcare portal</td><td colspan="2">Standardized mobile Internet (wireless) environment</td></tr>
<tr><td rowspan="2">TELUS Mobility[2]</td><td rowspan="2">Service-level management</td><td>Wireless WAN help-desk services</td><td>Wireless IP WAN solution</td><td rowspan="3">Wireless point-of-care device component</td></tr>
<tr><td>Connector help-desk services</td><td>Cable/Card solution</td></tr>
<tr><td rowspan="3">Palm Canada Inc.[3]</td><td rowspan="3">Incident management</td><td>PDA help-desk services</td><td>Device solution</td></tr>
<tr><td>Browser help-desk services</td><td colspan="2" rowspan="2">HTA presentation rendering component</td></tr>
<tr><td>HTA help-desk services</td></tr>
<tr><td rowspan="3">Compuware Corporation[4]</td><td rowspan="2">Release management</td><td>Mobile Internet (wireless) infrastructure deployment services</td><td colspan="2" rowspan="3">HTA quality assurance component</td></tr>
<tr><td>HTA production release services</td></tr>
<tr><td>Availability management</td><td>Application management services</td></tr>
<tr><td rowspan="3">NetManage Inc.[5]</td><td rowspan="2">Change management</td><td>HTA development</td><td colspan="2" rowspan="3">HTA presentation selection and rendering component</td></tr>
<tr><td>HTA enhancement</td></tr>
<tr><td>Problem management</td><td>HTA defect recovery service</td></tr>
</table>

healthcare institution, and physician, to name a few—should expect the information exchanged during such a mobile transaction process to be secure. Specifically, parties should expect the data associated with a transaction to be *confidential* and delivered without violation of *integrity* or *authenticity*. Further, all parties should similarly expect *nonrepudiability*; neither the client nor the hospital should be able to deny the completion of the transaction if such a transaction, in fact, occurred.

Possible security vulnerabilities can occur at several points including: (1) between the physician and the mobile device; (2) between the mobile device and the mobile infrastructure operator; (3) between the mobile infrastructure operator and the wireless application gateway of the merchant; and (4) between the wireless application gateway and the Web services of the hospital. There are similarities and differences between a mobile and a fixed (those not using a mobile device) transaction with respect to security concerns. As in the case of a fixed transaction, *content* of a mobile transaction must be protected from unauthorized access and alteration whether it is in storage or being transmitted. However, protecting data that move is a bit more difficult than data that do not move since one has to worry about parameters that do not come into play in a fixed network, such as (a) location management, (b) mobility management, and (c) radio resource management.

Figure 5. (Source: Wickramasinghe & Goldberg, 2004, pp. 140–156)



Security and privacy are important considerations. In the United States, all healthcare institutions must be HIPAA compliant in this regard. Yet, while HIPAA details what is required, it does not discuss how this should be addressed. We believe that a sound approach to achieve a high degree of end-to-end security in a mobile/wireless environment is to develop a suitable mobile trust model using symmetric keys (Goldberg et al., 2002e). This will not only ensure HIPAA compliancy but will also ensure at least as secure an environment as its wired counterpart. We are confident that these security challenges can be addressed and then healthcare organizations can enjoy the full power and potential of their mobile/wireless solutions.

Security is not the only challenge, however. Other important challenges include compatibility issues and the ability to share information throughout the existing hospital networks. We believe this is less problematic because of the advantages inherent in constructing the healthcare portal. Furthermore, in the designing of the wireless, infrastructure directly builds on top of existing hospital infrastructure as can be seen in Figure 5.

INET's studies (Goldberg et al., 2002a, 2002b, 2002c, 2002d, 2002e) have shown that this approach appears to circumvent several problems pertaining to information sharing. Clearly though incorporating techniques such as information management and knowledge management would serve to enable even more efficient and effective use of all necessary information (Wickramasinghe & Goldberg, 2004). We also suggest that some business process reengineering would be required to ensure that processes are as effective and efficient as possible and make full use of the potentials of the wireless technologies.

DISCUSSION

"Today's hospitals demand zero tolerance for software defects and downtime. That's where Compuware comes into play," says Cathy Lippert, Distributed Systems Product Management Director, Compuware Corporation, Farmington Hills, Michigan. "Our tools can set a new pace for accelerating the testing and quality assurance processes for handheld-technology applications based on mobile Internet standards. This means lower costs, reduced risk, a shorter learning curve, and the confidence that applications can be scaled from one user to thousands of users."

In an effort to improve medical information systems within 30-day cycles and help establish international mobile Internet standards, the WTC is creating a virtual research lab in Ontario. The virtual lab will play a key role in setting new standards for the delivery of high-impact handheld technology applications to clinicians. "The WTC's establishment of a virtual lab designed to help establish global mobile Internet standards should rapidly accelerate the development of handheld applications across a broad spectrum of healthcare industries," says Matthew Hickey, director, enterprise solutions for Palm Canada. "Already, Palm handhelds are at the forefront of this revolution and we expect to continue pushing the envelope in terms of innovation and collaboration." "Our 2.5G iDEN technology allows the seamless transfer of information from wireless handheld devices to hospital servers," says Jawad Shah, manager of Wireless Data, TELUS Mobility, Scarborough, Ontario. "The virtual lab can examine the proper implementation of firewalls, authentication servers, biometrics, intrusion detection, virus protection, and virtual private networks (VPN) to assure that patient-sensitive data can be transmitted securely over the entire TELUS Mobility iDEN network." The preceding quotations from various qualified people in the field are presented to illustrate some key points. First, wireless is here to stay and this technology offers us many advantages. Second, healthcare requires a new way to manage current challenges and thereby provide cost-effective quality care to us all. Finally, the combination of wireless in healthcare appears to be a real solution. Clearly we are technically capable of delivering wireless solutions to healthcare institutions as we have outlined in this chapter. Now appears to be the opportune time to do so. We strongly believe that hospitals of the 21^{st} century should be wireless to some extent at least and a mobile healthcare delivery solution will facilitate the necessary revolutionary changes required by healthcare to contend with today's challenging environment.

CONCLUSION

Wireless technology advancements have made it possible for physicians to use handheld computers to produce diagnostic imaging reports on demand and within seconds. Along with improving access to medical information and the overall speed of healthcare delivery, mobile Internet wireless technologies can play a significant role in lowering IT infrastructure costs for hospitals and preventing disruption to healthcare delivery due to technology problems. The preceding has served to demonstrate that (a) we have the capabilities to offer wireless solutions and (b) wireless solutions will enable healthcare institutions to enjoy many advantages. By detailing the studies conducted by INET, we have shown how doable wireless solutions are in healthcare. Furthermore, we have highlighted the many benefits including cost savings and the ability to deliver a higher level of quality care. We have also endeavored to highlight the key challenges that any wireless initiative must address so that the full power and potential of such an iterative can be realized and thereby truly making mobile healthcare delivery a necessary component of any healthcare arsenal.

We close with a strong message about the importance of thinking of wireless as a necessity in healthcare delivery for the 21st century and we urge for more research in this area as well as for healthcare institutions globally to move forward in this direction. Finally, we also wish to underscore that the lessons learned from INET Inc.'s accelerated project approach have broader implications above and beyond the healthcare domain. Specifically, we believe that the e-business accelerated project techniques are useful to many wireless endeavors and we encourage researchers and practitioners in the field to explore these possibilities.

REFERENCES

Crossing the quality chasm. (2001). *A New Health System for the 21st Century*. Committee on Quality of Healthcare in America Institute of Medicine. Washington, DC: National Academy Press.

Goldberg, S., et al. (2002a). Building the evidence for a standardized mobile Internet (wireless) environment in Ontario, Canada. January 2002 update. Unpublished data.

Goldberg, S., et al. (2002b). *HTA presentational selection and aggregation component summary*. Unpublished data.

Goldberg, S., et al. (2002c). *Wireless POC device component summary*. Unpublished data.

Goldberg, S. et al. (2002d). *HTA presentation rendering component summary*. Unpublished data.

Goldberg, S. et al. (2002e). *HTA quality assurance component summary*. Unpublished data.

Pallarito, K. (1996). Virtual healthcare. *Modern Healthcare, March*, 42–4.

Wickramasinghe, N., & Goldberg, S. (2004). How M=EC2 in Healthcare. *International Journal of Mobile Communications*, *2*(2), 140–156.

Wickramasinghe, N., & Mills, G. (2001). MARS: The electronic medical record system: The core of the Kaiser Galaxy. *International Journal Healthcare Technology Management*, *3*(5/6), 406–423.

Wickramasinghe, N., & Silvers, J.B. (2003). IS/IT the prescription to enable medical group practices to manage managed care. *Health Care Management Science*, *6*, 75–86.

ENDNOTES

- This chapter has been adapted from an earlier version of a paper by Wickramasinghe, N., & Goldberg, S. (2004). How M=EC^2. *International Journal of Mobile Communications*, *2*(2), 140–156.

[1] Corresponding author.

ENDIX I: BACKGROUND INFORMATION ON CONSORTIUM MEMBERS

1. **About INET International Inc.**
   INET International Inc. is a Canadian technology management consulting firm. Currently the firm is focusing on how to use wireless technology to accelerate healthcare delivery improvements and leads a wireless technology consortium. INET International is not on the Web. If you need additional information, please do not hesitate in contacting Steve Goldberg at sgoldberg@sprint.ca, or t: 905 889-2704.
2. **About TELUS Mobility**
   TELUS Mobility, combining the national wireless operations of Clearnet, QuébecTel Mobilité, and TELUS, provides a full suite of wireless services to more than 2.3 million clients across Canada. For more information, please visit us at www.telusmobility.com.
   TELUS Corporation (TSE: T, T.A; NYSE: TU) is one of Canada's leading telecommunications companies, providing a full range of telecommunications products and services that connect Canadians to the world. The company is the leading service provider in Western Canada and provides data, Internet Protocol, voice, and wireless services to Central and Eastern Canada. For more information about TELUS, visit www.telus.com.
3. **About Palm, Inc.**
   Palm, Inc. is a pioneer in the field of mobile and mobile and wireless Internet solutions and a leading provider of handheld computers, according to IDC (December 2000). Based on the Palm OS(R) platform, Palm's handheld solutions allow people to carry and access their most critical information wherever they go. Palm™ handhelds address the needs of individuals, enterprises, and educational institutions through thousands of application solutions. The Palm OS platform is also the foundation for products from Palm's licensees and strategic partners, such as Franklin Covey, Handspring, IBM, Kyocera, Sony, Symbol Technologies, and HandEra (formerly TRG). The Palm Economy is a growing global community of industry-leading licensees, world-class OEM customers, and approximately 170,000 innovative developers and solution providers that have registered to develop solutions based on the Palm OS platform. Palm went public on March 2, 2000. Its stock is traded on the NASDAQ national market under the symbol PALM. More information is available at http://www.palm.com.
4. **About Compuware Corporation**
   With fiscal 2001 revenues of more than $2 billion, Compuware Corporation provides business value through software and professional services that optimize productivity and reduce costs across the application life cycle. Meeting the rapidly changing needs of businesses of all sizes, Compuware's market-leading solutions improve the quality, ease the integration, and enhance the performance of distributed, e-business, and enterprise software. Compuware employs approximately 13,000 information technology professionals worldwide. For more information about Compuware, please contact the corporate offices at 800-521-9353. You may also visit Compuware on the World Wide Web at http://www.compuware.com.
5. **About NetManage**
   Founded in 1990, NetManage, Inc. (NASDAQ: NETM) delivers information access, publishing, integration, and support software and services that maximize a company's investment in existing information systems and applications. It provides an instant bridge to e-commerce. NetManage offers a significantly broader range of application integration software, host access software, centralized management, and live interactive support solutions than competitors. Only NetManage instantly transforms corporate information assets into powerful e-business solutions. NetManage sells and services it products worldwide through its direct sales force, international subsidiaries, and authorized channel partners. For more information, visit www.netmanage.com, send e-mail to pr@netmanage.com, or call 408.973.7171 (Pacific Time).

This work was previously published in Unwired Business: Cases in Mobile Business, edited by S. J. Barnes and E. Scornavacca, pp. 187-204, copyright 2006 by IRM Press (an imprint of IGI Global).

Chapter 6.11
Mobile Business Process Reengineering:
How to Measure the Input of Mobile Applications to Business Processes in European Hospitals

Dieter Hertweck
University for Applied Sciences Heilbronn, Germany

Asarnusch Rashid
Research Center for Information Technology Karlsruhe, Germany

ABSTRACT

There is an ongoing debate about the value of mobile applications for the optimization of business processes in European hospitals. Thus finding satisfying methods to measure the profitability of mobile applications seems to be of great importance. Prior research had its focus mainly on general value dimensions concerning the medical sector or the usability and design aspects of hospital information systems. Conterminous to that, the authors chose a strictly process-oriented approach. They modeled the requirements of future mobile systems as an output of a profitability analysis based on activity-based costing. The cost savings defined as the difference between former and future business processes were used as an incoming payment for an ROI analysis. In a nutshell, the authors present a case study that highlights the value of their analyzing method as well as the enormous benefit of mobile applications in the area of food and medical supply processes in German hospitals.

INTRODUCTION: INCREASED DEMAND FOR EFFICIENT PROCESSES IN HEALTH CARE

Apart from the long-term decline of the population, a great challenge in the contemporary discussion turns out to be the increasing aging of the population in European industrial societies. This raises various difficulties for our welfare systems and reveals the necessity of long-term adjustment to this development. Aging describes the process of composition of the population shifting for the benefit of elderly people.

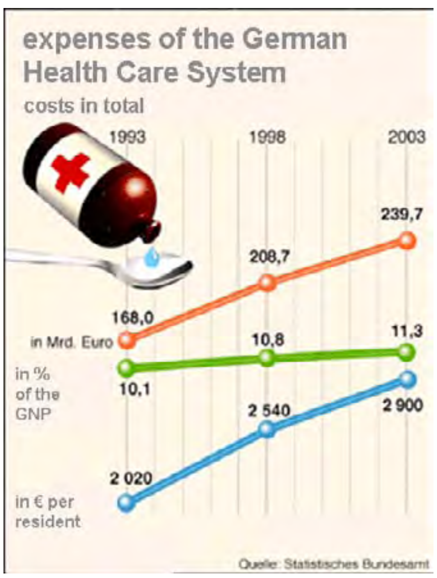
Thus, the decisive item is not the increasing number of the elderly but rather their increasing proportion of the population. For example as latest simulations for the development of the German population (Statistisches Bundesamt, 2003) reveal, the proportion of 65-year-old and older people will rise from 17.1 % today to 29.6 % in 2050. At the same time the percentage of geriatric people (80 years and older) will increase to 12 % which means a triplication.

This development causes serious problems in welfare and tax systems that are based on the income of a workforce. Less young people have to pay the pensions and health care of the elderly.

Furthermore, the productivity of our highly automated industry leaves an increasing number of people unemployed. So the real challenges of over aged European industrial societies will be to enhance the productivity of the existing education and health care systems.

And as productivity is defined as the relationship between output and input factors, there was an intensive discussion going on during the last 2 years about the input factor dimension. Even

Figure 1. Expenses of the German health care system



though the German health care system was able to perform quite well the last decades, from an input point of view the costs and resources to maintain the system were increasing dramatically (see Figure 1).

This development is not typical for the German health care system only; you will find similar developments in all Organisation for Economic Co-operation and Development (OECD) countries around the world as published at the OECD fact book (OECD, 2006).

One major initiative to stop this cost explosion was the release of a law for financing the clinical sector in 2002. According to this law, a hospital does not get paid for the duration a patient is being treated, but for the respective type of disease. The treatment of every illness is linked to a fixed price—documented in the Diagnosis Related Group (DRG)—the hospital is then paid by the health insurance companies. This system results in the effect that a hospital can only earn good money by improving the business processes in the treatment of the patient, without loosing quality. Since this time on, a big competition between hospitals and different clinical departments to enhance there business process productivity has taken place. Best practices achieving business processes improvements are defined as clinical pathways.

As a parallel to other new deregulation decisions invented by the government, new types of market players like, for example, the Rhön Clinical enterprise emerged. They act as business redeveloper, buying unproductive hospitals and now, by standardizing and optimizing their processes in relation to given DRGs to transfer them to profitable businesses.

This development has only just begun, but the trust in the potential of the business process optimizer is still unbowed, if you take a look to the 3 year curves of their stocks (see Figure 2).

Further potentials in optimizing business processes in the clinical sector are dependent on several issues like:

Figure 2. Chart of the Rhön clinical enterprise

- Deployment of best practices
- Availability of new technologies that enable high qualified staff to perform the processes in a new way
- Permanent will to benchmark and improve existing business processes

If you look on the infrastructural dimensions of business processes in hospitals today, you can still find a lot of weaknesses linked to the traditional clinical organization, like:

- A very low degree of computerization for support and administrational processes, deeply based in history where traditional administration leaders of the hospitals were not very powerful in comparison to their medical colleagues
- A very poor degree of existing IS integration between different hospital departments, which still causes a lot of medical problems and costs
- A very traditional hierarchical organization with powerful medicinal stuff, but seldom trained in economy
- A very restrictive type of regulation based on the influence of different interest groups and their negotiations in the past

On the other hand, good medical work needs a lot of different information just in time and close to the location, where the patient gets their treatment. From this point of view, one major success factor in business process improvement in the future will be based on technologies that are supposed to be able to transport a big variety of information from different data sources to the medical employee while they are treating their patients, analyzing their physical conditions, or support them with meals and medicine. This is a major reason why the elaborated use of mobile applications in hospitals will be essential for their future success in business

The focus of this paper will be showing you the economic potential of mobile applications, especially their capacity in leveraging the performance of supporting processes like meal and drug supply and how to measure these benefits.

BACKGROUND: THE ECONOMIC POTENTIAL OF MOBILE APPLICATIONS IN A CLINICAL ENVIRONMENT (RELATED WORK)

As mentioned before, it is perhaps, supposed to be a question of organizational survival if the hospitals in European countries are able to decrease there costs far to a level of where they are today. We also argue that only the invention of new, innovatively designed business processes will lead to this target.

But to what degree will and can mobile applications be part of value-added business processes? And how can we measure this value?

If we look at the related literature, we will find two different kinds of studies.

Studies that Deliver Theoretical Basics and Frameworks to the Topic of Value Generation of Mobile Applications

A good example for this kind of research was published by Nah, Siau, and Sheng (2005). Following their definition, value can be defined as: "the principles for evaluating the consequences of action, inaction, or decision" (p.85). The output of their study was a procedure modeled in Figure 3, which was meant to help creating a means-ends objective network including a distinction between basic and fundamental objects as shown in Figure 4.

The methodological way to gain these networks was based on interviews between researcher and employee in the field. A major result can be seen in Figure 4. The strength of this kind of research surely is the identification of relevant factors, which reveals the possibility of enhancing productivity of mobile applications for the whole enterprise.

On the other hand, their weakness can be seen in the fact that efficiency of an application system can only be measured in relationship to a supported business process, its core activities, and the output produced. A second weakness of their approach was more methodologically. The productivity model of the resulting means-ends objective network of the application system might be based on subjective cognitions, given by the interviewee during the interview situation, instead of time and cost savings measured in a real-world user scenario. A measurement of saved costs between the initial business processes and the redesigned one is not addressed.

Studies Focused on the Business-Oriented Design of Mobile Applications

This approach focuses on the question, what are the basic requirements or architectural patterns that drives the value of a mobile application system in relationship to supported business processes. In addition we can divide this related work in studies that start from an existing business process that

Figure 3. Procedures of value-focused thinking

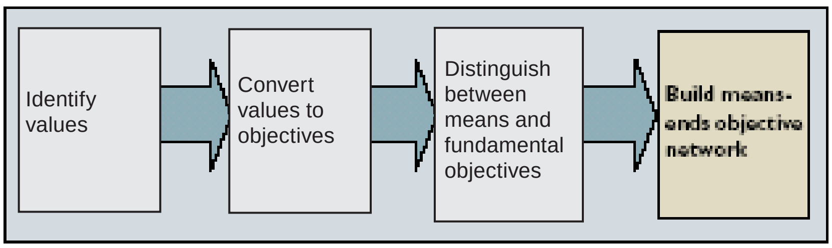


Figure 4. The means-ends objective network

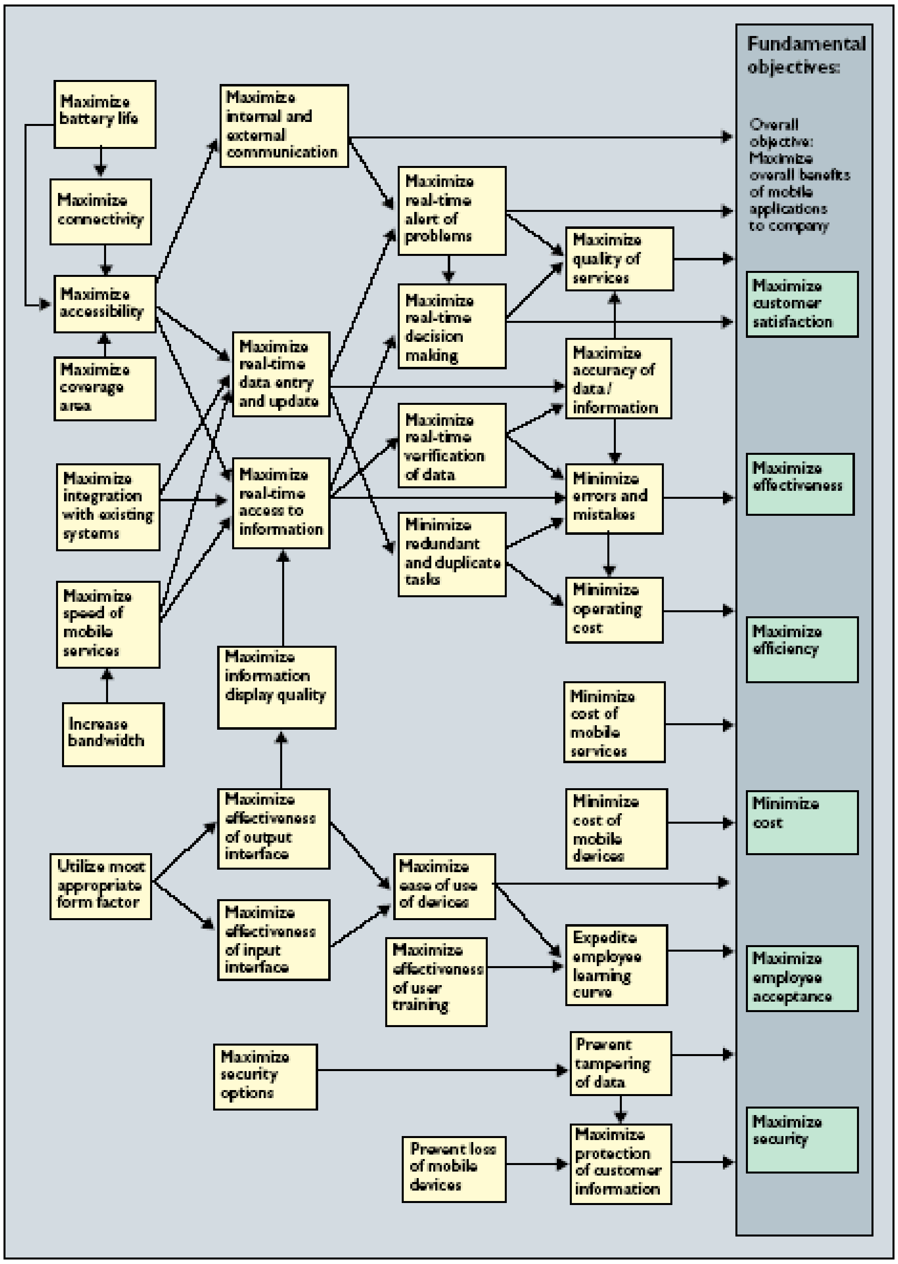


Figure 5. Delivered services by a mobile application (Shiffman 1999)

Delivered Service	Service Description
Recommendation	the determination of appropriate guideline-specified activities that should occur under specific clinical circumstances
Documentation	the collection, recording, and storage of observations, assessments, and interventions related to clinical care
Explanation	the provision of background information on decision variables and guideline-specified actions (e.g., definitions, measures of quality or cost) and the rationale that supports guideline recommendations, including evidence and literature citations
Presentation	the creation of useful output from internal data stores
Registration	the recording and storage of administrative and demographic data to uniquely identify the patient, provider(s) and encounter
Communication	the transmission and receipt of electronic messages between the clinician and other information providers
Calculation	the manipulation of numeric and/or temporal data to derive required information
Aggregation	the derivation of population–based information from individual patient data

has to be improved, and into studies that derive mobile application systems from target processes. There are various studies in the computer supported cooperative work (CSCW) area, which try to enhance the performance of existing business processes by supporting them with fitting mobile system architectures.

A typical research work in this area was published by Shiffman et al. (1999). He shows the benefits of different services (see Figure 5) delivered by a mobile application to a screening workflow for asthma deceases in child care. They described what kind of information management services should be delivered by a pen-based mobile application system to support decision-making workflows in a hospital. Although they get very deep into the interrelationships between system functionalities and the quality of business process support, they did not measure the resulting influence of the system in cost, time, and quality.

Another study from a more transaction network design perspective was published by Morton and Bukhres (1997). It focuses on the improvement of existing transactions in an ambulance scenario. Morton and Bukhres developed a network architecture that enables a hospital to use mobile applications successfully in the ambulatory service. A major challenge in this scenario is the fact, that a mobile host, for example, in an ambulance car could not be online all the time. So they developed an architectural solution, based on a so called base station agent (BSA) that is responsible for the monitoring of transaction of the mobile host during its execution. This architectural innovation developed from existing business processes by interviewing experts and measuring transaction time, which enables the ambulance to use mobile applications for accelerating their transaction times—a very successes critical benchmark in the area of lifesaving.

Different to Shiffman (1999), and Morton and Bukhres (1997), Wang, Van de Kar, and Meijer (2005) published a study that focuses on a mobile application system design, derived from a future target process. Their design methods are based on the collaborative business engineering (CBE) method from Hengst and De Vreede (2004), and tested for the design of a mobile online information system based on a PDA that supports the conductor on the railway delivering actual information and services to railway customers on the train. The output of their kind of research was a very good system design, derived from the requirements of the new, targeted business processes. Although their research design delivers the opportunity to do a business process benchmark between the performance of the current and the target process, an economically motivated measurement of the improvements was not operated in that research study either.

Conclusion: The Integrated Approach

Recurring to related researches as mentioned before, we were looking for a third way that started from an approach as published by Wang et al. (2005) for the necessary system design but also was combined it with a classical activity-based costing analysis. The aim of our approach was to gain data, which enabled us to measure the efficiency of the new designed and mobile application supported business processes. We found a suitable method named mobile process landscaping (MPL) developed by Köhler and Gruhn (2004) that supported our research in the mentioned way. MPL includes eight steps shown in Figure 6.

However, based on our long-years experience in the area of business process modeling at the Research Centre for Information Technologies (FZI) and a conclusion drawn in the unpublished doctoral thesis of Högler[1] (2006) we improved the MLP method and adapted it to the mobile business process development and profit analysis of a hospital nursing 300,000 patients a year.

We executed the analysis in the five major steps described in Figure 7:

Figure 6. Major steps of mobile process landscaping (MPL)

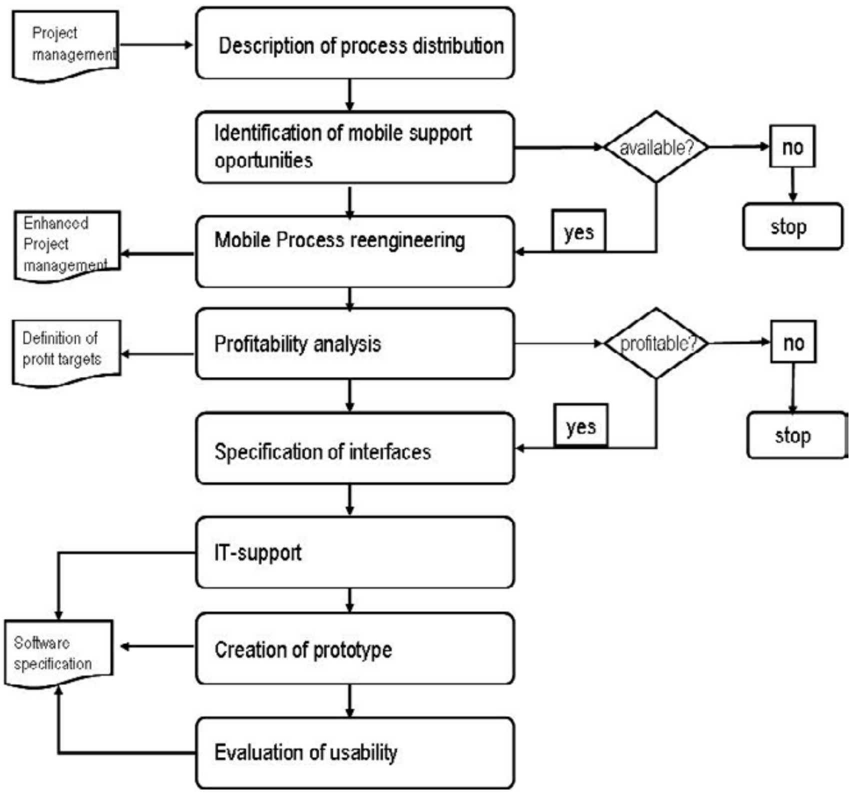
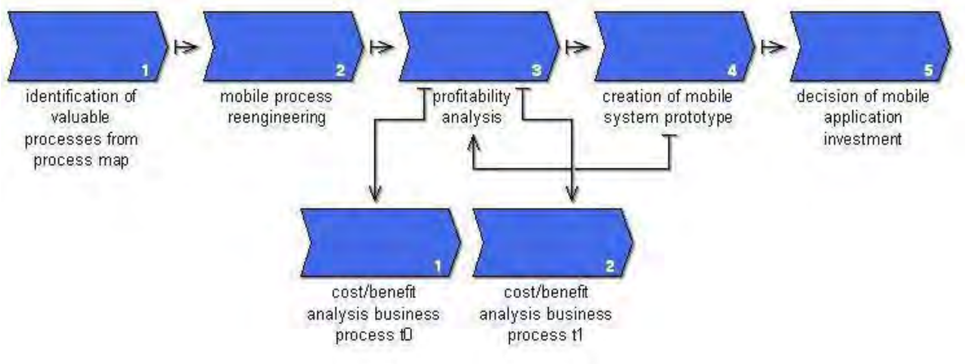


Figure 7. FZI mobile process reengineering approach



Step 1: Identification of highly valuable business processes in the business process map of the hospital (see Figure 8).

Highly valuable processes should have a significant cost-saving potential as well as a high degree of distributed activities, so that mobile application support may deliver a remarkable benefit. In our case these processes were easily to be identified, as they were delivered from external service providers, too. Benchmarks for meal and medical supply process performance were far ahead from the performance delivered in the observed hospital.

After quick-win business processes in the area of medical and meal supply had been identified as valuable, and the strategic decision made by the management to still internally deliver these services in the future, a second step in research had to be done—the analysis of the mobile potential of the identified supporting processes as well as the reengineering potential on an activity level (see Figure 7).

Step 2: Identification support opportunities on an activity level and the reengineering potential

The analysis of mobile potentials can be operated in two ways:

a. The first one will be *the identification of activities in an existing process, which could possibly be supported by mobile applications (improvement approach).* Typical indicators for supportable activities are:
   - the necessity of having access to actual information at a point work that is not yet connected to a network (information quality and delivery function), and
   - the necessity to improve existing data input procedures by delivering high qualitative information back to an integrated hospital information system electronically instead of manually (service quality function).

b. The second one will be *to reach predefined outputs of a needed service with totally new, less cost-intensive mobile business processes (reengineering approach).*

Especially the design of totally new business processes supported by mobile technologies in regard to their later implementation efforts, costs, and maintenance quality is a central research task the FZI has been working on for years now.

The application of steps 3 to 5 of the FZI Mobile Business Process Reengineering Approach

Figure 8. Business process map of the hospital

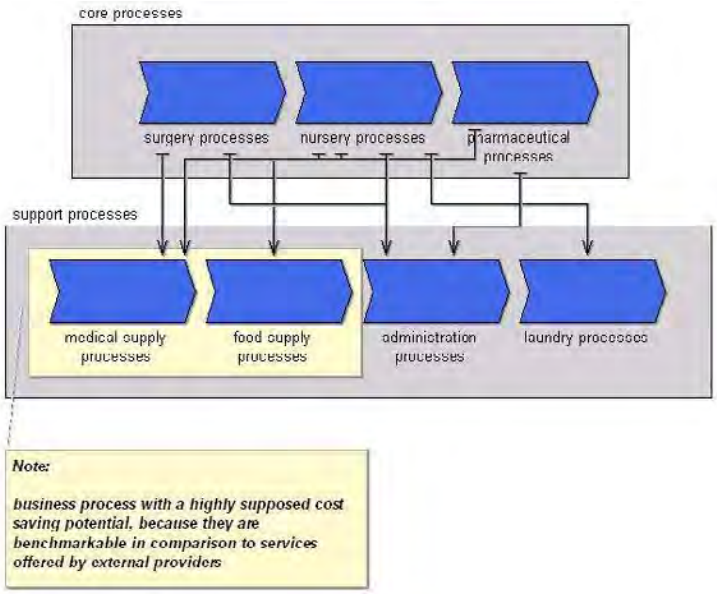


Figure 9. Example of a business process with reengineering potential

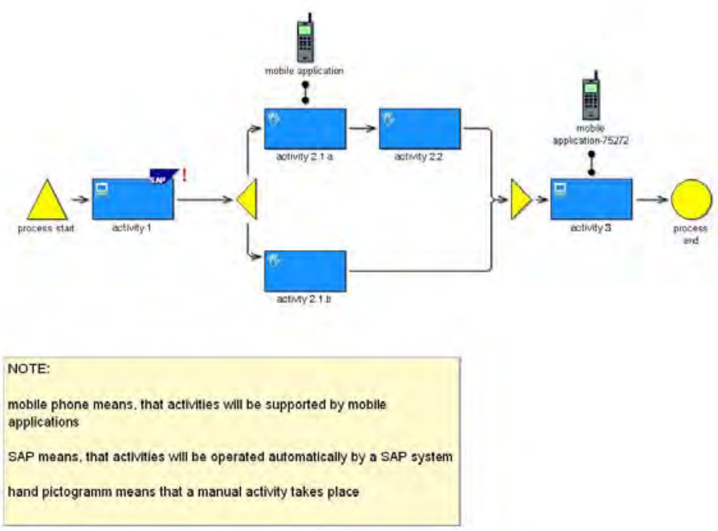


Figure 10. Comparison between conventional and focused ethnography

Conventional Ethnography	Focused Ethnography
Long-term field visits	Short-term field visits
Experientially intensive	Data/Analysis intensity
Time extensity	Time intensity
Writing	Recording
Solitary data collection and analysis	Data session groups
Open	Focused
Social fields	Communicative activities
Participant role	Field observer role
Insider knowledge	Background knowledge
Subjective understanding	Conservation
Notes	Notes and transcripts
Coding	Coding and sequential analysis

(MBPRA), shown in Figure 7, will be described in the case study in chapter 3. It includes the operation of a profitability analysis, and the design of a mobile system prototype fitting into the reengineered business processes. Once the prototype will be able to reach serious savings of activity-based costs, the investment decision for the implementation of the whole system has to be made.

A major focus in our case study will be regarding the question of data quality, which is necessary for modeling and simulating business processes in a way that allows the delivery of detailed information about their cost structure and process performance.

CASE STUDY: MOBILE TERMINALS FOR DRUG AND MEAL SUPPLY

This study was carried out from January until June 2004 at FZI by order of Karlsruhe City Hospital, the largest hospital under public law in the vicinity of Karlsruhe with more than 1,500 beds and 14 clinical departments, handling approximately 300,000 patients per year. The departments are allocated campus-like to numerous buildings. In this study the hospital's processes of drug supply in the pharmacy of the hospital as well as meal supply in the kitchen were examined. Key stakeholders were nurses, physicians, pharmacists, assistant medical technicians (MTAs), assistant pharmaceutical technicians (PTAs), cooks, cooking assistants, diabetes consultants, haulage service, and the controlling and executive committee.

Used Methodology

As every model of reality is as good as its input, our profitability analysis of mobile business processes was highly dependent on the data quality. As long as time for delivery or processing of activities was supposed to be measured or informal stocks at the wards to be detected, it does make sense to use a mixture of traditional ethnographic and focused ethnographic approach (see Figure 10) as claimed by Knoblauch (2005). In our role as researchers we were involved as participants in the working process, which is not the case in focused ethnography, where the researcher normally acts as an observer. But all the other techniques we used were based on the ethnographic approach of Knoblauch.

The challenge was to gather high qualitative data and transfer it to linear cost models, like Lazarsfeld, Jahoda, and Zeisel (1933) practiced it in their early studies. We used the following data gathering methods:

- **Participating observation:** we worked together with nurses and a druggist as well as kitchen and health care staff for 8 days and night shifts to get a deep understanding of the business processes including the used resources and materials.

- **Inventories:** Inventories of the formal and informal medical stocks at the drug store and the wards
- **Inside interviews:** 34 non-closed expert interviews with different process owners in the hospital (nurses, pharmacist, IT operators, IT management, IS consultants, cooks, diet cooks, garbage man)
- **Outside interviews:** 4 non-closed expert interviews with experts from other hospitals that had just invented mobile systems in the same processes to validate their own gathered data.

Initial Situation/Preface

Previous to this study the hospital was faced with the choice of an investment and integration of mobile devices on wards and pharmacy. These processes were chosen from the process map, because they promised a high potential for optimization. In 2003 the project team "Mobile Computing" was founded by the head of the pharmacy, the chief information officer, and the chief executive officer. In a pilot project they started to develop and test a mobile system in the pharmacy and on two selected stations (oncology and nephrology, each with more than 20 beds) in cooperation with a software company. The aim of the Mobile Computing project was improving the drug supply by reducing administrative activities just as avoiding sources of error by the dint of mobile terminals and their integration into the supply workflow. At the time the study was performed, the Mobile Computing project was in the productive test phase. In case of a positive result, it was planned to roll-out the system on all stations. The study's objective therefore was to provide a basis of decision making about the roll-out by dint of an economic analysis.

In order to examine the portability of the mobile system the study was made up of two scenarios. The first scenario dealt with the processes of drug supply, where, within the scope of the Mobile Computing project, the possibility of comparing past and new situations in a real environment was given. In the scenario two processes of the meal supply were chosen to demonstrate the ability of the mobile system being extendable. In contrast to the scenario, one of the analyses was limited to the investigation of a business process without any mobile devices due to the fact that there was no test implementation of a mobile system for the meal supply. The analysis of the future processes (with mobile devices) had to be designed with experiences gained from interviews of other hospitals using mobile devices for meal supply already.

By implementing the mobile system, traditional paper-based processes were supposed to be adopted and optimized where exchanging information should be converted from paper-based to electronic-based processes as far as possible. Improvements in workflow by automating administrative jobs (e.g., validity check, sorting of forms, calculating order quantity, etc.) and minimizing sources of error were expected.

In scope of the drug supply the orders were previously taken at the station's PC or at the pharmaceutical rack using PDAs and sent to the enterprise resource planning (ERP) system and the pharmacy's IS via docking station. In the pharmacy, the orders could be processed directly on the PDA by a wireless LAN (WLAN) without the necessity of using paper-based media.

At the meals supply process the paper-based order form was displaced by the digital entry via mobile devices. If necessary, the chosen data could be linked to the patient's incompatibilities or objections. The software was to check up on all data on their plausibility automatically and to send alerts in the case of incomplete or false orders. In the kitchen, all steps for receiving and handling meal orders were automated. Based on the demand of ingredients, the kitchen's staff sent all orders to the supplier. Additionally, the patient nameplates were printed and cut to mark the meal tablets.

Objectives

Of highest interest in this study was the economic potential that could be gained by implementing mobile terminals in a hospital whereat the profitability of the mobile system for the chosen processes for the supply of drugs and meals described previously had to be measured. For this reason processes of the past and future situation were to be documented and compared in order to ascertain possible profit and loss.

Another target was the development of a methodology for the evaluation of economic potentials achieved by the integration of mobile devices into clinical and economical pathways as well as for its testing. For this purpose the general conditions for the use of mobile devices in hospitals had to be considered and the applicability of different kinds of mobile devices to be detected. Another question to be answered was where and how mobile terminals could be integrated in existing business processes. The analysis of advantages and risks that could occur during the roll-out of the mobile system was targeted as well.

Central questions that had to be addressed in the study were:

- How could costs and efficiency of supply processes be measured (with and without the support of mobile devices)?
- What advantages and risks could be expected during the roll-out of mobile systems?
- What profit and loss could result from implementing the mobile system?

Scenario 1: Drug Supply

In the hospital all medicaments like drugs, infusions, and so forth were delivered by the hospital's own in-house pharmacy. The pharmacy is in charge of ordering, producing (in special cases), storing, and delivering drugs to the hospital stations. Pharmacist, MTA, PTA, and warehouseman have to work together closely. The pharmacist is responsible for quality control and consulting services in case of questions about drugs (e.g., compatibility, unlicensed drugs, etc.) for the hospital's physicians and nurses. Furthermore he/she has to determine the demand for drugs and in special cases produce special pharmaceuticals and infusion bags. Usually, wards order drugs every day.

Former Situation in Drug Supply

The processes of the former situation in the hospital's drug supply are presented in Figure 11. After the prescription had been inserted into the patient record by the responsible doctor, nurses ordered lacking medicaments by an order form. Before that, they had to walk along the medicine shelves on their wards and write down name and amount of the needed medicaments on a notepad. Afterwards they went over their notes and transferred them on the order forms mentioned previously. Due to the numerous classifications of medications (usual medication, infusion, dispensing, cytostatica, anesthetic, etc.) there were eight different types of forms available. The order form had to be signed by the responsible doctor and transported to the pharmacy finally by being thrown in the pharmacy's letter box, faxed, or in urgent cases by using pneumatic post.

In the pharmacy, the pharmacist first sorted the incoming order forms according to their type of form and to their posting station and checked the details on the order forms. Detected discrepancies had to be straightened out by telephone. In case of missing compulsory data (e.g., signature of the doctor) order forms were to be sent back to the station. In case of valid details drugs were collected and put in boxes together. Before delivery the drugs were registered in the ERP system of the pharmacy by scanning the barcode. The so called "Hol- und Bringdienst" ("catch and delivery service") performed the delivery to the stations. Additionally, nurses are able to fetch ordered drugs at the pharmacy counter.

Information was transmitted by paper-based information media like order forms and notepads as well as by fax, letter box, pneumatic post, telephone, and the staff itself (nurses, pharmacist).

Weak points in the former situation could be identified in several processes. Inefficient activities were the sorting of the order forms, queries by telephone because of non-explicit or uncompleted details, the recording of the anesthetic, output, and sale to hospital staff member as well as orderings from drug maker and wholesaler.

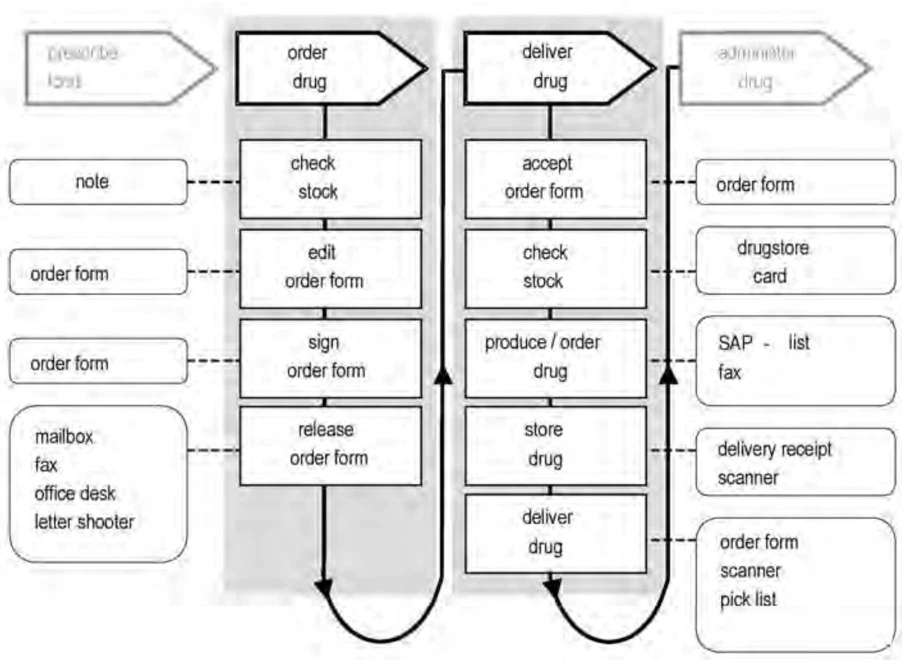
Using pneumatic post (ca. 40-60 times per day) required several time-consuming activities such as loading and unloading the sleeves. Furthermore the pneumatic post system often revealed little reliability and additionally caused high costs for maintenance (see Figure 12 and 13).

The material costs contain those for paper and printing, order forms, occupancy costs for archival storage of order forms, and of thrown away drugs because of the drugs were out of date/expired. The term factor represents the frequency of the respective activities per day. Costs can be calculated by multiplying labor costs by factor.

Present Scenario with Mobile Devices in the Drug Supply

In the present scenario processes of ordering drugs (see Figure 14 and 15) are supported by PDA (personal digital assistant) and by a Web-based order entry system. Nurses now walk along the drugs shelves on their station and fill out the order form directly on a PDA without any use of notepads. With the barcode scanner extension nurses just need to scan the barcode of the drug's package and choose the amount by pull-down menu. In order to save time, the barcodes labels itself are stuck on the shelves' front side, so that there is no more need of taking the drug packages out. Via docking station at the station's computer the order forms are directly transmitted to the ERP

Figure 11. Processes of the drug supply in the hospital (former situation)



system of the pharmacy. A Web front end enables nurses to check, correct, and activate the order. The responsible doctor merely has to check the Web-based order form and sign with his/her personal password to ratify its validity.

The team of the pharmacy now gains access to the order form in an electronic way. By computer and PDA they can view and work on all order forms on the display. Due to the pharmacy's WLAN connectivity the staff has ubiquitous access without any need to sort the order forms. The staff can walk along the shelves of the pharmacy and collect all ordered drugs. By scanning the barcode on the drug package the delivery items are automatically registered.

Review of the Process Improvements in the Drug Supply

The mobile system offered numerous forms of relief and advantages. It is possible to remedy the deficiencies of the former situation without IT support and to avoid disadvantages of stationary computer systems. Communication costs (telephone conversations) between pharmacy and station could be limited by eliminating unreadable handwriting and incomplete details. In the pharmacy numerous activities could be automated or usefully supported.

Because of several circumstances the field of application was limited in some issues. For example, ordering and delivering anesthetic had to be excluded from the mobile system, as dealing with anesthetic had to use paper-based documentation by statute. Another area where the mobile system was discussed to be applied was the pharmacy's stock control. But as the use of drugs could not be registered exactly (e.g., returned drug packages, which were partially used and no more needed), improvement of stock control by the mobile system was not possible either, nor was it to be expected that the amount of drugs that had to be disposed for exceed of expiration date could obviously be reduced.

Scenario 2: Meal Supply

In the hospital there exists one kitchen that supplies personnel, patients, and their dependants with meals. Patients get meals three times per day. One day before each meal, for example, lunch, nurses on the stations ask patients about their wishes for lunch the next day and transmit the orders to the kitchen. Thus, the kitchen's staff is able to plan and organize every meal up to 24 hours before.

Former Situation in the Meal Supply

Without the implementation of a mobile system, the supply of meals is a slow process (see Figure 16), afflicted with different kinds of errors. Orders are taken on the basis of menus and forwarded to the kitchen via meal vouchers. In the past these vouchers had been sorted, checked superficially, and afterwards were read in with a special PC-linked scanner. The associated software counted the orders per station and furthermore checked the vouchers' plausibility. Having gathered all orders the production schedule for the next day was printed and handed out to the cooks.

Future Scenario with Mobile Devices in the Meal Supply

On the basis of a new mobile system, the traditional paper-based processes can be adopted and improved (see Figure 17). The orders will be taken at the station's PC or directly at the patient via PDA. The name of the patient is chosen out of a patient administration system, linked to the chosen meal and via PC or docking station sent to the ERP system and the kitchen. If necessary, the chosen data can be linked to the patient's incompatibilities or objections. The software will check all data on their plausibility automatically and send alerts in case of incomplete or false orders. Additionally, it will check the patient's administration software in order to verify which

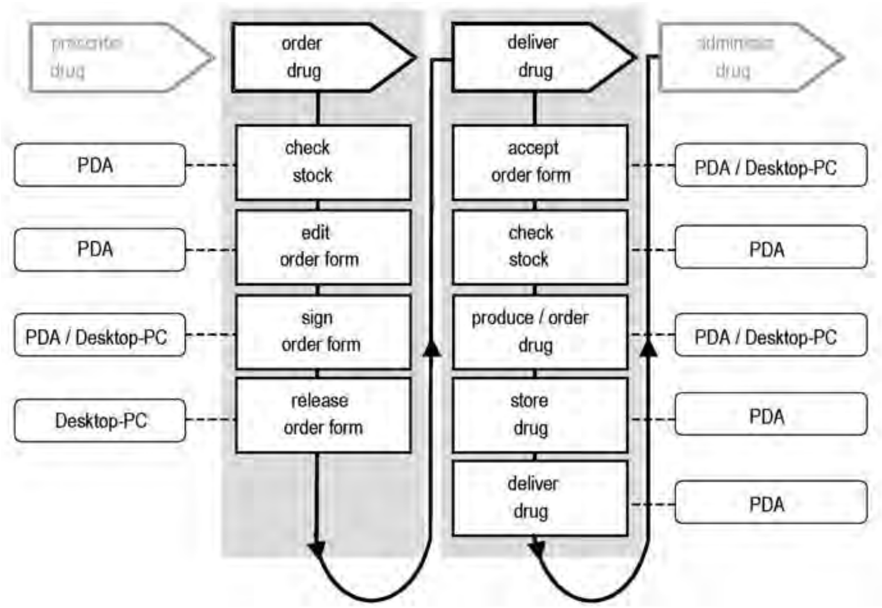
Figure 12. Cost of material for drug supply (traditional supply system)

	amount	unit price	cost of operation/per year
Drug documentation form	14064	0,0636 €	900 €
Archiving costs (drug store)	15 m²	72 €	1.600 €
Drug expiry costs	1	4.770 €	4.770 €

Figure 13. Personnel cost of drug supply (traditional supply system)

Drug order (each station)	duration [min.]	factor	costs [per day]
Fill out order form	6,36	1	3,08 €
Subscription by doctor	1,79	1	0,86 €
Walking to the mail box	2,18	1	1,05 €
Prepare pneumatic post	1,07	1	0,52 €
Edit returning form	0,45	1	0,22 €
Total costs of drug order (per station)			5,73 €
Drug delivery (pharmacy)	duration [min.]	factor	costs [per day]
Empty mail box and pneumatic post	64,11	1	51,73 €
Sort order forms	17,14	1	6,77 €
Correct orders, making inquiry calls	31,57	1	19,82 €
Store pick lists	9,00	1	3,80 €
Writing and attaching notes	2,39	1	1,50 €
Automated order	8,50	1	3,59 €
Restructure drug cards	12,00	1	5,06 €
Accounting of narcotics	0,50	6,31	2,63 €
Check of the narcotic accounts	30,00	1/30	0,83 €
Accounting of the sales to clinical employee	60,00	1/7	3,62 €
Data entry of Zytostatika in PC	90,00	1	75,00 €
Zytostatika-data exchange with ERP system	45,00	2/30	2,68 €
Register orders for wholesaler	14,00	1	11,67 €
Dictate wholesaler orders to secretary	20,00	1	16,67 €
Empty stock from employee drug sale	5,63	1/7	0,34 €
Additional efforts of employee drug sale	30,00	1/7	1,81 €
Preparation of order forms	2,03	1	0,80 €
Register industrial orders	10,79	1	4,56 €
Archiving of clipboard	27,22	1	11,49 €
Managing call backs of single charges	46,21	1/7	5,50 €
Total costs of drug delivery (pharmacy)			229,85 €

Figure 14. Processes of the drug supply in the hospital (present situation)



patients have left the hospital or have changed the station.

Review of the Process Improvements in the Meal Supply

These meal vouchers as well as false orders and insufficient deliveries emerge as the most important cost drivers within the old system. For example, it takes up a great deal of time to complete the meal vouchers, to arrange and to correct them as well as to import the vouchers into the system; errors emerging are redundant or needless orders that can not be cancelled. Further cost drivers are coordinative telephone calls between kitchen and hospital stations that have to be ascribed to the relocation or early release of patients, for example. 80-100 telephone calls per day are necessary due to short-term changes of orders.

These cost drivers add up to approximately 720,000 € each year for ordering activities between the kitchen and all stations of the hospital. This sum includes material costs like order vouchers and wasted meals due to mistakes as well as personnel costs for placing and taking the daily order.

With the help of the mobile system described previously, many of the former cost drivers in meal supply could be eliminated. Having implemented the mobile system, further costs for materials will only arise from the production of tablet cards and non-preventable false deliveries of meals. Non-preventable situations are, for example, when it is not foreseeable after which meal a patient will be released the next day or whether a meal can be taken after a surgery.

In Figure 18 and Figure 19 the costs accruing in the kitchen due to the old IS are listed.

The costs for materials compound of actualizing the software and the contract for maintenance of costs for paper and those for spare meals which can be ascribed either to preventable or non-preventable wrong orders.

Aside from retrenching working expenses, the mobile system also provides an obvious surplus in quality. In contrast to the former system, the patients' incompatibilities like, for example, allergies or possible objections can now be considered with much less effort. Figure 20 shows the personnel costs of the meal supply supported by the mobile system.

Figure 15. Cost of personnel for drug supply (present situation)

	duration [min.]	factor	costs [per day]
Drug order (each station)			
Fill out order form	2,00	1	0,97 €
Synchronize PDA/Scanner	1	1	0,48 €
Confirrm order	2,00	1	1,16 €
Subscription by doctor	1	1	0,82 €
Edit returning form	0,07	1	0,03 €
Total costs of drug order (per station)			2,98 €
Drug delivery (pharmacy)			
Print out pick-list	1	1	0,40 €
Send messages (e.g., quries, information)	0,60	1	0,31 €
Automated order	1,43	1	0,60 €
Accounting of narcotics	1	1	0,83 €
Check of the narcotic accounts	1	1	0,42 €
Data entry of Zytostatika in PC	10,00	1	8,33 €
Register orders for wholesaler	4,00	1	3,33 €
Register industrial orders	9,25	1	3,90 €
Archiving of clipboard	5,00	1	2,11 €
Managing call backs of single charges	8,00	1/7	0,95 €
Empty stock from employee drug sale	3,00	1/7	0,18 €
Synchonize PDA	1	5	2,56 €
Total costs of drug delivery (pharmacy)			23,94 €

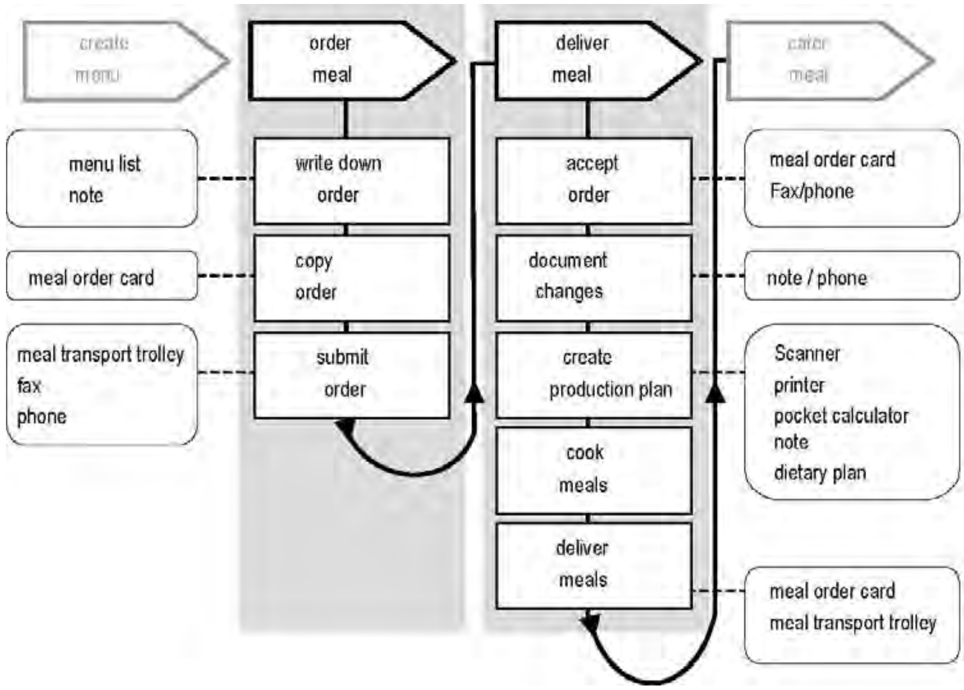
Economic Results

Process improvements were performed iteratively in several steps. During the process improvement, ideas and visions of staff members were considered and the experience gained in cooperative hospitals as well as in Mobile Computing projects in other lines of business were taken into account. Approved methodologies and rules for process improvement by Greiling and Hofstetter (2002, p. 96 et sqq.) were included. The profitability was calculated by the *net present value* (NPV) method (see Figure 21), where n means the time horizon of the project and t numeralizes the years. The term *discount rate* refers to a percentage used to calculate the NPV and reflects the time value of money at an actual average between 3% and 6%.

By dint of the methodology current and target processes of the traditional and the mobile system were modeled and documented, their costs and use in terms of activity-based costing were determined and compared within the scope of a profitability analysis.

The costs for the optimized drug and meal supply are scheduled in Figure 22. Just as in the drug supply, the one-of expenses for the mobile

Figure 16. Processes of the meal supply in the hospital (former situation)



system comprise of costs for investment and the non-recurring operating expenses. There is also a strong emphasis on staff's training activities to guarantee a maximum acceptance by the staff and a smooth IS implementation.

The investment into the mobile system and the scanner alternative gets profitable at least within one year (see Figure 23 and Figure 24), based on the assumption that operation expenses are lowered strongly and risks narrowed down. The discount rate is chosen on 10% by an initial investment of 200,000 €. In contrast the payback period is reached within 3 years time, if the implementation of the mobile system is not followed by an appropriate reengineering of the relevant processes. These facts can be ascribed to the integration of autonomous subsystems by mobile devices as well as to the removal of errors. Besides, parallel processes result in the decrease of labor time. Additionally, the mobile devices enable the controlling department to gather more and reliable data for the measurement of costs.

CONCLUSION

In comparison to the big amount of money, the European telecommunication industry has invested in Universal Mobile Telecommunications System (UMTS) and third generation (3G) mobile infrastructure; there still is a very slow adoption of mobile application in different domains, which might result from a lot of different reasons. Two major reasons surely apply: (1) lack of methods to demonstrate and measure the value creation potential of mobile business applications, and (2) the lack of potentially best practices and use cases in different domains.

If we go back to methodologies for measuring the value of mobile-supported business processes, we might be able to demonstrate that it is possible to show the benefits by business process modeling, design, and simulation methods, followed by an activity-based costing and traditional investment appraisal.

Figure 17. Processes of the meal supply with the mobile system (future situation)

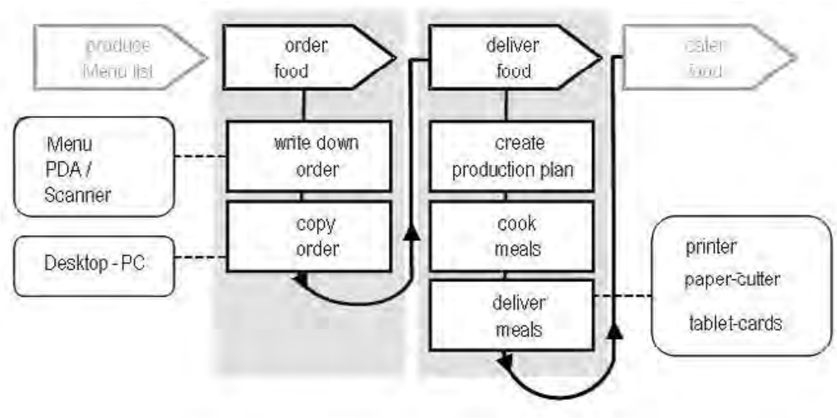


Figure 18. Material costs for meal supply (traditional scenario)

Invest (indispensable)	amount	unit price	Total costs
Software update (license paper based system)	1	3.100 €	3.100 €
Software update (installation)	1	5.200 €	5.200 €
Software update (training costs)	1	1700 €	1700 €
Additional operation costs (per year)	amount	unit price	Total costs
Software material costs	1.336.746,00	0,018 €	23.968 €
Software maintenance contract costs	1	911 €	911 €
Waste of meals	29.975	4 €	120.000 €

However, to do this in a proper way, there was a lot of effort necessary, for example, ethnographic studies and field work for the time of about 2 weeks, participating in the drug and meal supply processes, and recording and analyzing data.

On the other hand the quality of the resulting process model with time and cost data is strongly dependent on the quality of input. But if we look at the investment done in total and the potential savings, the time for research was worth its money.

The process-oriented research approach derived from the MPL method was able to deliver useful results. It was a further enrichment in comparison to research methods that focus on mobile system design (Wang et al., 2005) only, or on abstract means-ends objective networks that describe general value dimensions of mobile applications (Nah et al., 2004).

The business processes that had been chosen from the process map of the hospital turned out to be very fertile for optimization with mobile applications. An ROI of an investment of 200,000 € in mobile systems could be paid back in a time frame of 1 (progressive calculation) or maximum 3 years (conservative calculation).

Hospitals stand to benefit in different ways from implementing a mobile system: There is a high ROI and efficiency gain caused by the use of

Figure 19. Cost of personnel for meal supply (traditional supply system)

Meal supply process (per station)	duration [min.]	factor	costs [per day]
Produce collection of different menu	10,00	1/7	0,69 €
Discuss menu with involved employee	5,00	1/7	0,35 €
Prepare meal voucher	10	3	14,52 €
Sort meal voucher	1	3	1,45 €
Inform kitchen	2,14	1,00	1,04 €
Total costs of meal order process (per station)			18,04 €
Meal delivery process (kitchen)	duration [min.]	factor	costs [per day]
Retrieve meal voucher from stock	3	3	3,10 €
Sort meal voucher	120	3	123,86 €
Correct meal voucher	10	3	10,32 €
Prepare inward meal voucher for ambulant patients	5	1	1,72 €
Prepare meal voucher for staff.	5	1	1,72 €
Count token coins of staff	1	1	0,34 €
Operate inward statistics	20	1	6,88 €
Operate discharge statistics	20	1	6,88 €
Operate statistics for additional delivered fruits	2	1	0,69 €
Transport meal voucher to administration office	3	3	3,10 €
Start meal software	1	3	1,03 €
Scan paper-based meal voucher	35	3	36,13 €
Print out production plan	2	3	2,06 €
Save data	7	1	2,41 €
Total costs meal delivery (kitchen)			200,24 €
Meal delivery process (diet kitchen)	duration [min.]	factor	costs [per day]
Check and correct special diet meal voucher	45	1	19,86 €
Calculate portions for diet	30	1	13,24 €
Calculate ingredients	30	1	13,24 €
Label plates for special diets	30	1	13,24 €
Total costs meal delivery (diet kitchen)			59,57 €

mobile terminals. The restructuring of processes can reduce running time and thus the workload of health care and administrative personnel. By minimizing the number of errors that do occur during the recording and editing of orders, expenses for the drug and meal supply processes can be reduced, too. Information processing can be automated by the use of mobile electronic systems, thus the waste of material, labor time, and storage space for files (actually an important expense factor) can be significantly decreased.

Another remarkable improvement caused by the new mobile system is the extraction of very detailed information that can be used by the controlling department for the measurement of process performance, process costs, and occurring errors. Furthermore, by setting data in relation to patients, applications can be developed to

Figure 20. Cost of personnel for meal supply (with the mobile system)

	duration [min.]	factor	costs [per day]
Meal supply process (per station)			
Discuss menu with involved employee	2,50	1/7	0,17 €
Synchronize PDA	1,00	3,00	1,45 €
Confirm meal order	2,00	1,00	0,97 €
Total costs of meal order process (per station)			2,59 €
Meal delivery process (kitchen)			
Count token coins of staff	2,00	1,00	0,69 €
Print out production plan	1,00	3,00	1,03 €
Print out tablet cards	1,00	3,00	1,03 €
Cut tablet cards	10,00	3,00	10,32 €
Total costs meal delivery (kitchen)			13,07 €
Meal delivery process (diet kitchen)			
Calculate portions for diet	5,00	1,00	2,21 €
Calculate ingredients	1,00	1,00	0,44 €
Total costs meal delivery (diet kitchen)			2,65 €

Figure 21. Formula of the net present value according to Grob (1999)

$$NPV = \left(\sum_{t=1}^{n} \frac{\textit{Cash Flow}_t}{(1 + \textit{Discount Rate})^t}\right) - \textit{Initial Investment}$$

share patient-oriented information including high traceability and a high transparency in supply processes. For example in drug supply a mobile system can provide new possibilities for enhancing the medication: The hospital's pharmacy now plans to enhance the existing mobile system to a unit dose system, in which every patient gets their individual medication, beginning at the bed-side prescription by PDAs to the patient-related packaging and ending at the patient-related billing.

Nevertheless, economic potentials can be reduced significantly by the wrong choice of mobile terminals. This fact was revealed by the analysis of the barcode scanner alternative. Though the same processes are supported, the scanner alternative proved to be a less appropriate solution. Although barcode scanner costs half the price of PDAs, they cannot verify the input quality of data, nor display inconsistency checks or warnings. Furthermore, they are not able to alert staff members in case of patients' incompatibilities against drugs or meals. All input can only be checked at the wards' PC. This means, that in case of occurring problems, data has to be recollected again. Comparing the drug supply process scenario operated by the PDA or scanner supported alternative, it becomes clear that barcode scanners lead to additional expenses and to an increasing personnel workload.

In the context of the drug order and supply processes it has not become clear yet if the implementation of intelligent mobile devices can result in measurable advantages in comparison to the implementation of a scanner-supported scenario. This phenomenon is based in the short spatial distance between the ward PCs and the drug cupboards.

But it would make no sense to operate these processes without the PDA system, as it has no measurable negative effects and all different processes can be supported by one system. This

Figure 22. Cost of material for mobile system in drug and meal supply

	amount	unit price	Total costs
Invest			
PDA Symbol SPT1846	5	1.200 €	6.000 €
PDA Symbol SPT1550	181	500 €	91.000 €
Printer (kitchen, tablet cards)	2	5.000 €	10.000 €
Paper cutter (kitchen)	2	5.000 €	10.000 €
Printer (pharmacy)	3	750 €	2250 €
WLAN access points	2	200 €	400 €
Software development (ward)	1	39.000 €	39.000 €
Software development (kitchen)	1	32.000 €	32.000 €
Software development (in general)	1	1.000 €	1.000 €
Software development (sales tax.)	1	16.000 €	16.000 €
Course of training (pharmacy)	1	5.000 €	5.000 €
Course of training (station)	108	310 €	33.000 €
Course of training (kitchen)	1	200 €	200 €
Installation (pharmacy)	5	10 €	50 €
Installation (ward)	108	4 €	400 €
Installation (kitchen)	5	80 €	400 €
Additional running costs (per year)			
Costs of lost/ broken PDA (Symbol SPT1846)	0,5	1.200 €	600 €
Costs of lost/ broken PDA (Symbol SPT1550)	20	500 €	10.000 €
Costs of paper (pharmacy)	14.000	0,004 €	50 €
Costs of paper (kitchen)	400.000	0,004 €	1.600 €
Costs of hiring server (electronic data processing center)	1	10.000 €	10.000 €
Costs of data traffic (electronic data processing center)	108.000 MB	0,03 €	3.240 €

lowers the break in barriers for the staff members, who have to be trained only in one system with only one kind of mobile device.

If we go further in the business process map of the hospital, we surely will find other business processes that could be supported by the available mobile system, like, for example, the coordination process for the hospitals laundry.

Once the system's use is established among staff members, the next step might be to support medical business processes, too, if it ends up in a higher efficiency. But before doing that, business process performance analysis will be essential. Nowadays there already exists mobile applications for medical business processes, like decision support systems, e-learning, and telemedical communication systems. One of the prominent mobile systems in Germany is the Stroke Angel system (Holtmann, Rashid, Weinhardt, Gräfe, & Griewing, 2006) in which paramedics use PDAs in emergency ambulance vehicles in case of a stroke to send information about their patient to the targeted hospital. Medicals in this hospital can begin their analysis and make preliminary preparations.. Furthermore, paramedics fill out checklists by PDAs that enable hospitals to make a

Figure 23. Value calculation with the net present value method

Invest (Total)	274.750 €
PDA Symbol SPT1846	6.000 €
PDA Symbol SPT1550	91.000 €
Printer (kitchen)	10.000 €
Paper cutter	10.000 €
Printer (pharmacy)	2.250 €
WLAN access points	400 €
Software development	116.000 €
Training and installation costs	39.100 €
Single savings	10.000 €
Additional operation costs (per year)	25.250 €
Savings per year (material costs)	
Meal voucher and maintenance	24.900 €
Waste of meals	30.000 €
Drug order forms	900 €
Costs for needed archiving space	1.000 €
Savings per year (personnel costs)	
½ x diet cook	21.000 €
3x cook assistants	99.000 €
Labor time (general)	591.000 €
Total savings per year (material and personel costs only)	176.800 €
Total savings per year (with additional savings of labour time)	767.800 €
Period	3 years
Bank rate	10 %
„net present value" (NPV) (without process time)	108.700 €
„net present value" (NPV) (with process time savings)	1.581.000 €

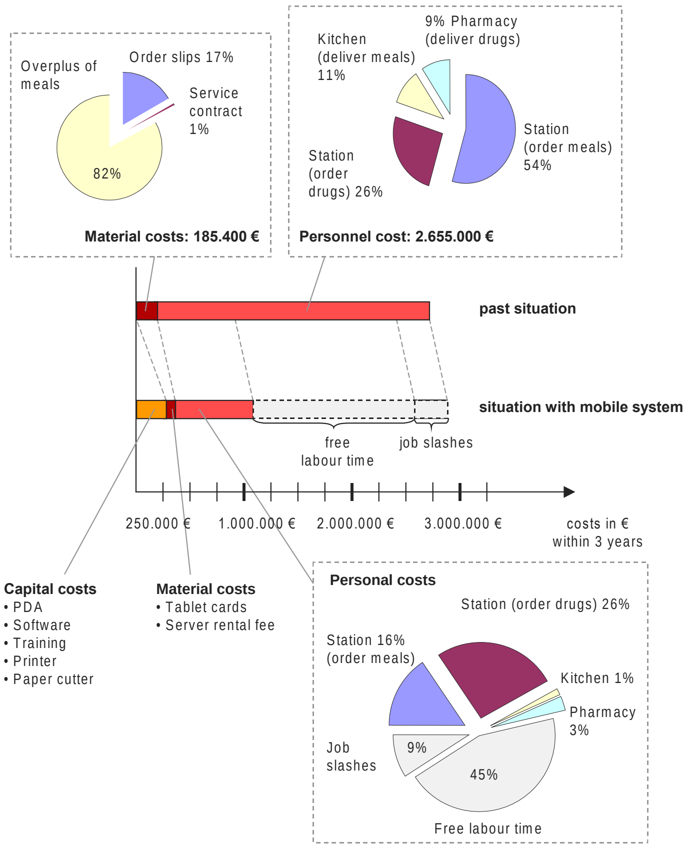
more precise analysis and collect data for improving the emergency management. Up to present there exists no business process performance analysis, neither of the rescue service nor of the emergency units, in hospitals.

A lot of German hospitals are still not able to tell you how much a treatment belonging to a given and predefined DRG costs. Modeling the existent business processes and enriching them with process costs and time data will be a major first step of improvement—the utilization of mobile applications for performance leverage the second.

Otherwise a hospital might go bankrupt by delivering services in a high quality but to very uncompetitive costs. For example, if a hospital is very famous for its kidney surgery and is planning to enhance it without knowing the cost/benefit structure it is possible to cause serious trouble if the profit contribution is negative, as happened to a competitor of the clinic we had research on.

However, to introduce mobile business process modeling, simulation, and redesign together with activity-based costing will be a major contribution of European hospitals to reduce the costs of their health care systems. Supporting or redesigning distributed clinical processes with mobile applications will be substantial items. This should be an inspiring signal for the telecommunication companies in OECD countries to become an informed partner for the health care sector with fitting mobile information applications and consultancy services.

Figure 24. Schematic diagram of the profitability analysis



REFERENCES

GetProcess AG. (2006). *Income process designer*. Retrieved May 25, 2006, from http://www.get-process.de

Greiling, M., & Hofstetter, J. (2002). Patientenbehandlungspfade optimieren—Prozessmanagement im Krankenhaus. Baumann Fachverlag, Kulmbach (Germany).

Grob, L. (1999). *Einführung in die Investitionsrechnung* (3rd ed.). München, Germany: Vahlen.

Hengst, M. den, & De Vreede, G. J. (2004). Collaborative business engineering: A decade of lessons from the field. *Journal of Management Information Systems*, *20*(4), 87-115.

Högler, T. (2006). Framework für eine holistische Wirtschaftlichkeitsanalyse mobiler Systeme. In *Proceedings of the MKWI Multikonferenz Wirtschaftsinformatik 2006*, Universität Passau, Germany.

Holtmann, C., Rashid, A., Weinhardt, C., Gräfe, A., & Griewing, B. (2006). Time is brain—Analyse der Rettungskette im Schlaganfall. In *Proceedings of the 5th Workshop of the GMDS Workgroup Mobiles Computing in der Medizin*, Frankfurt, Germany: Shaker Verlag.

Köhler, A., & Gruhn, V. (2004, February 2-3). Mobile process landscaping am Beispiel von Vertriebsprozessen in der Assekuranz. Mobile economy: Transaktionen, Prozesse, Anwendungen und Dienste. In *Proceedings of the 4th Workshop Mobile Commerce*, Universität Augsburg, Germany.

Knoblauch, H. (2005). Focused ethnography. *Qualitative Social Research, 6*(3). Retrieved May 24, 2006, from http://www.qualitative-research.net/fqs-texte/3-05/05-3-44-e.htm

Lazarsfeld, P., Jahoda, M., & Zeisel, H. (1933). *Die Arbeitslosen von Marienthal. Ein soziographischer Versuch über die Wirkungen langdauernder Arbeitslosigkeit.* Germany: Suhrkamp Leipzig.

Morton, S. & Bukhres, O. (1997). Utilizing mobile computing in the Wishard Memorial Hospital ambulatory service. In B. Bryant, J. Carroll, J. Hightower, and K. M. George, (Eds) *Proceedings of the 1997 ACM Symposium on Applied Computing* (SAC '97) (pp. 287-294). San Jose, California, United States. New York, NY: ACM Press.

Nah, F., Siau, K., & Sheng, S. (2005). The value of mobile applications: A utility company study. *Communications of the ACM, 48*(2), 85-90.

Organisation for Economic Co-operation and Development (OECD). (2006). *OECD fact book: Total and public expenditures in health.* Retrieved May 25, 2006, from http://thesius.sourceoecd.org/vl=5439459/cl=16/nw=1/rpsv/factbook/data/10-01-04-t01.xls

Shiffman, R. N., Karras, B. T., Nath, S., Engles-Horton, L., & Corb, G. J. (1999, August). Pen-based, mobile decision support in healthcare. *SIGBIO Newsl, 19*(2), 5-7.

Statistisches Bundesamt. (2003). *Im Jahr 2050 wird jeder Dritte in Deutschland 60 Jahre oder älter sein.* Retrieved May 25, 2006, from http://www.destatis.de/presse/deutsch/pm2003/p2300022.htm

Wang, Y., Van de Kar, E., & Meijer, G. (2005). Designing mobile solutions for mobile workers: Lessons learned from a case study. In *Proceedings of the 7th international conference on Electronic commerce ICEC'05.*

ENDNOTE

[1] She will publish her PhD thesis including improvements of the MPL method this year. First results had been presented in 2006 at the German Conference of Information Systems (WKWI).

This work was previously published in Global Mobile Commerce: Strategies, Implementation and Case Studies, edited by W. W. Huang, Y. Wang, and J. Day, pp. 174-198, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 6.12
E-Health Dot-Coms' Critical Success Factors

Abrams A. O'Buyonge
Creighton University, USA

Leida Chen
Creighton University, USA

BACKGROUND OF THE BUSINESSES

The increasing use of the Internet by consumers gave rise to an information boom to health-care consumers. Not only could the Internet be used as a communication tool to provide information that would allow patients to make informed decisions, but it could also be used to generate revenue for investors. The dot-com boom of the late 1990s exploited this opportunity, targeting the health-care system, a $1.7 trillion market in the United States alone. Overall, the health-care system is wasteful and costly (Itagaki, Berlin, & Schatz, 2002), and as a result, health-care IT was touted as the magic pill for cutting costs. The Internet boom of the late 1990s saw the emergence of e-health: the delivery of health services and health information through the Internet and Internet-related technologies (Eysenbach, 2001). Leading the many entrepreneurs and venture capitalists who stepped in to seize a piece of the health-care industry cake were WebMD Corp., an online provider of medical information for doctors and consumers in Elmwood Park, New Jersey, and DrKoop.com, an Austin, Texas-born company that later moved to Santa Monica, California, and began doing business as Dr. Koop LifeCare Corp.

Dr. C. Everett Koop, the former U.S. surgeon general, had spent over 6 decades in the medical profession. He envisioned the Internet as an opportunity to change the health-care delivery system in order to empower individuals to take charge of their own health care (Musselwhite, 2002). With this vision and his reputation as an advocate for health-care reform, along with the help of two budding entrepreneurs, Don Hackett and John Zacarro, the trio opened a business-to-consumer Internet portal: DrKoop.com. The portal was designed to provide health information to consumers in areas such as chronic illness, food and nutrition, fitness, and medical breakthroughs. At the beginning, the Web site was an overwhelming success, receiving a million hits per month after 2 years of operation, and about 4 million unique visitors per month at its peak. The portal included a personal medical-records system that facilitated the cross-referencing of medications for interactions, as well as the storage of medical reports that could then be accessed by both patients and physicians.

DrKoop.com's public woes began in February 2000 when its auditor, PricewaterhouseCoopers, issued a "going concern qualification," an ominous warning that highlighted the precarious financial situation the Internet-based health service was in Cleary (2000). By the end of 2000, DrKoop.com was still struggling, and in the first 9 months of 2001 alone, the company's losses were nearly 3 times its revenue. According to the Securities and Exchange Commision (SEC) filings, from January 1999 until the service's liquidation in September 2001, DrKoop.com's losses stood at $193.6 million, dwarfing the $41 million revenue generated during the period. At the site's peak in July 1999, DrKoop.com's stock rose to $45.75 per share on the NASDAQ, but was worth $0.12 at the time of bankruptcy filing. In July 2002, Vitacost.com, a privately held online seller of nutritional supplements, paid a paltry $186,000 in cash for DrKoop.com's assets, which included the brand name, trademarks, domain names, the Web site, and the e-mail addresses of its registered users.

WebMD, originally called Healtheon/WebMD, was founded by Jim Clark, who also founded Silicon Graphics and Netscape. Clark's vision was to connect insurance companies, doctors, and patients over the Internet in order to lower costs and reduce paper trails. Rather than building its own products and services, Healtheon used its highly valued stock to finance acquisitions of leading companies in the industries it targeted. In 1999, it acquired WebMD.com and OnHealth, both leading health portals, giving it access to the consumer health market (Salkever, 2000). Though WebMD lost $6.5 billion on revenue of $530.2 million in the first 9 months of 2001, it still continued to expand long after DrKoop.com had dropped off the radar screen. For the fiscal year ending in December 2003, WebMD reported revenues of $964 million, an increase of 10.6% on the previous year's revenues, which totaled $871.7 million. Of the 11 health-care mergers and acquisition deals in the first 7 months of 2004, valued at $900 million, WebMD was the leading acquirer (Abrams, 2004). Two of WebMD's high-profile acquisitions in 2004 were the $160 million cash purchase of ViPS, a privately held provider in Baltimore, Maryland, of information technology to the government, Blue Cross-Blue Shield, and other health-care insurers; and the $40 million acquisition of Dakota Imaging Inc., a private company in Columbia, Maryland, that offered automated health-care claims processing technology.

As industry leaders, WebMD and DrKoop.com faced competition from both health-care information portals (such as HealthGrades.com, MDConsult, ZoeMed.com) and online pharmacies that provided consumers with one-stop shopping for medications and medical information (Walgreens.com, drugstore.com, Webvan.com). The threat from the health-care information portals, nevertheless, was minimal due to their limited brand recognition and information coverage. In the online pharmacy sector, however, Walgreens.com gained a substantial market share by combining the best of both worlds: complementing its physical stores located throughout the country by offering online customer service, convenience, and real-time access to a health library that provided comprehensive information on prescription drugs, insurance, and health issues.

Table 1. Financial overview of WebMD (Source: Company Reports)

Financial Overview, WebMD 2003 vs. 2002 (in millions of $ dollars)			
	2003	2002 C	hange
Transaction services	$505.7	$466.8	8.3%
Physician services	302.6	275.3	9.9%
Portal services	110.7	84.3	31.3%
Plastic Technologies	71.9	65.8	9.3%
Adjustment -	27.0 -	20.5	31.5%
Revenues	$963.9	$871.7	10.6%
Net Income	$17.0	$49.7 N	A
Source: Company Reports			

Table 2. SWOT analysis of WebMD (Source: WebMD Corporation, 2004)

Strengths	**Weaknesses**
• Large footfall • Growing financial strength • Aggressive acquisition strategy	• Large restructuring charges • Acquisition-driven strategy • Investors unsure of strategy
Opportunities	**Threats**
• Shed unprofitable alliances • Drive core strength and diversify • Excess cash to pursue acquisition • Self-directed health-plan market	• Low demand from practitioners • Increasingly innovative competitors • Power shift in the market

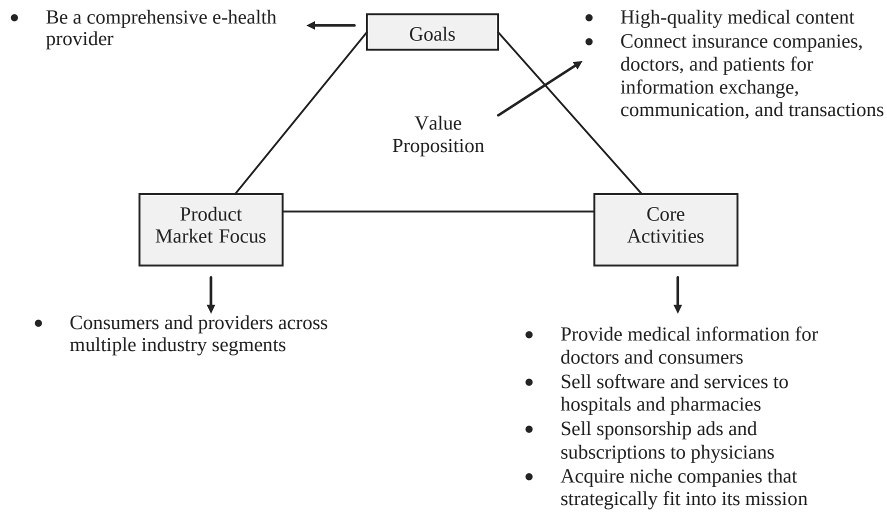
DESCRIPTION OF THE BUSINESSES

Products and Services

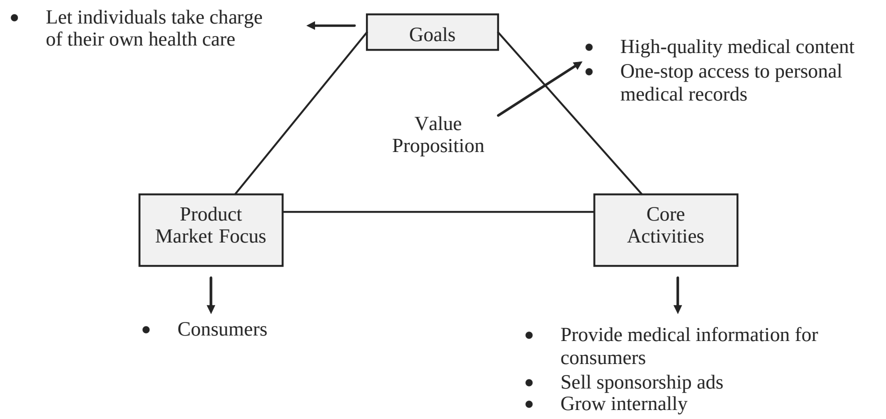
DrKoop.com's vision for its online venture was a Web-based health-care information portal on which individuals and their doctors could have one-stop access to personal medical records, prescription history, medical exam results, and general health-care information (Itagaki et al., 2002; Musselwhite, 2002). According to Musselwhite, the venture's marketing strategy was to differentiate itself as "the premier healthcare information website," and would accomplish this via the four *C*s: content (which would be the driver), community (bringing people together to discuss issues pertaining to health), cool tools (such as Drug Checker), and commerce (enabling people to purchase drugs online from strategic e-commerce partners such as Drug Emporium) (p. 6). In order to generate revenue, DrKoop.com sold advertising rights to health-care-related companies that would pitch sales and promote their products to the portal's users and communities. Unfortunately, DrKoop.com failed to implement checks and controls on advertising, and some of the companies began making promotional claims that were not always substantiated.

WebMD's products and services, on the other hand, were more diversified with the overall objective of facilitating information exchange, communication, and transactions between consumers, physicians, and health-care institutions. Through a series of mergers and acquisitions, WebMD was able to offer a wide variety of health-related

Figure 1. Value proposition frameworks for WebMD and DrKoop.com



WebMD



DrKoop.com

products and services, including health-care information to consumers and physicians, billing and transaction processing for physician practices and hospitals, and handheld services (Itagaki et al., 2002).

Management and Business Models

DrKoop.com was plagued with management problems from the start. None of the founders had business experience, but with less than $1 million of lifetime revenue, DrKoop.com still raised $84.4 million in its initial public offering (IPO). The company swelled from a handful of employees to 200 after the IPO, and an 80,000-square-foot office was acquired in a long-term lease. At a time when competition was fierce, advertising revenue alone could not support operational expenses, but the corporation continued its free-spending practices. Employees' in-house massages cost $9,000 per month, and free, catered Friday lunches cost $15,000 each week, contributing to a monthly burn rate of $7 million (Hawkins, 2001). The company signed a 3-year, $58 million agreement with Disney's Go Network to be the exclusive provider of health-care information on that network (Musselwhite, 2002). This was later followed by a 4-year agreement with AOL for a whopping $89 million to be featured on the AOL Web site, a sum that surpassed the IPO capitalization. Such evidence of capital mismanagement sent the wrong signals to investors, who began to question the viability of DrKoop.com. The business model envisioned by the company—a health-care information portal whose content would attract users, and in which advertisers would be attracted to the users—soon began to crack. By April of 2000, DrKoop.com was trailing its two major rivals in visitors per month and was spiraling into the dot-com abyss. In July, two class-action lawsuits were filed against management alleging the withholding of financial information for the purpose of inflating its stock price. The lawsuit was bolstered by other questionable actions by executives, including the selling of shares before the end of a 6-month holding period, which gave the impression of insider profiteering. Subsequently, two top executives, the chief operations officer and the chief financial officer, were forced to resign. In August, the stock fell below $1 a share and the downward trend never reversed. Public trust and investor goodwill on DrKoop.com were dealt a final blow with more reports of unethical practices. The company had received revenue from a pharmaceutical firm for referring patients for clinical trials, undisclosed commissions were received on health products and services sold through the Web site, and companies had paid money to be listed exclusively in some sections of the Web site, thus misleading customers (Noble, 1999).

Unlike DrKoop.com, WebMD adopted a comprehensive business model, targeting both consumers and providers across multiple industry segments. WebMD began a multibillion-dollar shopping binge from the start, acquiring several firms specializing in medical-related ventures such as electronic medical-claims processing, medical software, medical information, charge capture, physician practice management, and so forth. As the company's financial report shows (see Table 1), revenues from all areas were growing at a healthy pace. WebMD's revenues were expected to top $1 billion in 2004, and after several years of systematic cost cutting, operating earnings were expected to rise to at least $150 million (Tsao, 2004). The strength, weakness, opportunigy, threat (SWOT) analysis from Datamonitor (WebMD Corporation, 2004; see Table 2) indicates a favorable competitive environment for WebMD's continued growth. Using the value proposition framework (Crossan, Fry, & Killing, 2002), the goals, product market focus, core activities, and value propositions of both WebMD and DrKoop.com are compared in Figure 1. Compared to Dr.Koop.com, WebMD has more diversified product market focus and core activities. As a result, WebMD has been able to achieve more enhanced and sustainable value propositions.

LESSONS LEARNED

Apart from the irrational exuberance exhibited by investors, the success or failure of dot-coms can be attributed to other factors. Itagaki et al. (2002) identified four critical factors that predict the success or failure of e-health companies: a compelling value, an unambiguous revenue model, competitive barriers to entry, and the organizational structure for cost control. These factors can be matched to the traits described by Walters (2002): the "lots of money, but no vision" dilemma; the "business-as-a-website" approach; the burn rate; and the speed trap. The following sections discuss how these critical factors played out in the cases of WebMD and DrKoop.com, and the lessons that can be learned.

A Compelling Value or Vision: In the end, both WebMD and DrKoop.com succeeded in providing valuable, easy-to-understand, and useful health-care information to patients and physicians. However, DrKoop.com began to deviate from the original vision and mission, focusing on internal growth and fast-tracking to the IPO without adequate planning for the aftermath, a trap that befell many of the dot-coms that are now virtually extinct. WebMD developed a clear vision (to be a comprehensive e-health provider) and stuck to it, focusing on diversifying assets in order to augment its value. As of July 2004, WebMD Health provided the content for two of DrKoop.com's high-profile customers, the health channels for MSN and America Online, and had teamed up with the federal government to make consumer health information more widely available via the Department of Health and Human Services (HHS) channel on the WebMD Health portal (Henkel, 2004).

The lesson here is that instead of quickly rushing to the IPO, companies should first have a clear vision of where they want to be, a long-term strategy on how to get there, and the realization that agility is the key to survival in the fluid e-commerce marketplace. DrKoop.com lacked a long-term strategy and was forced to take poorly calculated risks. As a result, the stock price plummeted, massive layoffs soon followed, and in a desperate move to save the sinking ship, the company began engaging in questionable practices.

An Unambiguous Revenue Model and Avoiding the Business-as-Web-Site Approach: DrKoop.com built a technology infrastructure that relied heavily on an unproven revenue model and did not have alternative strategies for generating revenue. As a result, expenses soared and, with increased competition, Web-site traffic declined. WebMD, on the other hand, aggressively diversified its revenue sources via strategic mergers and acquisitions of firms in health-related industries. Consequently, WebMD began generating most of its revenue from the sale of software and services to hospitals and pharmacies (Bulkeley, 2003).

The lesson here is that companies should have a viable plan and alternative sources of revenue instead of putting all their eggs in one basket. WebMD's growth from a health-information portal to a comprehensive e-health site enabled the company to effectively deter the threat from substitute products and services. By expanding its offerings to both patients and physicians, and aggressively seeking business opportunities that are in sync with its corporate strategy, WebMD continues to thrive in the volatile e-commerce marketplace.

Competitive Barriers to Entry and Avoiding the Speed Trap: Both WebMD and DrKoop.com were able to create high-quality medical content, utilizing experts to ensure that the information was accurate. DrKoop.com went a step further by branding itself using the identity of a prominent physician. However, by aggressively pursuing a getting-to-market-first approach, DrKoop.com was lean on contingency planning and was thus not able to deal with the shocks that soon followed, such as the droughts of funding and declining customer demand fueled by competition. WebMD, however, was favored by economies of scale: Its large size and comprehensive services made it a dominant force and, in essence, was an entry barrier to competitors. By focusing on expansion, capturing new markets, and buying niche companies that strategically fit into its mission, the company was able to shield itself from competition and market downturns.

The lesson from this is that companies should be alert and ready to adapt to change, or perish. WebMD expanded its base by building relationships with its customers (physicians and general-health-information consumers), industry partners, and the government. Even though fast market penetration has competitive advantages, companies should also exercise patience, ensuring they have a solid and well-tested business model and contingency plan before they set sail into the expanse of the e-commerce market. Competitors who wait and take advantage of the new and improved technologies and business processes often surpass first movers in this dynamic technical and business environment. Marketing gurus emphasize the importance of capitalizing on the first-to-market advantage, but the history of e-commerce has taught us that competitive advantages created by IT were often unsustainable and could be easily duplicated by competitors (Porter, 2001).

Organizational Structure for Cost Control and Controlling the Burn Rate: Cost control was the nemesis of many dot-coms, including WebMD and DrKoop.com. The pay-per-click revenue model was fragile and unpredictable. Companies that did not have alternative revenue sources soon realized that their expenses could not be met. Within the first 6 months following the IPO, DrKoop.com spent $50 million while raising revenues of only $8 million. DrKoop.com's extravagance, with little consideration for earning a profit, further contributed to its demise. WebMD's struggle with cost control was complicated by the arduous task of harmonizing operations among the many business units it owns, but it had accumulated cash reserves that have enabled it to grow.

The lesson here is that management should not abuse its investors' trust by spending irresponsibly. Costs should be controlled and, in essence, aligned to revenues so that expenses do not spiral out of control. The surviving e-businesses spent funds wisely because they recognized that in the e-marketplace, the bargaining power of customers is very strong and demand is ever changing.

The dot-com boom was short lived, and following the bubble burst, many of the dot-coms never recovered. For a time, DrKoop.com seemed to defy the odds, but eventually it succumbed as well. WebMD managed to ride the tide and is now on its way to profitability. In this article, we have explored the factors that were critical to the failure and success of DrKoop.com and WebMD, respectively, and discussed some of the lessons learned. As can be gleaned from the demise of DrKoop.com and the continuing growth of WebMD, success in Internet commerce cannot be realized without a compelling value and vision, a clear-cut and effective revenue-generating strategy, cohesive and effective organization, and a cost-control system. Profit is not the goal but the requisite for any business, thus any dollars spent for creating and keeping customers should be measured and tracked against incremental revenue. Executives must watch the cash flow, income, and loss statements closely, and make changes where necessary to ensure that unnecessary expenses are curbed.

REFERENCES

Abrams, S. L. (2004). Healthcare IT: Hot M&A area. *Investment Dealers' Digest, 70*(30), 12-14.

Bulkeley, W. M. (2003, July 2). Dot-com hope: Akamai, others discover new life. *Wall Street Journal* (Eastern ed.), p. B1.

Cleary, M. (2000). Drkoop.com prospects dim. *Inter@ctive Week, 7*(32), 36.

Crossan, M. M, Fry, J. N., & Killing, J. P. (2002). *Strategic analysis & action* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Elias, P. (2000). DrKoop.com gets a second opinion. *Vault.* Retrieved April 29, 2004, from http://www.vault.com/nr/main_article_detail.jsp?article_id=53176&cat_id=0&ht_type=5_

Eysenbach, G. (2001). What is e-health? *Journal of Medical Internet Research, 3*(2), Article e20.

Retrieved April 29, 2004, from http://www.jmir.org/2001/2/e20/

Hawkins, L. (2001, January 15). Give me your Tired.com, your Poor.com. *BusinessWeek Online*. Retrieved April 29, 2004, from http://www.businessweek.com/2001/01_03/b3715091.htm

Henkel, J. (2004). FDA.gov. *FDA Consumer*, *38*(4), 39.

Itagaki, M. W., Berlin, R. B., & Schatz, B. R. (2002). The rise and fall of e-health: Lessons from the first generation of Internet healthcare. *Medscape General Medicine*, *4*(2), 35-39.

Musselwhite, C. (2002). *Another dot-com disaster*. Retrieved April 29, 2004, from http://www.sph.unc.edu/phli/changeforum/chap9.pdf

Noble, H. (1999, September 4). Hailed as a surgeon general, Koop criticized on Web ethics. *New York Times*. Retrieved May 30, 2004, from http://www.sci.tamucc.edu/~physweb/ethics/Koop.html

Porter, M. E. (1979, March/April). How competitive forces shape strategy. *Harvard Business Review*, *57*(2), 137-145.

Porter, M. E. (2001). Strategy and the Internet. *Harvard Business Review*, *79*(3), 63-78.

Salkever, S. (2000, March 30). Healtheon: Coming down with a case of acute competition? *BusinessWeek Online*. Retrieved April 29, 2004, from http://www.businessweek. com/bwdaily/dnflash/mar2000/nf00330e.htm

Tsao, A. (2004, April 8). WebMD's Achilles' heel. *Business Week Online*. Retrieved April 29, 2004, from http://www.businessweek.com/bwdaily/dnflash/apr2004/nf2004048_2383_db014.htm

Walters, J. S. (2002). Why dotcoms failed (and what you can learn from them). *The CEO Refresher*. Retrieved April 29, 2004, from http://www.refresher.com/!jswdotcom.html

WebMD Corporation. (2004). *Datamonitor*. Retrieved April 29, 2004, from http://www.datamonitor.com

KEY TERMS

Business-to-Consumer (B2C): An e-commerce business model that facilitates transactions between a company and a consumer, as opposed to a transaction between companies (called B2B) or a transaction between consumers (C2C).

E-Business: Any financial or non-financial transaction involving an electronic process using the Internet or Internet technologies e.g. creating a map with directions on Mapquest.com.

E-Commerce (electronic commerce): The buying and selling of products and services by businesses and consumers over the Internet. E-commerce involves a direct financial transaction in the electronic process using Internet technologies. E-commerce encompasses business-to-business or B2B (Cisco), business-to-consumer or B2C (Amazon.com), and consumer-to-consumer or C2C (eBay).

E-Health: The use of emerging technologies, especially the Internet, to improve or enable health and healthcare-related services. The Journal of Medical Internet Research defines e-health as "an emerging field in the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the Internet and related technologies" (Eysenbach, 2001).

Handheld Services: The use of hand-held devices, such as Personal Digital Assistants (PDAs) and Pocket PCs, as an extension of workstation resources through the use of client/server based software for access and synchronization in conjunction with wireless access service provider. Handheld services are used for electronic prescribing, real-time drug references, scheduling, charge capture, etc.

Pay-per-Click: A search engine advertising strategy that allows for companies to bid for a website ranking based on the price they are willing to pay per click-through (when a visitor clicks on a Web ad as a result of keywords used when performing the search, e.g. on Google.com). The client chooses the keywords to appear on his website when a search is performed.

Portal: An electronic environment that provides a secure, single point of interaction with diverse information, business processes, and people, personalized to a user's needs and responsibilities. Examples of web portals include About.com, MSN, and Yahoo.

This work was previously published in Encyclopedia of E-Commerce, E-Government, and Mobile Commerce, edited by M. Khosrow-Pour, pp. 379-384, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 6.13
Healthcare Network Centric Operations:
The Confluence of E-Health and E-Government

Dag von Lubitz
MedSMART, Inc., USA & Central Michigan University, USA

Nilmini Wickramasinghe
Illinois Institute of Technology, USA

ABSTRACT

Healthcare has yet to realize the true potential afforded by e-health. To date, technology-based healthcare operations are conducted chaotically, at a wide variety of nonintegrated fronts, with little or no long-term strategy, and at a tremendous and ever increasing cost. This chapter proposes that in order for healthcare to ever reap the full benefits from e-health, it is imperative for the development of a doctrine of healthcare network centric operations. Otherwise, millions if not billions of dollars will be spent on a futile chase of the definitions of how and when the computer, healthcare provider, and healthcare administrator interact most efficiently and at least expense. The concept of a doctrine, "conceptual platform," that outlines the consequent, goal-oriented way forward, and integrates all constituent elements into a smoothly operating whole, is utilized to great effect in the military. Drawing upon the strategies and techniques employed by the military to develop a network centric doctrine, the chapter outlines the essential components necessary for the establishment of the doctrine for healthcare network centric operations (HNCO), and in so doing not only highlights the integral role played by information computer and communication technologies (IC^2T) but also the pivotal role of policy makers and governments. In fact, HNCO underscores the important yet rarely acknowledged confluence of e-health and e-government.

INTRODUCTION

The rules of competition are changing as a result of the growth of global markets, the increased speed of business transactions, the technological revolution, and continued change in customer expectations. The growth, integration, and sophistication of information computer and communication technologies (IC²T) are changing our society and economy. Consumers and businesses have been particularly quick to recognize the potential and realize the benefits of the Internet and Internet-facilitated computer networks. The resultant "e-revolution" changed many aspects of the traditional "way of doing business," facilitated substantial changes in internal and external management styles, enabled increased efficacy of virtually all production stages and operations, and helped to extend the customer reach. As e-commerce or the application of IC²T to business matures, more attention is being placed on maximising its potential benefits to all areas of society. Two areas where much focus is now being placed regarding the use of IC²T to improve access to information and provide better access to services include public institutions and governments and healthcare.

E-government is defined as the use of IC²T to provide citizens and organizations with more convenient access to government services and embrace interactions within governments (government-to-government), between governments and citizens (government-to-citizen), and between governments and businesses (government-to-business) (Turban, King, Lee, & Viehland, 2004). An analogous definition holds for e-health, which involves the use of IC²T to provide all participants within the healthcare domain with better access to information and services (Wickramasinghe, Geisler, & Schaffer, 2005).

IC²T are without doubt the source and the platform of one of the greatest transformations of society since the invention of print and permit-free flow, access, and exchange of information, and the development of universal means of contact among humans. In practical terms, these technologies offer the possibilities for vast improvement of efficiency and cost reduction in business, provide a platform for dissemination of high quality education, facilitate healthcare delivery, and limit the potential for conflict. Already today, the impact of the increasingly more intensive IC²T use can be measured in the way local, national, and global political, economical, or social transactions are conducted. Yet, with the growing employment of IC²T in daily operations, it is also apparent that neither the optimal pattern of use has been developed (we need look no further, for example, than the productivity paradox (Haag, Cummings, & McCubbrey, 2004; Jessup & Valacich, 2005; Laudon & Laudon, 2004; O'Brien, 2005), nor a philosophy guiding such use has been contemplated despite the emerging chaos that threatens the future growth of the field. The impeding state of chaos is exemplified by the collapse of e-businesses in the 1990s (Affuah & Tucci, 2001; Kalakota & Robinson, 1999; Stiglitz, 2003; Wickramasinghe & Sharma, 2004). The need for optimization and doctrine development is most evident in the arena of healthcare that, at the moment, is the most costly budgetary item in the world's economy (McGown, Overhag, Barnes, & McDonald, 2004; Reinhardt, Hussey, & Anderson, 2002; Stats & Facts, 2002; von Lubitz & Wickramasinghe, 2005a; Wickramasinghe & Ginzberg, 2001; Wickramasinghe & Mills, 2002).

Irrespective of the particular healthcare system adopted by a country, governments will always be one of the major influencing forces and key actors of the healthcare stage. It should therefore be a logical extrapolation that e-government and e-health should share a significant overlap. One might even go to the extreme and suggest that e-health, especially for countries which only have a public healthcare system, is in fact a subset of e-government. And yet, the two are rarely if ever discussed together. This confluence between e-health and e-government is best highlighted in the doctrine of healthcare network centric operations (HNCO). We believe that by acknowledging this confluence, more effective policies and protocols can be developed which will facilitate the adoption and diffusion of HNCO and in turn will lead to more effective and efficient healthcare delivery.

In healthcare, despite their relatively late arrival, computer-based technologies will play the major role in management of services at all levels. Yet, while the concept of "e-health" brings the promise of improved economy, increased efficiency, greater equality, and wider range and availability of services, the increasingly diverging goals and philosophies, disparity in the emerging approaches, and frequent incompatibility or limited nature of the available or developed electronic platforms mitigate against such optimistic future. Moreover, the tendency to convert precise terms describing individual components of "e-operations" (operations or activities where IC^2T have been applied to make them electronic) into often meaningless "household" terminology adds superficial linguistic grandeur to quite non-glamorous and very common-sense notions, and transforms the latter into profoundly sounding expressions of a trivialized credo. The misconceptions serve to produce chaos that, in turn, diffuses the true value of the "electronic" in the context of healthcare, reduces its impact, and, consequently, delays its most significant contribution toward elimination of spatial and temporal barriers, and energized operational facilitation.

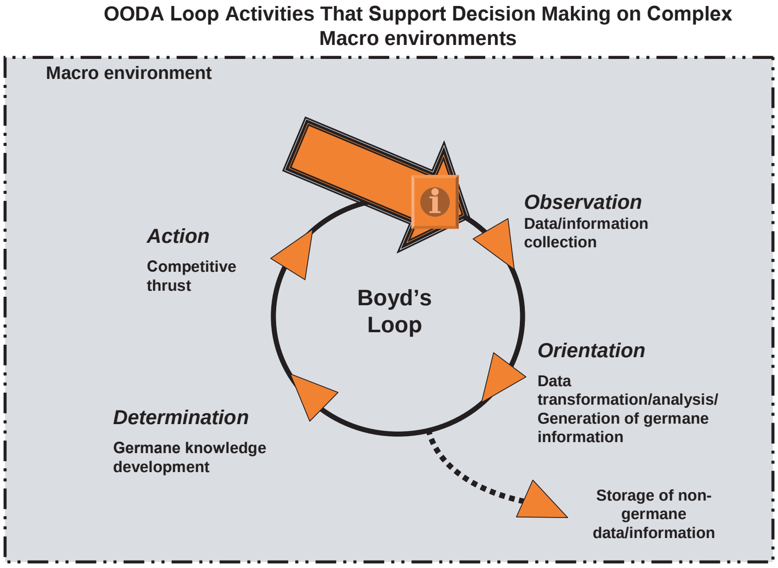
Presently, existing networks and e-health initiatives in healthcare represent distinctive and disconnected entities whose operation is, essentially, platform-centric, that is, concentrates on the operations of a single system (platform) without any regard for the operational interaction among different systems (a "system of systems" or systems network, see von Lubitz & Wickramasinghe, 2005a, b, c, 2006). The consequent fragmentation and broad incompatibility of individual efforts has, in turn, a major influence on the access range to and sharing of high quality information existing among (or even within) the existing individual systems (ibid). For these reasons, despite providing very significant advantages to the local users (Wickramasinghe & Ginzberg, 2001; Wickramasinghe & Mills; 2002), the overall impact of the electronic information systems on either national or international healthcare operations continues to be relatively limited (von Lubitz & Wickramasinghe, 2005a) and nor does it provide the promised superior quality of care (Institute of Medicine, 2001).

Healthcare operations are conducted in a complex operational space. The limitations of a platform-centric approach in such environments are compounded by the fact that the limited (or inadequate) information extraction, gathering, and manipulation capabilities of individual systems are amplified by their frequent lack of interoperability (von Lubitz & Wickramasinghe, 2005a, b, c, 2006). Such limitations severely degrade the possibility of integrating inputs of complex, highly relevant data both from a wide variety of relevant (clinical data banks, for example) as well as seemingly nonessential sources (e.g., political, economical, geographical, ethnographic data pools) and has been the justification for taking a systems approach (Checkland & Scholes, 1990; Churchman, 1968). Environmental complexity of healthcare operations is often magnified by the presence of multiple actors (agencies, governmental bodies, global organizations, etc.) performing within the same space but using a wide variety of independent and non-intercommunicating platform-centric tools. As a consequence of the resulting chaos, the attainment (mission) of healthcare goals (objectives) is uncertainty, rather than information-driven (von Lubitz & Wickramasinghe, 2005c).

HEALTHCARE NETWORK CENTRIC OPERATIONS (HNCO)

The *network-centric approach* stands in direct apposition to platform centricity (Alberts, Garstka, & Stein, 2000; Cebrowski & Garstka, 1998; von Lubitz & Wickramasinghe, 2005a, b, c, 2006). Its underlying conceptual framework is drawn

Figure 1. Boyd's (OODA) Loop depicting how the processes for making critical decisions can be framed in terms of the four key stages of orientation, observation, determination, and action.



Note: The loop revolves both in time and space, and that termination of each revolution (Action stage) modifies the environment with which the Observer interacts, and imposes a new set of "unfolding circumstances" that generate "outside information" and shape the subsequent "Orient" and "Decide" stages. (Adapted from von Lubitz & Wickramasinghe, 2005a, 2005b, 2005c)

from the pioneering work of Boyd (1987), who synthesized decision making, interaction with and control of a fast paced and unpredictably changing environment into a loop-like process (OODA Loop) (Figure 1) (Boyd, 1987; von Lubitz & Wickramasinghe, 2005b, c, d, 2006). Decision making in complex macroenvironments, particularly those characterized by a very rapid revolution cycle and a vast array of multispectral information inputs, is the primary beneficiary of the practical applications of Boyd's Loop. Application of Loop methodology allows for the interpretation of the environment of multidimensional action space (in this case, healthcare) as a set of simultaneous and intertwined events that characterize the space, and are both influencing and are influenced by the actor (e.g., a healthcare organization, clinician, or disease). Boyd's Loop-based thinking forces the actor into continuous extraction and synthesis of pertinent information and germane knowledge. Without such extraction / synthesis, information chaos and overload will rapidly ensue and force the operator into subjective selection of *seemingly* relevant inputs and disregard of other inputs *that appear, at the time of the analysis, to be inconsequential* (von Lubitz & Wickramasinghe, 2005c, 2006; von Lubitz et al., 2004). The ensuing responses to the objective pressures will thus derive from increasingly subjective data interpretation, leading to suboptimal decision making and the ultimate danger of catastrophic errors.

Application of Boyd's Loop and the associated continuous extraction and analysis of high quality information from the environment of the "operational space" provides yet another major advantage: the development of "information superiority" and the reduction of "information asymmetry" relative to the environment.

Each environment contains a complete set of data (information) that describe it. By extracting this information, the actor shifts the knowledge of the given environment from fully unknown (i.e., the environment containing all information that is hidden from the observer) to fully known (the observer uncovers the entire information content of the environment). In the process of uncovering the informational content of the environment, the actor shifts information asymmetry (from environment contained to actor extracted) in his favour and, at the same time, attains the state of "information superiority" (the actor knows progressively more about the faced environment, while the density of the unknown information contained within the environment diminishes during the extraction process). Hence, the highest possible rate of the shift of information asymmetry in favour of the actor and the maximum reduction of time needed for the attainment of valid information superiority are the principal goals of the actor acting in a complex, dynamically evolving environment. Both actor-biased information asymmetry and information superiority are also the principal countermeasure to the "information chaos" or information overload (Lehto, 1991; Prietula, Feltovich, & Marchak, 2000; Tole, Stephens, Harris, & Eprath, 1982) — elements that also characterize many aspects of today's healthcare (Arellano & Weber, 1998; d'Alessandro & Kreiter, 1999; Geiger, Merriles, Walo, Gordon, & Kunov, 1995).

In summary, information superiority provides asymmetric operational advantage not only assures complete control of the direction and tempo of all activities in a collaborative yet highly coordinated manner, but also facilitates attainment of the objective in the most effective and economical way possible.

The state of information superiority can be attained only through the effective use of IC^2T — a critical architectural element of doctrine of "network-centric warfare" created and currently implemented by the U.S. Department of Defense (Stein, 1998). The approach calls for the development of interconnected information grids creating a multilayered, robust network that facilitates information sharing among all participants within the operational space (Arellano & Weber, 1998; d'Alessandro et al., 1999). Drawing upon these ideas and the underlying logic of Boyd's OODA, a doctrine for healthcare network-centric operations has been proposed recently (von Lubitz & Wickramasinghe, 2005a), of which the principal element is the creation and employment of an effective integrated worldwide collection/dissemination/exchange IC^2T grid that will allow free flow of standardized/structured data/information among worldwide actors within the healthcare space necessary to allow the development of global information superiority state.

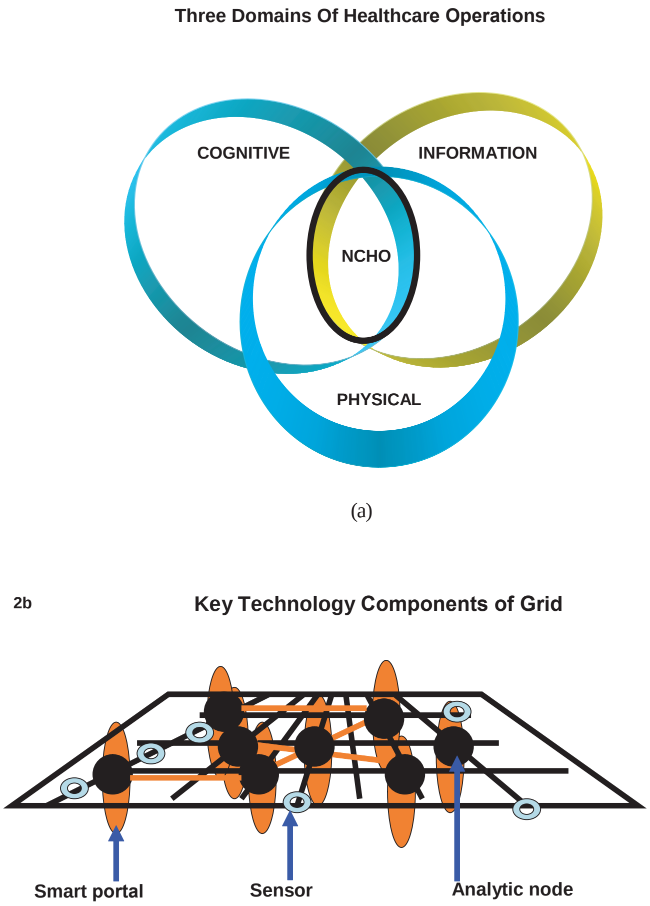
The doctrine of HNCO is thus defined as:

unhindered networking operations within and among all three domains that govern all activities conducted in healthcare space and are based on free, multidirectional flow and exchange of information without regard to the involved platforms or platform-systems and utilizing all available means of ICTs to facilitate such operations. (von Lubitz & Wickramasinghe, 2005a)

The three distinct domains of networkcentric operations are very closely interrelated, and the space within which the domains overlap provides physical and conceptual infrastructure within which all activities are conducted (Figure 2a):

- **Physical domain:** the target that the healthcare intends to influence directly or indirectly. In addition to the most obvious goal of eliminating disease, other aspects of healthcare, for example, research and its policies, fiscal issues, political environment of healthcare, provider personnel, and patient education, belong to this domain. The data/information/knowledge within physical domain relate to its *present rather than future state* and are the easiest to collect. Cumulatively, the physical domain

Figure 2. (a) The relationship among the operational domains in health care. Network-centric health care operations (NHCO) exist only in the territory where all three domains overlap; (b) The critical components of WHIG, including the smart portal, analytic nodes, and sensors (Adapted from von Lubitz & Wickramasinghe, 2005a, 2005b, 2005c, 2005d)



(b)

represents *the current state of healthcare reality.*

- **Information domain:** consists of all elements needed to generate, store, manipulate, disseminate/share/transform information, and disseminate/share the product of such transformation as knowledge in all its forms. All sensory inputs gathered are gathered within the information domain with the sensors representing a wide variety of entities (people and devices.) The output of the sensors may have a wide variety of forms/sensitivity (e.g., basic research data, diagnostic data from individual/multiple patient encounters, electronic health records, etc.) The information existing in this domain may not fully represent the current state of reality. However, all knowledge about that state emerges from and through the interaction with the information domain, and all communications about that occur through interactions within this domain. The information domain is particularly sensitive to incursions that may affect the quality of information contained within the domain. Hence, it must have appropriate security measures that prevent hostile intrusions in any form.

- **Cognitive domain:** constitutes all human factors that affect healthcare operations, such as education, training, experience, personal engagement (motivation), "open-mindedness," or even intuition of individuals involved in relevant activities. All these represent highly intangible quantities that are difficult to measure but form the basis for the selection of right timing, right place, and right effect of all actions undertaken within the healthcare. Even in the presence of well developed decision support systems provided by the information domain, deep situational awareness is created and decisions are made within the cognitive domain.

The technological backbone of network-centric healthcare is the World Healthcare Information Grid (WHIG) (von Lubitz & Wickramasinghe, 2005a, b, c, 2006) which utilizes Web and related Internet technologies, for example, HTML, TCP/IP, Web, JAVA, XTML, and so forth (Eysenbach, 2001; Eysenbach & Diepgen, 1998; Glaser, 2002; McGown et al., 2004; PricewaterhouseCoopers, 2003; Sharma & Wickramasinghe, 2004a; Sieving, 1999; Umhoff & Winn, 1999; Wickramasinghe & Ginzberg, 2001; Wickramasinghe & Mills, 2002). Figure 2a depicts the WHIG with its three distinct yet interconnected domains, each made up of interconnecting grids while Figure 2b identifies the key elements of the grid with its smart access portals, analytic nodes, and intelligent sensors (von Lubitz & Wickramasinghe, 2005b). Implementation of principles based on ASP (application software provider) philosophy and fusing them together with the concept of smart portals allows direct worldwide access to the WHIG that is essentially independent of the technology existing at the user site and that opens network-centric healthcare operations to practically all involved entities — whether individual healthcare providers or national/international bodies providing healthcare governance, monitoring, and policy development.

FUTURE TRENDS: IMPLICATIONS FOR POLICY MAKERS AND GOVERNMENTS

WHIG provides the technology backbone of HNCO; however, for WHIG to function as intended, providing pertinent information and germane knowledge in a seamless, effective, and efficient fashion to a decision maker anywhere anytime, various protocols and procedures must be developed at a global level. Without such standardization even the simplest of functions, such as the exchange of documents and other procurement information, connectivity, and e-commerce enabled benefits, become problematic while the critical goal of decreasing information asymmetry becomes unattainable. Unfortunately, standardization is woefully lacking in too many areas of healthcare let alone e-health. Given the global nature of WHIG, it is here that global bodies, such as the World Healthcare Organization (WHO) in conjunction with governments and policy makers, must develop policies, procedures, and standards that will enable the seamless functioning of WHIG. We identify four key areas that must be addressed:

1. **Information computer communication technology (IC^2T) architecture/infrastructure:** The generic architecture for most e-health initiatives is similar to that required by WHIG. However this infrastructure that consists of phone lines, fiber trunks and

submarine cables, T1, T3 and OC-xx, ISDN, DSL and other high-speed services used by businesses as well as satellites, earth stations and teleports must be available globally.

A sound technical infrastructure is an essential ingredient to the undertaking of e-health initiatives by any nation. Such infrastructures should also include telecommunications, electricity, access to computers, number of Internet hosts, number of ISPs (Internet service providers) and available bandwidth and broadband access. To offer a good multimedia content and thus provide a rich e-health experience, one would require a high bandwidth. ICT considerations are undoubtedly one of the most fundamental infrastructure requirements.

Networks are now a critical component of the business strategies for organizations to compete globally. Having a fast microprocessor-based computer at home has no meaning unless you have high bandwidth based communication infrastructure available to connect computers with the ISP. With the explosion of the Internet and the advent of e-commerce, global networks need to be accessible, reliable, and fast to participate effectively in the global business environment. Telecommunications is a vital infrastructure for Internet access, hence also for e-commerce. One of the pioneering countries in establishing a complete and robust e-health infrastructure is Singapore which is in the process of wiring every home, office, and factory up to a broadband cable network which will cover 98% of Singaporean homes and offices (APEC, 2001; A report prepared for Asia Pacific Foundation, 2002a).

It is estimated that over 70% of world telephone lines are in countries with 15% of the world's population (Parker, 1998). The average number of telephone lines in industrialized countries is one for every two persons as compared to 13 lines per thousand people in emerging economies such as India and China (ibid). A sound technical infrastructure is a key ingredient to the future economic health of any given nation and should be a priority for all governments.

For example, in the Asia-Pacific region (APEC, 2001; Beal, 2000; Oxley & Yeung, 2001; A report prepared for Asia Pacific Foundation, 2002a; A study report on Thailand, 2001; Turpin, 2000), there are a number of "digital divides," not only between the richer and poorer countries, but divides between urban and rural populations and between more and less educated or affluent groups. Although the number of female Internet users in a number of countries in the region is catching up quickly with those of male users, women are less present when it comes to the actual use of the new technologies. Furthermore, in many of these countries, women make up the majority of the rural population, which is often marginalized in areas of telecommunication infrastructure, education, and training. Therefore, governments should formulate national IC^2T and e-commerce strategies that help to ensure universal access for all socioeconomic groups. A number of prerequisites for access to IC^2T include education and training, local content, sociocultural awareness, and a stable social, economic, and political environment. Appropriate technologies need to be developed to address the needs of disadvantaged communities.

2. **Building human capital/resources:** The disparity between educational standards of developed and developing nations in the world is significant (Roquilly, 2002; Sharma & Wickramasinghe, 2004a, b; UNCTAD, 2002). In order for healthcare network-centric operations to fully maximize the potential afforded by WHIG significant enhancement of human capital through education and

training is required. Many of the developing nations through investment and foreign aid need to concentrate on developing human capacity, and increasing basic access to IC^2T, as well as rapidly increasing the presence and access to the Internet. Moreover, the governments of disadvantaged countries must also exercise significant international pressure to lower costs of hardware and software while, at the same time, contemplating to form larger multinational user blocks capable of effective implementation of ASP principles. E-commerce adoption in general requires improved e-commerce knowledge and skills, and improved language proficiency (especially in English as it is as the recognized principal technology / commerce language). For example, governments in many countries have started to introduce basic education in digital literacy in primary and secondary schools. Such training courses should be constantly updated as new innovations and practices emerge very rapidly in this field. Institutions responsible for human resource development of e-commerce personnel should provide appropriate incentives for e-commerce courses to be kept up-to-date. Increasing the number of programs or activities for human resource development for IC^2T and especially Internet use will only be effective if the education and training matches the changing needs of the industries concerned. It is important to note that only medical specialists require medical training but all people globally who need to interact with WHIG must have basic familiarity, training, and proficiency with use of IC^2T. Once again much of the onus falls on governments to set the agenda and priority for such an emphasis.

3. **Fostering consumer trust:** Consumers are also concerned about a number of dimensions of trust; trust in the security of value passed during electronic transactions with organizations that are "virtual" in a disconcertingly ineffable way and trust in the privacy of personal data arising from electronic transactions (Fjetland, 2002; Ghosh & Swaminath, 2001; Panagariya, 2000; Roquilly, 2002). Other than in North America, Japan, and integrated Europe, the infrastructure for e-commerce is not in place for effective e-commerce transactions. The key reason for slow penetration has been the scale of investment in infrastructure, and the small volume of transactions over which to amortize (Dutta, 1999; A report prepared for the Asia Pacific Foundation, 2002b).
The security issue in particular is perceived as very important across the Asia-Pacific region, and the majority of SMEs, for instance, have a fear of electronics. This is primarily due to the low level of technology diffusion and awareness making it a psychological barrier for SMEs as confirmed in various reports (APEC, 2001; Beal, 2000; A report prepared for the Asia Pacific Foundation, 2002b). Many of these SMEs do not have technical expertise, and are not convinced that the technology standards, such as encryption, will protect them. Thus, SMEs are not willing to use electronic payment systems.
Given such a general lack of trust in developing nations with using e-commerce, it is likely that the trust issue will be a significant factor in e-health initiatives given the general sensitivity of healthcare information. Once again government and national bodies must make every attempt to foster trust. This will not be an easy task since in many developing countries people have a low level of trust for governments, corruption tends to be significant, and especially in rural areas, people tend to be very superstitious and leery of "modern methods." However, in many instances WHIG and healthcare network

centric operations will offer these groups of people the best healthcare alternatives and thus significant effort must be made to build and foster trust.

Ethical concerns are inevitably related to healthcare operations based on network centricity. The notion of the governmental bodies having ready access to healthcare information of the citizens is among the major concerns in the U.S., and similar reservations are also voiced in Europe. The possibility of security breaches, similar to those that recently affected millions of credit card customers in the United States, demands a very stringent layer of protective layers that will assure prevention of commercial misuse of healthcare data. The seriousness of such intrusions is emphasized by the dynamic development of molecular biology and genetics that allow early determination of the likelihood to develop long-term debilitating diseases at the later stages of life. Misappropriation of such information may then affect the subsequent chances of employment, nature and availability of health insurance, and even acquisition of credit. While laws preventing such discrimination already exist in several Western countries, their enforcement is not simple and the fear of clandestine stratification into "health risk" classes whose quality and freedom of life is directly related to the long-term risk is a threat that needs to be addressed with utmost vigor now. At a less threatening, albeit far more practical level, the problems of the responsibility for the delivery of healthcare also need clarification.

Among the major obstacles to the development of an efficient national telemedicine network in the U.S. is the difficulty of assigning responsibility for treatment conducted across state borders. The dilemma becomes even more pertinent when healthcare expertise is projected across national borders. While there is no doubt that the rule of the "best medical practice" ought to be the prevailing one, the best practice has different definitions in different countries and what is considered the standard of care in one country may not be so in another. There is thus a clear need to develop generally accepted standards of delivery and the current acceptance of the concept of "evidence-based medicine" may offer the most useful foundation for further deliberations of such standards. These are but two of several problems that already either influence or are about to influence e-health in all its forms. The magnitude of network centricity and its global nature will add even greater urgency to the need for their effective operational rather than merely political or legislative solution. At a purely technical level, while network centricity strengthens the efficacy and economy of global healthcare, it may also have an impact on access to it. Ideally, network-centric operations will support a continuously broadening access to high quality healthcare across the entire globe. In practice, however, there is a chance it may support improvements there where the level of the preexisting technology and sophistication in its use are already high while leaving those who lag behind, yet needing rapid improvement of even basic healthcare in the same position they are now — desperately struggling to satisfy the most elementary healthcare needs of the local population. It is thus evident that introduction of network centricity and the creation of WHIG require a fundamentally different approach to the one currently practiced: they can be made functional only when a state of intense, multilateral, and determinedly synergistic collaboration exists between corporate and governmental entities not only at the national but, even more importantly, at the international level. Ultimately, transparency, international synergism, and adherence to the highest ethical standards of operations will be among the most instrumental factors to convince the users (i.e., both deliverers and recipients of healthcare) that network centricity in healthcare offers tangible benefits that have been unattainable prior to the implementation of the concept.

4. **Creating synergies between national and regional economic blocks:** For WHIG and HNCO to be fully functional, it is vital that synergies are created between national and economic blocks. For example, in similarity to the EU, all countries in the Asia-Pacific region should improve access to IC^2Ts, design the necessary legal and institutional frameworks to educate the governments, businesses, and civil society to more efficiently use IC^2Ts in their daily practices, and create an environment for further development of IC^2T use in general and in particular with a focus on healthcare. There is a tremendous need to support the integration and interoperability of regional e-commerce initiatives through various international and regional economic blocks such as the Asia-Pacific Economic Cooperation (APEC), North American Free Trade Agreement (NAFTA), and the European Union.

   Another critical aspect that requires significant consideration to enable and facilitate the required synergies is that pertaining to legal and ethical issues. In general the Internet community has developed its own distinctive culture and a great part many of these behavioural norms function outside the realm of International Law given the current maturity of International Internet Law (Commission of the European Communities, 2003; De Ly, 1992, 2000; Nielson, 2000; Polanski, 2005; Polanski & Johnston, 2002). Given the sensitive nature of much of the information contained with healthcare interactions, this poses a significant problem and major stumbling block for the adoption of HNCO. This is clearly an area that requires significant discussion and will form the center of future research; we mention it here only in passing to note that it remains as yet a key area that needs to be addressed

CONCLUSION

It is beyond the scope of this chapter to discuss the details of the technology configuration of WHIG or provide guidelines for practioners. Interested readers can refer to other research (von Lubitz & Wickramasinghe, 2005a, 2005b, 2005c, 2005d; von Lubitz, Wickramasinghe, & Yanovsky, 2005) for details. The primary focus here is to highlight the key role for governments if we are to ever realize the full benefits of e-health.

The idea of WHIG and HNCO is to span many parties and geographic dimensions. To enable such a far reaching coverage, significant amounts of document exchange and information flows must be accommodated. Standardization is the key for this. Once a country decides to undertake e-health initiatives and become a "WHIG member," standardization policies, protocols, and procedures must be developed at the outset to ensure the full realization of the goals of e-health. The transformation to e-health by any country cannot be successfully attained without the deliberate establishment of standardization policies, protocols, and procedures which play a significant role in the adoption of e-health and the reduction of many structural impediments (Panagariya, 2000; Samiee, 1998). Fortunately, the main infrastructure of WHIG is the Internet which imposes the most widely and universally accepted standard protocols such as TCP/IP and http. It is the existence of these standard protocols that has led to the widespread adoption of the Internet for e-commerce applications.

Access to the technologies of e-commerce is defined by the WTO (World Trade Organization) as consisting of two critical components: (1) access to Internet services and (2) access to e-services (Health Insurance Portability and Accountability Act, 2001); the former deals with the user infrastructure, while the latter pertains to specific commitments to electronically accessible services. The user infrastructure includes number of Internet hosts and number of Web sites, Web users as a percent of the population as well as ISP availability and costs for consumers, PC penetration level, and so forth. Integral to user infrastructure is the diffusion rate of PCs and Internet usage. The United States and the United Kingdom have experienced the greatest penetration of home computers (Samiee, 1998). For developing countries, such as India and China, there is, however, a very low PC penetration and teledensity. In such a setting it is a considerable challenge then to offer e-health, since a large part of the population is not able to afford to join the e-commerce bandwagon. Countries, thus have to balance local call charges, rentals, subscription charges, and so forth, otherwise the majority of citizens will find these costs a disincentive. This is particularly significant for developing and emerging nations where access prices tend to be out of reach for most of the population. Upcoming new technologies hold the promise to increase the connectivity as well as affordability level and developing countries will need to seriously consider these technologies. In addition to access to PCs and the Internet, computer literacy is important and users must be familiar not only with the use of computers and pertinent software products but also the benefits and potential uses of the Internet and World Wide Web (ibid).

The key challenges regarding e-health use include (1) cost effectiveness that is less costly than traditional healthcare delivery; (2) functionality and ease of use: that is, they should enable and facilitate many uses for physicians and other healthcare users by combining various types and forms of data as well as be easy to use; and (3) they must be secure. One of the most significant legislative regulations in the U.S. is the Health Insurance Portability and Accountability Act (HIPAA) (2001).

Given the nature of healthcare and the sensitivity of healthcare data and information, it is incumbent on governments not only to mandate regulations that will facilitate the exchange of healthcare documents between the various healthcare stakeholders but also to provide protection of privacy and the rights of patients (Dyer, 2001). Some countries, such as China and Singapore, even control access to certain sites for moral, social, and political reasons while elsewhere transnational data flows are hindered by a plethora of regulations aimed at protecting domestic technology and related human resource markets (Ghosh & Swaminatha, 2001; Gupta, 1992; Samiee, 1998). Irrespective of the type of healthcare system; that is, whether 100% government driven, 100% private or a combination thereof, it is clear that some governmental role is required to facilitate successful e-health initiatives.

HNCO also serves to underscore the inextricable connection and intertwining of e-health and e-government which to date has rarely been researched let alone acknowledged. Moreover, for HCNO to become adopted successfully, it requires governments to develop policies and protocols which will in turn facilitate its usability. We identify four key areas that will have an important impact on the development of the necessary policies and protocols as IT education, morbidity, cultural/social dimensions, and world economic standing as elaborated upon below. It is interesting to note that these areas also impact the development of numerous e-government initiatives (Turban et al., 2004), once again highlighting the inextricable link between e-health and e-government.

- **IT education:** A sophisticated, well educated population boosts competition and hastens innovation. According to Michael Porter, one of the key factors to a country's strength in an industry is strong customer support (Porter, 1990). Thus, a strong domestic market leads to the growth of competition which leads to innovation and the adoption of technology enabled solutions to provide more effective and efficient services such as e-health and telemedicine. As identified earlier, the health consumer is the key driving force in pushing e-health initiatives. We conjecture that a more IT educated healthcare consumer would then provide stronger impetus for e-health adoption.

- **Morbidity rate:** There is a direct relationship between health education and awareness and the overall health standing of a country. Therefore, a more health conscious society, which tends to coincide with a society that has a lower morbidity rate, is more likely to embrace e-health initiatives. Furthermore, higher morbidity rates tend to indicate the existence of more basic health needs (World Health Organization, 2003) and hence treatment is more urgent than the practice of preventative medicine and thus e-health could be considered an unrealistic luxury and, in some instances, such as when a significant percentage of a population is suffering from malnutrition related diseases, is even likely to be irrelevant, at least in the short term. Thus, we conjecture that the modifying impact of morbidity rate is to prioritize the level of spending on e-health vs. other basic healthcare needs.
- **Cultural/social dimensions:** Healthcare has been shaped by each nation's own set of cultures, traditions, payment mechanisms, and patient expectations. While the adoption of e-health, to a great extent, dilutes this cultural impact, social and cultural dimensions will still be a moderating influence on any countries e-health initiatives. Another aspect of the cultural/social dimension relates to the presentation language of the content of the e-health repositories. The entire world does not speak English, so the e-health solutions have to be offered in many other languages. The e-health supporting content in Web servers/sites must be offered in local languages, supported by pictures and universal icons. This becomes a particularly important consideration when we look at the adoption and diffusion of evidence-based medicine as it will mean that much of the available evidence and case study data will not be easily accessible globally due to language barriers.

Therefore, for successful e-health initiatives it is important to consider cultural dimensions. For instance, an international e-commerce study by International Data Corp. indicates that Web surfing and buying habits differ substantially from country to country (Wilson, 1999) and this would then have a direct impact on their comfort to use e-commerce generally and e-health in particular, especially as e-health addresses a more fundamental need. Hence, the adoption of e-health is directly related to ones comfort with using the technology and this in turn is influenced in a major way by cultural dimensions. Also connected with cultural aspects is the relative entrepreneurial spirit of a country. For example, a study (Hofstede, 1984) indicates that in a cultural context, Indians score high on "uncertainty avoidance" criteria when compared to their Western counterparts. As a result, for example, Indians do not accept change very easily and are hostile towards innovation. This then would potentially pose a challenge to the starting up of e-health initiatives whose success depends on widespread adoption for their technological innovations. Thus, we conjecture that fear of risk and absence of an entrepreneurial mindset as well as other cultural/social dimensions can also impact the success of e-health initiatives in a given country.

- **World economic standing:** Economies of the future will be built around the Internet. All governments are very aware of the importance and critical role that the Internet will play on a country's economy. This makes it critical that appropriate funding levels and budgetary allocations become a key component of governmental fiscal policies so that such initiatives will form the bridge between a traditional healthcare present and a promising e-health future. Thus, the result of which would determine success of effective e-health implementations and consequently have the potential to enhance a country's economy and future growth.

A healthy society is in the interests of all governments. As labor costs increase and productivity is critical to economic development, a healthy population is naturally able to contribute more significantly to increasing GDP. Coupled with the fact that healthcare remains the most expensive item for any government, it behooves governments to make healthcare a top priority on their agenda. The doctrine of HNCO serves to outline a coherent and systematic approach for harnessing the full potential of IC^2T for healthcare and the ability to realize the promises of e-health. However, if this is to occur, e-health must become a top priority on all government agendas. Moreover, governments must go further than just acknowledge the importance of e-health to focusing their energies and efforts to address the key impediments we have identified and that currently preclude the ubiquitous adoption of HNCO.

The World Economic Forum's Global competitiveness ranking measures the relative global competitiveness of a country. This ranking takes into account factors such as physical infrastructure, bureaucracy, and corruption. It is a simple extrapolation of the combination of these factors to postulate that the combination of a weak physical infrastructure with high levels of bureaucracy and corruption will constitute a significant impediment to the establishment of successful e-health initiatives. Surely, network-centric healthcare operations will not change these realities. On the other hand, the possibility for a rural healthcare provider working in an even most oppressive political and economical environment to enter WHIG and benefit from the combined global healthcare expertise will be without doubt a step in the right direction. And many steps in the right direction will ultimately help to redress the level of current inequalities in access to the most fundamental right of all humans across the globe — health.

REFERENCES

Affuah, A., & Tucci, C. (2001). *Internet business models and strategies.* Boston: McGraw-Hill Irwin.

Alberts, D. S., Garstka, J. J., & Stein, F. P. (2000). *Network centric warfare: Developing and leveraging information superiority* (CCRP Publication Series, pp. 1-284). Washington, DC: Department of Defense. Retrieved May 5, 2006, from http://www.dodccrp.org/publications/pdf/Alberts_NCW.pdf

APEC. (2001). *The new economy and APEC.* Report to the Economic Committee of APEC. Retrieved May 5, 2006, from http://www.diw.de/deutsch/produkte/veranstaltungen/docs/apec-report.html

Arellano, M. G., & Weber, G. I. (1998). Issues in identification and linkage of patient records across an integrated delivery system. *Journal of Healthcare Information Management, 12*, 43-52

Beal, T. (2000). *SMEs and the World Wide Web: Opportunities and prospects*. In A. M. Asri (Ed.), *Small and medium enterprises in Asia Pacific* (Vol. III Development Prospects, pp. 102-134). Commack, NY: Nova Science Publishers.

Boyd, J. R. (1987). *COL USAF, in "Patterns of Conflict", unpublished briefing*. Retrieved from "Essence of Winning and Losing", http://www.d-n-i.net

Cebrowski, A. K, & Garstka, J. J. (1998). Network-centric warfare: Its origin and future. *US Nav. Inst. Proc., 1*, 28-35

Checkland, P., & Scholes, J. (1990). *Soft systems methodology in action*. New York: Wiley.

Churchman, C. (1968). *The systems approach*. New York: Del Publishing.

Commission of the European Communities (2003). *First report on the application of Directive 2003/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular, electronic commerce, in the Internal Market (Directive on electronic commerce)* (pp. 1-25). Brussels.

d'Alessandro, D. M., & Kreiter, C. D. (1999). Improving usage of pediatric information on the Internet: The Virtual Children's Hospital. *Pediatrics, 104*, e55.

De Ly, F. (1992). *International business law and Lex Mercatoria*. Amsterdam, London, New York, Tokyo: T. M. C. Asser Instituut — The Hague.

De Ly, F. (2000). Emerging new perspectives regarding Lex Mercatoria in an era of increasing globalization. In *Festschrift für Otto Sandrock* (pp. 179-204). Heidelberg: Verlag Recht und Wirtschaft.

Dutta, A. (1999). The physical infrastructure for electronic commerce in developing nations: Historic trends and the impacts of privatization. *International Journal of Electronic Commerce, 2*(1), 63-82.

Dyer, K. A. (2001). Ethical challenges of medicine and health on the internet: A review. *JMIR, 3*(2). Retrieved from http://www.jmir.org/2001/2/e23

E-Health and the Practicing Physician. (n.d.). Retrieved May 5, 2006, from http://www.medscape.com

E-Health in the Medical Field. (n.d.). Retrieved May 5, 2006, from http://www.americantelemed.org/ehealth/ehealthH.htm

Eysenbach, G. (2001). Journal of Medical Internet Research is now indexed in Medline. *Journal of Medical Internet Research, 3*(3), e25.

Eysenbach, G., & Diepgen, T. L. (1998). Towards quality management of medical information on the Internet: Evaluation, labeling, and filtering of information. *BMJ, 317*, 1496-1500.

Fjetland, M. (2002). Global commerce and the privacy clash. *Information Management Journal, 36*(1), 54-58.

Geiger, G., Merriles, K., Walo, R., Gordon, D., & Kunov, H. (1995). An analysis of the paper-based health record: Information content and its implications for electronic patient records. *Medinfo, 8*(1), 295.

Ghosh, A. K., & Swaminatha, T. M. (2001, February). Software security and privacy risks in mobile e-commerce. *Communications of the ACM, 44*(2), 51-58.

Glaser, J. (2002). *The strategic application of information technology in health care organizations* (2nd ed.). San Francisco: Jossey Bass.

Gupta, U. (1992). Global networks: Promises and challenges. *Information Systems Management, 9*, 28-32.

Haag, S., Cummings, M., & McCubbrey, D. (2004). *Management information systems for* *the Information Age* (4th ed.). Boston: McGraw-Hill Irwin.

Health Insurance Portability and Accountability Act (HIPPA). (2001). *Privacy compliance executive summary.* Washington, DC: Protegrity Inc.

Health Technology Center. (2000). *A survey conducted for the Health Technology Center (HealthTech) by Harris Interactive in cooperation with Pricewaterhouse Coopers and the Institute for the Future (IFTF).* Retrieved May 5, 2006, from http://www.ncddr.org/cgi-bin/good-bye.cgi?url=http://www.healthtechcenter.org

Hofstede, G. (1984). *Culture's consequences, international work related values.* Beverly Hills: Sage.

Institute of Medicine. (2001). *Crossing the quality chasm: A new health system for the 21st Century Committee on Quality of Health Care in America Institute of Medicine.* Washington, DC: National Academy Press.

Jessup, L., & Valacich, J. (2005). *Information systems today* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Kalakota, R., & Robinson, M. (1999). *E-business roadmap for success.* Reading, MA: Addison-Wesley.

Laudon, K., & Laudon, J. (2004). *Management information systems* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Lehto, M. R. (1991). A proposed conceptual model of human behaviour and its implications for design of warning. *Percept. Mot. Skills, 73*, 595-611.

McGowan, J. J., Overhag, J. M., Barnes, M., & McDonald, C. J. (2004). Indianapolis I3: The third generation integrated advanced information management systems. *J Med Libr Assoc, 92*, 179-187.

Nielsen, J. (2000). *Designing Web usability: The practice of simplicity.* Indianapolis, IN: New Riders Publishing.

O'Brien, J. (2005). *Management information system* (6th ed.). Boston: Irwin McGraw-Hill.

Odyssey Research. (n.d.). Retrieved May 5, 2006, from http://www.odysseyresearch.org

Oxley, J. E., & Yeung, B. (2001). E-commerce readiness: Institutional environment and international competitiveness. *Journal of International Business Studies, 32*(4), 705-723.

Panagariya, A. (2000). E-commerce, WTO and developing countries. *The World Economy, 23*(8), 959-978.

Parker, B. (1998). *Globalization and business practices.* Beverley Hills: Sage.

Polanski, P. (2005, June 17-19). Common practices in the electronic commerce and their legal significance. In *Proceedings of the 18th Bled E-Commerce Conference*, Bled, Slovenia (pp. 170-189).

Polanski, P. P., & Johnston, R. B. (2002, January 5-8). International custom as a source of law in global electronic commerce. In *Proceedings of the 35th Hawaii International Conference on System Sciences* (CD-ROM), Big Island, HI.

Porter, M. (1990). *The competitive advantage of nations.* New York: Free Press.

PricewaterhouseCoopers Healthcare Practice. (2003). Retrieved May 5, 2006, from http://www.pwchealth.com

Prietula, M. J., Feltovich, P. J., & Marchak, F. (2000). Factors influencing analysis of complex cognitive tasks: A framework and example from industrial process control. *Human Factors, 42*, 56-74.

Reinhardt, U. E., Hussey, P. S., & Anderson, G. F. (2002). Cross-national comparisons of health systems using OECD data. *Health Affairs, 3*, 169-181.

Report prepared for Asia Pacific Foundation. (2002a). *Asia-Pacific E-Commerce: B2B & B2C.*

Retrieved May 5, 2006, from http://www.gii.co.jp/english/em11033_asia_ec_toc.html

Report prepared for Asia Pacific Foundation. (2002b). *E-commerce and Asia Pacific, Castle Asia Pacific*. Retrieved May 5, 2006, from http://www.gii.co.jp/english/em11033_asia_ec_toc.html

Roquilly, C. (2002). Closed distribution networks and e-commerce: Antitrust issues. *International Review of Law, Computers & Technology, 16*(1), 81-93.

Samiee, S. (1998). The Internet and international marketing: Is there a fit? *Journal of Interactive Marketing, 12*(4), 5-20.

Sharma, S., & Wickramasinghe, N. (2004a). E-health with knowledge management: Areas for tomorrow. In N. Wickramasinghe, J. Gupta, & S. Sharma (Eds.), *Creating knowledge-based healthcare organizations* (pp. 110-124). Hershey, PA: Idea Group Publishers.

Sharma, S., & Wickramasinghe, N. (2004b). Obstacles to SMEs for e-adoption in the Asia Pacific region in ebusiness, egovernment & small and medium enterprises: Opportunities and challenges. In Al-Qirim et al. (Eds.), *E-business, E-government & Small and Medium Enterprises: Opportunities and Challenges* (pp. 112-133). Hershey, PA: Idea Group Publishing.

Sieving, P. C. (1999). Factors driving the increase in medical information on the Web: One American perspective. *Journal of Medical Internet Research*, 1. Retrieved May 5, 2006, from http://www.jmir.org/1999/1/e3

Stats and facts. (2002). Global health cost and care comparisons. *Manage Care Interface, 15*, 66-77.

Stein, P. (1998). *Observations on the emergence of network centric warfare*. Retrieved May 5, 2006, from http://www.dodccrp.org/research/ncw/stein/_observations/steincw.htm

Stiglitz, J. E. (2003). *The roaring nineties*. New York: Norton & Co.

Study report on Thailand. (2001). *SMEs and e-commerce*, Asia Pacific Foundation.

The higher the connection speed, the higher the value: Broadband use as an indication of value amongst e-health users. (2001). *Cyber Dialogue Cybercitizen Health Trend Report 2001*, No. 49.

Tole, J. R., Stephens, A. T., Harris, R. L., Sr., & Eprath, A. R. (1982). Visual scanning behavior and mental worlkload of aircraft pilots. *Aviat. Space Environ. Med., 53*, 54-61.

Turban, E., King, D., Lee, J., & Viehland, D. (2004). *Electronic commerce*. Upper Saddle River, NJ: Prentice Hall.

Turpin, T. (2000). SMEs and technology diffusion in Asia Pacific-Pacific economies: Indonesia a case-study. In *Proceedings of the Annual Conference of the Academy of International Business Southeast Asia Pacific Region, International Business Research Institute,* University of Wollongong, Australia (CD-ROM).

Umhoff, B., & Winn, J. (1999, May 1). The healthcare profit pool: Who stands to gain and lose in the digital economy. *Health Forum Journal*.

von Lubitz, D., & Wickramasinghe, N. (2005a). Healthcare and technology: The doctrine of networkcentric healthcare. *International Journal Healthcare Technology Management, 4*, 322-343.

von Lubitz, D., & Wickramasinghe, N. (2005b). Networkcentric healthcare: Outline of the entry portal concept. *International Journal of Electronic Business, 1*, 16-28.

von Lubitz, D., & Wickramasinghe, N. (2005c). Networkcentric healthcare and bioinformatics. *International Journal on Expert Systems*.

von Lubitz, D., & Wickramasinghe, N. (2005d). Creating germane knowledge in dynamic environments. *International Journal of Innovation Learning.*

von Lubitz, D., & Wickramasinghe, N. (2006, January). Networkcentric healthcare: Strategies, structures and technologies for managing knowledge. In *Proceedings of the IRMA Conference*, Washington, DC.

von Lubitz, D., Wickramasinghe, N., & Yanovsky, G. (2005). Networkcentric healthcare operations: The telecommunications structure. *International Journal on Global Technology.*

von Lubitz, D., et al. (2004). Medical readiness in the context of operations other than war: Development of first responder readiness using OODA. In *Proceedings of the Loop Thinking and Advanced Distributed Interactive Simulation Technology Symposium*, Istanbul, Turkey. Retrieved May 5, 2006, from the Defence and National Intelligence Network, http://www.d-n-i.net/fcs/pdf/von_lubitz_1rp_ooda.pdf

Web Based Second Opinion. (n.d.). Retrieved May 5, 2006, from http://www.eclevelandclinic.org

Wickramasinghe, N., Fadlalla, A., Geisler, E., & Schaffer, J. (2004). A framework for assessing e-health preparedness. *International Journal on Electronic Healthcare (IJEH)1*(3), 316-334.

Wickramasinghe, N., Geisler, E., & Schaffer, J. (2005). Assessing e-health. In Spiel & Schuring (Eds.), *E-health systems diffusion and use: The innovation, the user and the use it model* (pp. 294-323). Hershey, PA: Idea Group.

Wickramasinghe, N., & Ginzberg, M. (2001). Integrating knowledge workers and the organization: The role of IT. *International Journal Healthcare Quality Assurance, 14*(6-7), 245-253.

Wickramasinghe, N., & Mills, G. (2002). Integrating e-commerce and knowledge management: What does the Kaiser experience really tell us? *International Journal of Accounting Information Systems, 3*(2), 83-98.

Wickramasinghe, N., & Sharma, S. (2004). Key factors that hinder SMEs in succeeding in today's knowledge based economy. *International Journal of Management and Enterprise Development, 2*(2), 141-158.

Wilson, T. (1999). Not a global village after all? Consumer behavior varies widely by country. *Internetweek, 792*(13), 792.

World Health Organization (WHO). (2003). Retrieved May 5, 2006, from http://www.emro.who.int/ehealth/

This work was previously published in Global E-Government: Theory, Applications and Benchmarking, edited by L. Al-Hakim, pp. 127-147, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 6.14
Inter-Organizational E-Commerce in Healthcare Services:
The Case of Global Teleradiology

Arjun Kalyanpur
Teleradiology Solutions, India

Firoz Latif
Teleradiology Solutions, India

Sanjay Saini
Harvard Medical School, USA

Surendra Sarnikar
University of Arizona, USA

ABSTRACT

Advances in healthcare information technology have enabled new models for electronic delivery of healthcare services. In this article, we present the case of electronic delivery of radiological services and describe the market-based and technological factors that have led to the development of Internet-based service models for flexible delivery of radiological services. Specifically, we describe the technical, regulatory, and security issues that affect teleradiology and propose a service delivery model for providing cost-effective and flexible radiological services.

INTRODUCTION

The continued wide-scale adoption of information technology in the healthcare industry is facilitating the electronic delivery of healthcare related services. Following the integration of medical infrastructure with information and communication technologies (ICT), patient-facing, and inter-organizational healthcare services can now be delivered using Internet-based applications and e-commerce platforms. In order to understand the factors influencing the evolution of e-commerce models in healthcare services, we present the case of global teleradiology.

Teleradiology refers to the electronic transmission of radiological images such as x-rays, computed tomograms (CT's), and magnetic resonance images (MRI's) across geographical locations for the purposes of interpretation and consultation. The digital radiological images are typically transmitted using standard telephone lines, satellite connections, or wide area networks (WANs). Teleradiology is an empowering technology and a facilitator for enhanced medical care. It enables a single radiologist to simultaneously provide services to several hospitals independent of their location and allows the exploitation of global time differences to provide emergency night coverage by personnel in a different time zone working a day shift. Additionally, the quality of care delivered by an alert physician working a day shift is far superior to that provided by a radiologist who is up all night. In addition, teleradiology also enables the delivery of subspecialty opinions to remote locations, where otherwise expertise is not available. In this article, we describe the current state-of-art in teleradiology, the benefits of the clinical practice of teleradiology, and the technical, regulatory, and security issues related to teleradiology. We begin by discussing relevant work in healthcare e-commerce and subsequently the evolution of global teleradiology.

E-Commerce in Healthcare Services

In the past decade, several new businesses have developed that deliver healthcare related services over the Internet. They can be broadly classified into four different forms of e-health business (Parente, 2000). These include portal, connectivity, B2B, and B2C applications. Portal, connectivity, and B2C commerce in healthcare typically involves either the provisioning of information to consumers via advertising supported Web sites, or the Web-based ordering services for prescription drugs (Zehnder, Bruppacher, Ruppanner, & Hersberger, 2004). In B2B e-commerce, the healthcare industry is mostly focused on the procurement and supply of medical devices and equipment (Arbietman, Lirov, Lirov, & Lirov, 2001; Smith & Correa, 2005). The exchange of services has been mostly limited to non-medical services such as billing and claims processing.

The exchange of medical services over the Internet has received little attention both in the industry and in academic literature. However, recent advances and developments in the healthcare industry and medical information technology are now enabling several new models for delivery of medical services over the Internet (Siau, 2003), primary examples of which include telemedicine and teleradiology. In addition, initiatives on data standardization and standardization of quality are also driving the development of outsourcing-based models in the healthcare industry (Segouin, Hodges, Brechat, 2005). Following the development of enabling technologies and the emergence of favorable market factors for medical service outsourcing, the outsourcing of medical services is now a major topic for discussion and research (Wachter, 2006). Among the first applications of medical service outsourcing and offshoring is the teleradiology service, and is now practiced by firms like NightHawk (http://www.nighthawkrad.net/), virtual radiologic (www.virtualrad.com), teleradiology solutions (www.teleradsol.com), etc.

Although the concept of teleradiology was first tested and clinically utilized in the late 1950's (Gershon-Cohen & Cooley, 1950), the high cost of transmission and the variability in digital imaging protocols limited the widespread adoption and application of teleradiology applications. However, the rapid progress in digital communication technologies and the development of efficient Internet-based software for image transmission, storage, and display in the 1990s has significantly reduced the technical barriers to teleradiology adoption. In addition to the previous developments, the universal adoption of the DICOM standard, as required by the ACR-NEMA (American College of Radiology and the National

Electronic Manufacturers Association), has enabled the wide spread adoption of teleradiology applications.

Global Teleradiology

Today, teleradiology has become both global and online. The precedent for globalization of teleradiology was set by the information technology services industry, which has pioneered the concept of the global office where the work "follows the sun" (i.e., divisions in different time zones are connected over a WAN to provide a 24-hour workforce without the need for having individuals to work a night shift at any location). Extended to teleradiology, this means that the interpreting radiologist for a given hospital can potentially be located anywhere on the globe and day-night time differences can be exploited to staff the emergency room (ER) night shift. A paper published by Kalyanpur, Weinberg, Neklesa, Brink, and Forman (2003) from Yale University showed that this was both technically and clinically feasible (Kalyanpur et al., 2003).

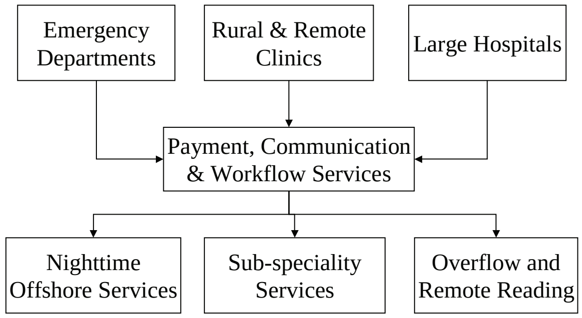
Several market-based factors such as staffing shortages (Sunshine, Maynard, Paros, & Forman, 2004), increases in imaging volumes (Bhargavan & Sunshine, 2002), new technology, and insurance and regulatory changes have also contributed to the growth and development of teleradiology. We briefly describe these factors next.

- **Radiologist Staffing Shortage:** A significant shortage of radiologists became apparent in the U.S. toward the turn of the millennium, with a large cohort of senior radiologists retiring from practice and training programs not having grown adequately to keep pace with the increased needs (Sunshine et al., 2004). In many other parts of the world, a similar shortage of trained/skilled radiologists exists. As might be expected in a time of staffing shortages, it is the night shift in the community hospital that is typically hit the hardest, although academic departments also stand to benefit from having their off-hours/overflow work done remotely by faculty based offshore (Kalyanpur et al., 2004).
- **Increase in Imaging Volumes:** The rapid evolution of faster imaging technologies such as multi-slice CT, faster MRI scanners, and sequences, coupled with an ageing population in the U.S. has led to a consistent per year increase in imaging volumes in recent years. A recent survey of radiology utilization in Medicare population indicates a 13% increase in the utilization of radiological imaging (Martin et al., 2003).
- **Increase in Trauma Imaging Utilization:** The development of newer applications of CT in the emergency setting has necessitated a large increase in 24-hour radiologist coverage at hospital emergency rooms (Spigos, Freedy, and Mueller, 1996). In addition, the Health Care Financing Administration (HCFA) requires that in order for the services to be billable, overnight coverage for radiological services has to be provided by a fully trained and certified radiologists, rather than residents/trainees (Federal Regulation 60: 63124). This has resulted in a large increase for overnight radiologist services.

A Service Delivery Model for Teleradiology

Recent advances in information and communication technologies such as Web services and workflow technologies enable new extensions to teleradiology that facilitate the delivery of radiological services across various time zones and geographical regions. The foremost application of teleradiology is in the emergency setting. In emergencies, teleradiology facilitates a prompt response by bringing the emergently performed images to the off-site radiologist that allows for

Figure 1. A service delivery model for teleradiology



timely diagnosis and the timely administration of appropriate treatment.

However, using secure Internet-based communication mechanisms and workflow technologies, several new applications of teleradiology can be developed. An overview of the service delivery model is presented in Figure 1. A key feature of the model is an e-commerce platform consisting of payment, communication, and workflow services. Under typical current arrangements for emergency teleradiology, the radiology group outsources its night calls to a teleradiology provider, somewhat like a locum relationship. The radiology group then reimburses the teleradiology provider at a predetermined rate for the preliminary emergency report provided by the latter and then bills the insurer for the final read that the former performs. By integrating a workflow system with the remote teleradiology service, all the previous payment, billing, and routing services can be automated.

In addition to automating the billing, payment, and routing services, an Internet-based service delivery model enables the delivery of radiological services to emergency departments, large hospitals, and remote rural clinics. We further describe these application areas next.

- **Emergency teleradiology services:** Teleradiology within the U.S. led to the development of the "nighthawk" concept wherein a radiologist is able to simultaneously provide services to multiple hospitals via teleradiology links to a central reading facility (often the radiologist's home). This enables a single radiologist to simultaneously staff the night shift at multiple hospitals. The process is cost-effective to hospitals, as the need to recruit night shift personnel is minimized.
- **Radiology to remote locations:** Teleradiology also enables the provision of radiological services to remote locations where the technical infrastructure for radiological scanning exists, but a radiologist is not available on site. Field studies indicate that teleradiology implementations at rural hospitals have been highly successful and cost-effective in providing radiological services even in real-time situations and have lead to improved patient care (Franken et al., 1995; Lee et al., 1998).
- **Optimization of workflow:** Teleradiology also increases the efficiency of a radiologist by ensuring that he or she spends the most part of his or her time delivering quality care to the maximum number of patients.

This is achieved by bringing the images to the radiologist rather than vice versa, thereby saving physician commuting time and increasing the range and reach of the radiologists' expertise. Within large radiologist groups servicing multiple hospitals and imaging centers, teleradiology also permits the optimal distribution of work based on need and the availability of radiologists.

- **Subspecialty consultations**: Teleradiology enables the wider availability of subspecialty consultations wherein images of a specific body region/modality need to be referred to the radiologist with expertise in the interpretation of that type of study. Past studies indicate that opinions of subspecialty radiologists obtained through teleradiology have resulted in significant improvements in the quality of patient care (Franken et al., 1997; Kangarloo et al., 2000).

Technical Requirements for Teleradiology

The key components of a teleradiology system include a picture archiving and communications system (PACS), a radiology information system (RIS), and a reliable and secure high-speed connectivity between the remote sites. These put together with standards for imaging and systems/procedures for security and contingency practices complete the technical aspects of teleradiology. These are further discussed in the following sections.

Picture Archiving and Communications System

An efficient Web-based PACS is the cornerstone of a clinical teleradiology practice. PACS is the information system used for the acquisition, storage, communication, archival, viewing, and manipulation of radiological images and related data. We provide a brief discussion on each of these components next.

- **Acquisition devices:** Acquisition devices are modalities with digital output capabilities or devices such as frame grabber and digitizers that convert the analog output from imaging modalities to a digital format.
- **Storage:** The storage components in a PACS system include both short-term as well as long-term archival solutions. This allows radiologists to have easy access to relevant prior studies for comparison. Typical radiological images range from 0.5MB to 4 MB or higher in size and depending on the number of images storage requirements can run to several gigabytes of storage per month. Since long-term archives do not require instant accessibility, optical disks or tape drives could be used for this purpose. The duration for which images need to be stored is dependent on the image type and varies from organization to organization. In the emergency setting, images are not usually stored for longer than 30 days. However, if outpatient imaging is being performed where comparisons are required, longer periods of storage may be necessary.
- **Communication:** PACS requires high-speed connectivity to enable rapid transfer of images to viewing workstations over LAN and WAN (the latter is what constitutes teleradiology). Transmission of images is based on protocols of imaging standards called DICOM. Transmission of non-image data like text uses HL7 standards. We further discuss DICOM and HL7 in later sections.
- **Software:** Image viewing and manipulation software are most often an integral part of PACS. Image viewers can be on a diagnostic workstation, review workstation, or could be done utilizing a Web-based module. Diagnostic workstations are high-end systems with high-resolution flat panel displays while

standard desktop PCs can be used as review workstations or as Web viewers.

- **Image Compression:** Given the large file size of typical radiological images, an important feature required facilitating rapid image transfer and throughput in teleradiology is compression. Compression algorithms used may be industry standard like JPEG 2000 or could be proprietary to the vendor. It has been noted that compression settings of up to 10:1 can be tolerated in clinical teleradiology without the compromise or loss of clinically relevant data, for review of CT images (Kalyanpur, Neklesa, Taylor, & Daftary, 2000). In the case of plain radiographs, even higher settings may be tolerated.

Radiological Information Systems

RIS is often a subsystem of hospital information system (HIS), but can also be a stand-alone entity and may or may not be connected to PACS. While PACS mainly deals with images, RIS/ HIS deals with data associated with patient demography, studies, and reports. RIS is often what guides the workflow of a teleradiology practice. Though RIS was used as a report generation and distribution tool earlier, commercially available RIS packages currently have integrated many features like voice recognition, staff scheduling, work distribution, invoicing, etc. At teleradiology solutions, a RIS has been developed by an in-house software development team to meet the requirements of an offshore teleradiology practice.

Connectivity Requirements

An ideal connectivity solution for teleradiology providers includes dedicated commercial broadband connection. The high-speed connections allow teleradiology service providers to serve clients half way around the globe, with report turnaround times comparable or even superior to the ones from local radiologists. A single radiologist working from home may use DSL connectivity, provided adequate throughput is confirmed prior to clinical use. For example, a single ultrasound image would take about 2 seconds to transfer over a typical DSL connection; however, larger plain film images can take between 1 and 4 minutes to transfer. Apart from bandwidth, one would also require other networking components such as routers, firewalls, VPN concentrators, and intrusion detection and prevention systems.

As the service being provided is a clinical service, typically in the emergency setting, a high level of communication between the site of origin and interpretation of the images is mandatory. This involves the utilization of fax systems capable of handling high volume data, direct telephonic contact and video and teleconferencing facilities.

Data Standards

DICOM is the industry standard used for transfer of radiological images between different hosts claiming conformance. DICOM is a standard developed by a joint committee set up by American College of Radiology (ACR) and National Electrical Manufacturers Association (NEMA). HL7 (Health Level Seven) is the standard for the exchange, management, and integration of electronic healthcare information like clinical and administrative data. Using HL7 compatible software streamlines the workflow considerably. For example, if PACS and RIS of a teleradiology practice are HL7 enabled, as soon as a new study is received by the PACS, a new order can be created on the RIS using patient demography and study details from the DICOM file, thereby eliminating duplication of work.

Securing Patient Data

Teleradiology providers are covered under the Health Insurance Portability & Accountability Act (HIPAA) and therefore need to implement adequate privacy and security practices as protected health information (PHI) and electronic protected health information (EPHI) are transmitted over public networks on a regular basis. While offshore teleradiology providers are not required to comply with HIPAA laws, the HIPAA act requires that "covered entities execute contracts that consist specific provisions for protection, use, and disclosure of health information" (Hilger, 2004). The privacy rule deals with all forms of patients' protected health information, whether electronic, written, or oral, while, the security rule covers only protected health information that is in electronic form, including EPHI that is created, received, maintained, or transmitted.

With the availability of a large number of commercial off the shelf network security solutions and technical safeguards, building and managing a secured network are relatively easy to implement. However, unauthorized access to patient data due to administrative breaches are difficult to monitor and implement, and hence the bulk of the HIPAA privacy rules have to do with administrative and physical safeguards. These breaches can only be prevented by educating employees about privacy and security practices followed by the organization and ensuring that these practices are implemented.

The security rule does not prescribe any specific technologies--being technology neutral allows the HIPAA covered entities to choose solutions based on their specific requirements. Technical safeguard standards include access control, audit controls, data integrity, person or entity authentication, and transmission security. Our in-house developed RIS solution (Tele-RIS) is used as a case in point here to illustrate/demonstrate how it complies with the above standards.

Access control mechanisms are used to ensure that a person or software can only access and modify resources for which the person is granted authorization. Access controls standard require the implementation of several features such as unique user identification, automatic logoff, and data encryption procedures. In addition, the following functionality also needs to be implemented:

- **Emergency access procedure:** This implementation specification requires a covered entity to establish procedures to retrieve EPHI during emergencies. Tele-RIS has a provision for adequate backup for power or network outages, but in case of a total outage, a backup procedure for the entire workflow from order entry to report distribution is well defined. Operating from different geographical locations also helps in dealing with contingencies.
- **Audit controls:** Are useful for recording and examining information system activity, especially when determining if a security violation occurred. Any addition, deletion, or remarkable modification on Tele-RIS is logged for audit purposes.
- **Data integrity:** Data could be altered during transit or in storage. "EPHI that is improperly altered or destroyed can result in clinical quality problems for a covered entity, including patient safety issues" (Health Insurance Portability & Accountability Act). This could happen by both technical and non-technical sources. A user can change the data, accidentally or maliciously. Tele-RIS addresses this problem using audit logs and techniques like versioning and digital signing.
- **Person or entity authentication:** Authentication involves confirming that users are who they claim to be. Tele-RIS uses a username/password combination for authentication. Many applications now incorporate the use of biometrics for authentication.
- **Transmission security:** This standard requires a covered entity to "Implement technical security measures to guard against unauthorized access to electronic protected health information that is being transmitted

over an electronic communications network" (Health Insurance Portability & Accountability Act).

Barriers to Adoption

Although teleradiology provides several benefits and can result in improved patient care and the technology exists for the large scale implementation of Internet-based radiological services, several barriers exist in the healthcare services industry that prevent the wide adoption of Internet-based radiological and medical service delivery. We briefly discuss these factors next.

- **Technology:** Although the technological capability to enable the delivery of medical services exists, the availability of the technology and trained technicians to operate them is a major barrier to wide-scale adoption. For example, the availability of reliable Internet connections in remote locations is a major barrier to the successful implementation of teleradiology and telemedicine projects. For a detailed discussion on the technological barriers to telemedicine, please refer to Paul, Pearlson, and McDaniel (1999).
- **Credentialing:** A major impediment to the rapid adoption of teleradiology services is the credentialing process. In addition to state-wise licensing requirements and board certifications, a hospital appointment for the consulting radiologist with the client hospital is necessary for providing radiological services. The current regulatory and credentialing structure is designed for a physical presence-based delivery model of radiological services, and needs to be redesigned for remote radiological or medical service delivery to develop.
- **Malpractice liability:** The medical malpractice liability laws differ from state to state and its implications to the practice of telemedicine are not clear. In addition, malpractice claims could also arise from improper use of equipment and errors in data transmission. For a detailed discussion on the legal issues related to the practice of telemedicine and teleradiology, the reader is referred to Gantt (1999).
- **Billing and reimbursement:** Although teleradiology consultations are now eligible for reimbursement under Medicare, HFCA rules limit the eligibility to patients who reside in a designated health professional shortage area (HPSA). In addition, regulations exist that prevent the billing of Medicare claims outside of United States, thus preventing the offshore reading services for Medicare related cases.
- **Health Insurance Portability & Accountability Act:** The confidentiality and security of patient information when transmitted across public networks is of major concern during the implementation of teleradiology solutions. Adequate safeguards need to be implemented in order to prevent unauthorized access to patient information at both local and remote locations. In addition to textual data, new encryption mechanisms need to be implemented to secure image data while maintaining their reliability (Cao, Huang, & Zhou, 2003).
- **Availability and work-satisfaction:** The availability and work-satisfaction of radiologist is a major factor in determining the success of teleradiology services. While a recent study of teleradiology satisfaction indicates a positive job-satisfaction for radiologists (Krupinski, McNeill, Haber, & Ovitt, 2003), the long effect of teleradiology on job satisfaction of radiologists needs to be evaluated.

CONCLUSION

In this article, we describe the state-of-art in teleradiology and present a service delivery model for providing cost-effective and flexible radiological services. The proposed model can enable the delivery of radiological services to several hospitals independent of their location and allows the exploitation of global time differences to provide emergency night coverage in emergency departments. In addition, we also describe the technical, regulatory, and security issues related to teleradiology and discuss barriers to the adoption of Internet-based teleradiology services. We conclude that while the technological capability exists to enable the remote delivery of radiological and other medical services, several legal, regulatory, and operational barriers remain that prevent the wide-scale adoption of Internet-based radiology and medical services.

REFERENCES

Arbietman, D., Lirov, E., Lirov, R., & Lirov, Y. (2001). E-commerce for healthcare supply procurement. *Journal of Healthcare Information Management*, *15*(1), 61-72.

Bhargavan, M., & Sunshine, J. (2002). Workload of radiologists in the United States in 1998-1999 and trends since 1995-96. *American Journal of Roentgenol*, *179*, 123-1128.

Cao, F., Huang, H. K., & Zhou, X. Q. (2003). Medical image security in a HIPAA mandated PACS environment. *Computerized Medical Imaging and Graphics*, *27*(3), 185.

Franken, E. A., Berbaum, K. S., Brandser, E. A., D'Alessandro, M. P., Schweiger, G. D., & Smith, W. L. (1997). Pediatric radiology at a rural hospital: Value of teleradiology and subspecialty consultation. *American Journal of Roentgenology*, *168*, 1349-1352.

Franken, E. A., Berbaum, K. S., Smith, W. L., Chang, P. J., Owen, D. A., & Bergus G. R. (1995). Teleradiology for rural hospitals: Analysis of a field study. *Journal of Telemed Telecare*, *1*(4), 202-208.

Gantt, A. H. (1999). Telemedicine and healthcare e-commerce. *The Maryland Bar Journal*, *32*(6).

Gershon-Cohen, J., & Cooley, A. G. (1950). Telognosis. *Radiology*, *55*, 582-587.

Goldberg, M. A., Sharif, H. S., Rosenthal, D. I., Black-Schaffer, S., Flotte, T. J., Colvin, R. B., & Thrall, J. H. (1994). Making global telemedicine practical and affordable: Demonstrations from the Middle East. *American Journal of Roentgenol*, *163*(6), 1495-500.

Health Care Financing Administration Medicare Program. (1995). Revisions to payment policies and adjustments to the relative value units under the physical fee schedule for the calendar year 1996 Fed Reg 60: 63124

Health Insurance Portability & Accountability Act, http://www.hhs.gov/ocr/hipaa/

Hilger, D. D. (2004). The HIPAA privacy rule and HR/benefits outsourcing: Does the business associate label belong on your recordkeeper? *Benefits Quarterly*, *20*(1), 44-52.

Kalyanpur, A., Casey, S., & Michel, E. (2002). Impact of overnight staff coverage of a university hospital emergency room via international teleradiology. RSNA, Chicago, IL.

Kalyanpur, A., Neklesa, V., Brink, J., Pham, D., Stein, S., & Forman, H. (2004). Implementation of an international teleradiology staffing model. *Radiology*, *232*, 415-419.

Kalyanpur, A., Neklesa, V. P., Taylor, C. R., & Daftary, A. (2000). Evaluation of JPEG and Wavelet compression for teleradiology transmission of direct-digital body CT images. *Brink Journal of Radiology*, *217*, 772-779.

Kalyanpur, A., Weinberg, J., Neklesa, V., Brink, J., & Forman, H. (2003). Emergency radiology coverage: Technical and clinical feasibility of an international teleradiology model. *Emergency Radiology*, *10*(3), 115-118.

Kangarloo, H., Valdez, J. A., Yao, L., Chen, S., Curran, J., Goldman, D., Sinha, U., Dionisio, J. D., Taira, R., Sayre, J., Seeger, L., Johnson, R., Barbaric, Z., & Steckel, R. (2000). Improving the quality of care through routine teleradiology consultation. *Acad Radiol*, *7*(3), 149-55.

Krupinski, E., McNeill, K., Haber, K., & Ovitt, T. (2003). High-volume teleradiology service: Focus on radiologist satisfaction. *Journal of Digital Imaging*, *16*(2), 203-209.

Lee J. K., Renner, J. B., Saunders, B. F., Stamford, P. P., Bickford, T. R., Johnston, R. E., Hsaio, H. S., Phillips, M. L. (1998). Effect of real-time teleradiology on the practice of the emergency department physician in a rural setting: Initial experience. *Acad Radiol*, *5*(8), 533-8.

Maitino, A. J., Levin, D. C., Parker, L., Rao, V. M., & Sunshine, J. H. (2003). Practice patterns of radiologists and nonradiologists in utilization of noninvasive diagnostic imaging among the Medicare population 1993-1999. *Radiology, 228*, 795-801.

Parente, S. T. (2000). Beyond the hype: A taxonomy of e-health business models. *Health Aff*, *19*(6), 89-102.

Paul, D. L., Pearlson, K. E., & McDaniel, R. R. (1999). Assessing technological barriers to telemedicine: Technology-management implications. *IEEE Transactions on Engineering Management*, *46*(3).

Segouin, C., Hodges, B., & Brechat, P. H. (2005). Globalization in health care: Is international standardization of quality a step toward outsourcing? *International Journal of Quality Health Care*, *17*(4), 277-9.

Siau, K. (2003). Health care informatics. *IEEE Transactions on Information Technology in Biomedicine*, *7*(1), 1-7.

Smith, A. D., & Correa, J. (2005). Value-added benefits of technology: E-procurement and e-commerce related to the health care industry. *International Journal of Health Care Qual Assur Inc Leadersh Health Serv, 18*(6-7), 458-73.

Spigos, D., Freedy, L., & Mueller, C. (1996). 24-hour coverage by attending physicians: A new paradigm. *American Journal of Roentgenol, 167*, 1089-1090

Sunshine, J., Maynard, C. D., Paros, J., & Forman, H. P. (2004). Update on the diagnostic radiologist shortage. *AJR*, *182*, 301-305

Wachter, R. M. (2006). The “dis-location” of U.S. medicine--The implications of medical outsourcing. *New England Journal of Medicine*, *354*(7), 661-665.

Zehnder, S., Bruppacher, R., Ruppanner, H., & Hersberger, K. E. (2004). Swiss community pharmacies’ on the Web and pharmacists’ experiences with e-commerce: Longitudinal study and Internet-based questionnaire survey. *Journal of Med Internet Res. 6*(1).

This work was previously published in Journal of Electronic Commerce in Organizations, Vol. 5, Issue 2, edited by M. Khosrow-Pour, pp. 47-56, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 6.15
A Metric for Healthcare Technology Management (HCTM):
E-Surveying Key Executives and Administrators of Canadian Teaching Hospitals[1]

George Eisler
BC Academic Health Council, Canada

Joseph Tan
Wayne State University, USA

Samuel Sheps
Western Regional Training Centre (WRTC) for Health Services Research & University of British Columbia, Canada

ABSTRACT

Among key drivers of healthcare reform in Canadian society are the challenges faced by the rapid rate of technological change and its impact on organizational performance in terms of efficiency, cost-effectiveness, and innovation in business and operational processes. However, despite the noted significance of the impact of technological change on healthcare organizations, the challenge of healthcare technology management (HCTM) has received only scattered and marginal attention in the technology management (TM) literature. The lack of formalization in HCTM construct, attributes, and measures motivated an empirical study to develop a metric for HCTM. This metric was then used to assess HCTM practices in teaching hospitals across Canada. The project began with an analysis of developments to date in the fields of Management of Technology and Management of Medical Technology. An extensive literature content analysis generated a set of definitions and attributes of the conceptual TM construct, which was eventually extended to

HCTM. A measuring instrument was developed through a formal design process involving expert panel review, pilot testing, instrument refinement, and field-testing to extract and measure HCTM performance indicators. Administration of this metric with the help of the Association of Canadian Academic Health Organizations via a Web-based survey of senior healthcare administrators provided insights into the HCTM status of Canadian teaching hospitals and its relationship with organizational performance.

INTRODUCTION

Amazingly, the history of digital computing and automated information processing technology has unfolded before us, for only a brief period of no more than several decades. In the last decade, we began to witness a gradual convergence of computing, telecommunications, and Web-based networking technologies, augmented by high-speed global access to an explosion of information on the Internet and increased global competition (Geisler, 2000; Ramanathan, 1990; Ulhoi, 1996). A key challenge with this trend is to unravel how management of technology innovation can impact business and operational performance of healthcare organizations in today's environment.

Business strategists, analysts, and researchers as well as economists have pointed to technology innovation as a catalytic change agent in the structure of industries and competition. Indeed, technological innovation can shift the competitive balance within an industry and create opportunities for growth. A technology management (TM) problem arises when business strategy does not fully incorporate technology-based threats and opportunities. Countries around the world are recognizing that the competitiveness of their private sector industries in the global market depends on their attention to TM. Technology has become a competitive tool in national and corporate survival, especially in an environment of global and more intense competition (Perrino & Tipping, 1989; Sharif, 1994). The issues are similar in the public service arena in that, "public expectations for the level and quality of government services… have grown while satisfaction with their fulfillment has steadily declined. In the past few years, it has become evident that cutting fat, eliminating waste, and preventing abuse is not nearly enough. Government needs to rethink its methods and restructure its approach to public services." As identified by Canadian healthcare CEOs, the task of managing healthcare organizations and systems is particularly complex, demanding that healthcare executives master many different skills, including government relations, community liaison, human resources, finances, patient care, research, and teaching. Communication skills, culture management, creativity, shared leadership, and alliance building rank among the top (Armstrong, Brunelle, Angus, & Levac, 2001). Healthcare technology management, therefore, adds one other dimension to these challenges.

In this paper, we review the extant literature with an attempt to link the more established field of Management of Technology (MOT) to emerging fields of Management of Medical Technology (MMT) and Healthcare Technology Management (HCTM). HCTM research domain and questions, focusing on the lack of alignment between healthcare technology strategies and healthcare systems goals and objectives, are first reviewed. Next, we outline the significance of HCTM research and attempt to link the published literature on MOT, MMT, and HCTM. We then discuss the methodological approach and statistical techniques used in the study. In analyzing and reporting the findings, we attempt to provide insights into characteristics and comparative performance of Canadian healthcare organizations and address key questions raised by this research with respect to aggregate views and perspectives of senior teaching hospital executives on HCTM. We close the paper with concluding remarks and directions for future studies.

LITERATURE REVIEW

In recent years, there has been increasing attention on the need to diffuse HCTM practices and expertise in developed as well as developing nations. The World Health Organization (WHO), for example, had alluded to serious shortcomings in the performance of health systems around the globe (World Health Organization, 2000). WHO's attempts to introduce components of a HCTM-like system in the late 1980s into countries and regions around the world had not been very successful.

HCTM Research Domain and Questions

The lack of a working HCTM model or framework and the shortage of TM training, knowledge, and expertise among workers in developed as well as developing countries were identified as serious limitations. Without an accepted and proven HCTM model, TM component functions and activities, such as technology planning, life cycle management, technology assessment, and performance assurance had no lasting infrastructure support and impact. Even outside the healthcare sector, these deficiencies with respect to TM point to the crucial need to align technology strategy and business strategy in organizations across many business industry sectors (Kaounides, 1999; Neumann et al., 1999). The strengthening, linking, and aligning of technology planning and business planning in healthcare is the essential purpose of HCTM. In light of this, our research concentrated on the status of HCTM among Canadian teaching hospitals. Specifically, key questions to be addressed included:

1. What is HCTM? What is the status of HCTM practices in Canadian teaching hospitals?
2. What attributes constitute successful TM activities in industries other than the healthcare sector? How can these attributes be applied to Canadian teaching hospitals to improve services without increasing costs?
3. What primary factors are adopted by hospital executive teams in practicing HCTM efficiently, cost-effectively, and in response to trends in hospital governance, infrastructure, resource limitation, health informatics leadership and culture, as well as changing medical and technological capabilities?

In an attempt to provide some valid and reliable answers based on empirical data collection and evidence gathered from the literature, we first operationally restate these questions into a series of more specific investigative queries. In this sense, the following questions serve as specific research aims or project objectives:

- How can TM and HCTM be conceptualized? What are these constructs and how do we define them operationally? More specifically, what are key and specific dimensions and indicators of the HCTM construct and how can these be determined empirically?
- How can HCTM performance indicators be measured? How can these measures be aggregated, grouped appropriately, and developed into a reliable and valid instrument for easy and convenient administration on the Web (as a Web-based survey instrument)?
- What is the level of HCTM capacity in Canadian teaching hospitals?
- Where are potential "weaknesses" and "specific performance gaps" among HCTM measures as applied to Canadian teaching hospitals?

Study Rationale and Significance

This section addresses the legitimacy of transferring TM lessons from other industry sectors to healthcare and hospital management, focusing on the concept of strategic TM and the importance of technology innovation. It also discusses the challenges of sustainability, cost, and quality of healthcare and incentives for technological innovation in healthcare.

The complexity of the healthcare environment, the multitude of forces that shape technology decisions, and the uniqueness of the healthcare environment are part of the justification for the promotion of the academic discipline of HCTM as separate and unique from MOT in general. Compared to other industry sectors, such as banking, housing, and transportation, the healthcare environment is seen not only as more complex, but also more emotionally charged (Eisler, Sheps, Satuglu, & Tan, 2002; Geisler & Heller, 1996). This environment is characterized not only by the complexities inherent in the development and maintenance of a seamless system spanning the continuum of healthcare delivery, but also by the complexity of relationships between provider organizations, service and product vendors, third-party payers, funders, and insurers, patients (consumers), the general public, regulators, researchers, and educators. In a 2005 *Communications of the ACM* publication, Tan et al. (2005) pointed to characterizing the present day healthcare and services delivery systems as "Complex Adaptive Systems". The need to understand the strategic TM concept in such a modern, complex system environment is clearly warranted.

The healthcare industry is an industry in transition, driven by such factors as changing population demographics, changing technology, and changing economic conditions. These pressures have resulted in changes to structure, process, financing, and human resource management. These challenges are driven by attempts to ensure timely access to quality and cost-effective healthcare services facilitated by technological innovation. Healthcare systems in Canada, the U.S., and other OECD countries are expected to continue on a road of cost reduction and quality improvement. Among OECD members, many countries regard cost control as a paramount policy aim (Gray, 1992). In Canada, federal and provincial governments are responding to such data as that provided by the Conference Board of Canada[2]. The Board projects that the proportion of Canada's population over the age of 55 will increase 10% per year for the next 20 years, rising from 22% to 32% of the population. Public health expenditures are projected to rise correspondingly from 31% in 2001 to 42% in 2020 as a share of provincial government revenue, representing an annual growth in healthcare costs of 5.2% over that period, of which 1.7% are attributed to demographic changes, that is, aging and population growth. Adjusted for inflation, per capita spending on healthcare is projected to increase by 58% while per capita spending on all other government services is projected to increase by only 17% over the next 20 years. Given these economic predictions, a debate about the role of technology is taking place against the backdrop of the incentive structures in publicly-funded, not-for-profit, for-profit, and mixed healthcare environments. It is, therefore, significant at this juncture to do a study to assess the status of HCTM in Canadian teaching hospitals.

Technology can play a vital strategic role in healthcare, as it does in other knowledge-based service industries like banking, entertainment, and other professional and technical fields. This is particularly true for information technology (IT) and communications technology, which can contribute significantly to improved management, cost effectiveness, customer service, and support. These applications have strategic implications, creating opportunities for new services, or for delivering existing services in new ways. It is for these reasons that the BC government has maintained IT and tele-health applications on its list of priorities at a time of severe cost reduction responses to the economic situation (British Columbia Ministry of Health Planning, 2002).

MOT, MMT, and HCTM

The field of MOT has been developed as an academic specialty in the arena of bio-medical engineering and sometimes cross-disciplined between business and IT engineering based on an evolving concept about the role of technology in organizations. Technology and its management have long been seen as critical factors to an organization's success and survival, particularly in competitive environments (Badawy, 1996; Morton, 1991; van Wyk, 1988). In public sector environments, MOT concepts are equally relevant because they help organizations to provide better service within available resources, to optimize organizational objectives, such as seamless access and quality, and to address special needs, societal problems, or political goals.

In the past, many industries saw IT as playing only supportive roles, thereby contributing simply to overhead costs. In short, these technologies were not seen as central to corporate goals and objectives. Today, new perspectives are being championed: Emerging technologies are seen as significant core enabling assets with major strategic implications for business survival and success. The power of integrative and converging technologies is blurring the boundaries between administrative and core business assets. Many CEOs now believe that such enabling technologies, if managed appropriately, can contribute significantly to the achievement of organizational objectives. As well, they may change fundamentally the way an organization functions and the way it relates to its industry sector, its sponsors, suppliers, and, perhaps most importantly, its customers and clients.

A common case example is that of the Internet and related Web Services. Many companies, realizing the power of the Internet, have set up shop online. From a marketing perspective, such unpredictable and rapid technological development creates a volatile technology push on the input side of organizations. For example, companies such as CVS Pharmacy and Rite Aids have gone online to protect their chain stores from losing customers to its growing list of competitors that have virtual storefronts. On the output side, customers expect reliable, consistent, safe, effective, and efficient services. The convenience of online shopping means that they can change their royalty to companies more easily and faster than queuing up at the retail store counter waiting to pay for their purchases. To these e-consumers, the technology and its applications should be seamless. The challenge then is for executives, in the face of increasingly difficult internal and external constraints, to employ management strategies to enable the organization to continuously transform the turbulent technology input into a customer-focused and appropriate output (Tapscott, 1996; Tapscott & Caston, 1993). According to McGee and Thomas (1989), what has been missing "is a comprehensive view of how technological change can affect the rules of competition, and the ways in which technology can be the foundation of creating defensible strategies for firms." Driven by the need to compete more aggressively and efficiently in global markets, new ways of doing business has emerged, including shared responsibilities and programs restructuring, buyouts, and takeover campaigns, and various forms of online collaborative and joint venture arrangements. Studies have indicated that levels of companies' investments in technology may explain international differences in productivity and in shares of world markets.

In contrast to the disciplinary tradition of MOT, Canadian healthcare researchers have yet to pay attention to the more recent work on MMT. In an attempt to develop a new intellectual space, Geisler and Heller (1998a, 1998b) were among the first to suggest the application of the TM construct to the healthcare field. The scope of MMT advocated by Professor Eli Geisler encompasses both clinical and administrative technologies (and functions) of the healthcare system. The premise is that owing to economic pressures, the health services delivery system is in crisis unless it can further, and more substantially, benefit from the increasing role that technology has played. It is, therefore, argued that proper and better management of medical technology is critical in overcoming the challenges faced in future healthcare. Even so, the terminology and constructs in the MMT field are still poorly understood due to varied interpretations of the individual components of these constructs and the lack of commonly accepted definitions of the constructs. Before measures for these constructs can be established, Geisler (2000) argued that clear definitions and indicators have to be determined.

THE HCTM Construct

"Health", as defined by the WHO, refers to a state of total physical, mental, and social well-being, not just the absence of disease or infirmity. It is now recognized that population and individual health has many determinants not traditionally associated directly with the healthcare system. Correspondingly, the concept of "healthcare technology" includes applications of know-hows, methods, tools, and techniques that influence the environment, health information dissemination, healthy lifestyle, care protection, and disease prevention. It goes beyond technologies applied in modern acute care systems or for direct medical care. In this context, it applies broadly to facilities, information, devices, processes, and drugs, from the simplest to the most complex, along the entire continuum of healthcare. Technologies that may contribute to quality or sustainability of healthcare systems could be associated with direct patient care, infrastructure, or business processes. It seems, therefore, appropriate to refer to the management of these various and diverse forms of technologies as HCTM.

Notwithstanding, our content analysis of current MOT literature leads to the conclusion that there still is no commonly accepted definition of the TM construct, much less the HCTM construct. On the one hand, Anderson (1993) infers that the lack of agreement on a definition might be a barrier to the growth of the TM concept even though it is expected that a set of favored definitions will emerge over time. On the other hand, there has been some agreement that MOT is a multi-faceted discipline, linking engineering, science, and management concepts and theories in a holistic, systematic, and integrative fashion, and that TM should shape and support organizational strategic goals and objectives.

Following a comprehensive review of the literature across several industry sectors, we found that the multi-faceted TM concept yields a number of general but important observations (Eisler, 2002): (1) TM forms the basis for a technology strategy; (2) TM is characteristic of a technology-ready organization; and (3) TM encompasses the responsibilities and capabilities of the chief technology officer (CTO).

1. TM forms the basis for a technology strategy; the technology strategy is based on:

- The competitive environment, the organization, and technology; consideration of firm-specific factors, environmental factors, and customer preferences; creation of strategic advantage, technological expertise, the decision-making process, and organizational capabilities; comprehensive rethinking and readjustment of job descriptions, information systems, organizational structure, incentives, and decision-making processes; and
- Corporate objectives identified; organizational structure is one of the most important issues; policies hold together a decentralized, virtual workplace with direct access to global information; in addition to flexible organizational structures, management emphasizes information flow, incentives, and different performance assessment schemes;

centralization versus decentralization is one of the single biggest organizational issues.

2. TM is characteristic of technology-ready organizations; technology-readiness strategy is characterized by:

- Top management vision, foresight, and entrepreneurial spirit; leadership is the most critical aspect; commitment to knowledge acquisition rather than product development; alignment with strategic planning and integration of related functions; management must know what it wants, given the difficult-to-quantify costs and benefits of new technologies and attributes like flexibility; set the goals, understand the product/market interactions, being clearly aware of resources, constraints, and risks;
- Decisions and attitudes are based on analysis of competitive position, market intelligence, technical preferences of customers and internal capabilities; focus on the customer replaced organization-centered approaches; emphasis is on market pull rather than technology push;
- Management systems focus on an integrated enterprise; coordination across functional boundaries has priority over efficiency-driven divisions within functions; cross-functional approaches facilitate convergence of divergent views between technical- and marketing-oriented individuals; full and meaningful worker and customer participation in the production process is assured; and
- Process management has replaced product management; focus is on flexibility, adaptability, responsiveness, and effectiveness rather than efficiency and costs; competitive advantage comes from technology and strategy, not from savings in labor costs; able to change, to adapt, and to avail itself of new opportunities; effectiveness is considered with efficiency.

3. TM encompasses the responsibilities and capabilities of the CTO; characteristics of the CTO include:

- Being responsible for visionary leadership, organization, funding, alignment with objectives, bridging between and among operational units, planning, resource allocation, development of standards, rapid reorganization when necessary, leading adoption and implementation of fundamental organizational change; must also be steward of inter-networked leadership, be close to the business front line, and build an invisible enabling infrastructure;
- Assures that promises made on behalf of technology are kept; builds a viable, productive, and flexible technology asset base; gets back to basics and delivers the goods; takes responsibility for managing technology-driven organizational change, for learning what can be done and how to apply it, and for acting as a change champion; and
- Capabilities include managing in an environment of decentralized decision making with a high level of inter-functional coordination; must have commitment, technical competence; capable of skills to effect and manage change; conversant in business and organizational matters, as well as with technical information; understands the importance of systems that provide a competitive edge and the need for systems that support the goals of the organization.

RESEARCH METHODOLOGY

The study reported here was conducted in three phases. Phase I, content analysis, concentrates on identifying the constructs and related critical MOT capabilities and attributes. Phase II, the instrument development, emphasizes the selection of items, expert panel review, and metric refinement and validation. Phase III, the national survey, involves the administration of the field-tested instrument via a Web survey of hospital executives.

Phase I: Content Analysis

A rigorous content analysis was used to develop and contribute to a definition of the TM construct and to identify related critical MOT capabilities and attributes. Lewis (1993) refers to the methodology of "content analysis" as a "common technique employed in the social sciences to draw inference from text; it is executed by objectively and systematically extracting attributes from written communication and by analyzing those extracted parts". Essentially, a broad and comprehensive review of both published and unpublished sources forms the basis of this methodology[3].

Definition of the TM Construct

The database of 255 articles and dissertations was searched for the sampling unit "defin*" (for definition) in its abstract or full text, to locate discussions of the definition of "technology management". Thirty such articles were found, which were in turn scanned to eliminate those that did not in fact discuss or contain a definition, leaving seven articles. Although only one dissertation contained the root word "defin*" in its abstract, eight more dissertations with discussions about the definition were located by detailed reading of 47 dissertation abstracts.

A summative definition of the conceptual TM construct emerged from this exercise:

Technology management can be defined as a holistic and integrated application of engineering, science, and management capabilities to strategic life cycle management of new and relevant product and process technologies in order to shape as well as accomplish the goals and objectives necessary for organizational success.

From this definition, a schematic diagram for the TM Framework emerged, as shown in Figure 1. Organizational success in healthcare could, for example, be expressed as improved patient care within given resources or as protection of the principles of the *Canada Health Act*.

Phase II: Instrument Development

TM is a complex, multi-faceted process with interlinked activities, clearly containing more than one indicator. Following an abstraction ladder approach to metric development, the goal was to create a hierarchy of major dimensions, of variables that described the dimensions, and of indicators that gave shape to the variables. The measurement instrument itself was constructed from an exhaustive list of individual attributes, which were the result of the literature content analysis.

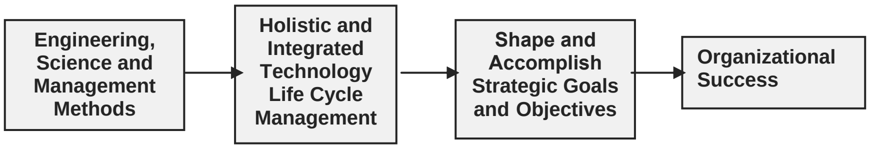
A long list of indicators was initially formed. These were grouped and regrouped iteratively until an intuitively sensible and comprehensible list of 26 indicators emerged. The list of indicators was far more detailed than similar reported attempts in the literature, although it became apparent that the indicators could be grouped into variables comparable in name to ones referred to in the literature.

The Pilot Test

Using the Delphi approach, senior healthcare administrators and management consultants were asked to rank and validate the attributes of the HCTM construct from the perspective of Canadian teaching hospitals with a focus on the responsibilities at the corporate or executive level. This pilot process constituted an expert panel review and a more specific content validity assessment.

In order to qualify as expert, an individual had to have been involved in healthcare management at senior levels for more than 10 years. Their collective professional backgrounds were to cover

Figure 1. Technology management framework



technical, administrative, and patient care arenas. They were provided with a draft survey and were asked to critique it from the perspectives of clarity and understandability. These experts were also asked to recommend unambiguous descriptors for the state of implementation of each statement category in the survey. The positions held by these experts at some point in their career were: Deputy and Assistant Deputy Minister, CEOs, VPs, and CIOs of hospitals, Directors and Executive Directors of Canadian health industry organizations.

The experts were informed about the national character and the expected target audience of the study. They were presented with the prototype questionnaire including 42 indicators, and were asked to:

- Complete the survey keeping any familiar large healthcare organization in mind:
- Identify if and how specific statements should be altered;
- Add statements that they thought were important but not included;
- Express concerns they may have about the appropriateness of the three main domains and the categories within each domain of the survey; and
- Identify examples of additional real and practical evidence or indicators that could be used to support statements about the extent of implementation in the various categories (strategic plans, job descriptions, budgets, etc.).

At this stage, the objective was to test the comprehensiveness and comprehensibility of the metric.

The Field Test

For further refinement of the HCTM metric, a field test was undertaken, incorporating feedback from the pilot study. It was targeted at institutions that would not be included in the survey of teaching hospitals. Senior managers of four regional referral organizations, who represented a variety of administrative responsibilities, participated.

The CEOs of four medium-size BC health service organizations provided the e-mail addresses of their executive teams and the individuals reporting to the executive team members (VPs, CEOs). A total of 64 senior managers, representing eight different areas of responsibility, received a cover letter serving as the consent form designed and based on recommendation of the respective Human Investigative Committees (HICs) as well as the survey instrument. With 33 responses from 64 individuals on the e-mail list, overall response rate was 52% and varied between 43% and 64% for the four participating institutions. As in the pilot test, respondents were asked to comment on content, wording, and structure of the instrument. No changes to the instrument were suggested. The importance of the HCTM topic itself was confirmed by a consistently high average rating from each organization. On the 6-point Likert scale, average ratings for each organization ranged from 5.6 to 5.9, where 5 means “great importance” and 6 means “very great” importance.

In addition to the Likert ratings for each indicator, a single measure was established to capture each respondent's perception of the "gap" between the ideal extent of implementation (reflected in a rating of 6 on the Likert scale) and perceived extent of implementation (1 to 6 rating) for each of the 26 indicators. Gap scores were weighted using the respondent's assessment of the importance of the respective indicator.

The result is a new variable (Gap Score) calculated by the formula:

$$GS = (6 - E) \times B,$$

where

GS = gap between ideal and perceived implementation
6 = ideal implementation rating
E = rating for perceived extent of implementation
B = rating for the perceived benefit/importance of implementation

A Gap Score (GS) was therefore calculated for each respondent. The term (6 – E) can be interpreted as an indication for "room for improvement", as an E score of "6" would indicate full implementation. The factor B becomes a weighting factor relative to the perceived importance of the item. For example, perceived full implementation of an indicator would yield a gap score of zero: $GS = (6 - 6) \times B = 0$.

A rating of "1" (meaning: not at all) for implementation would yield the highest gap score if the indicator was also rated as "6" for importance (meaning: very great): $GS = (6 - 1) \times 6 = 30$.

The possible range for gap scores was, therefore, 0 to 30. An average gap score (average of responses from a particular organization) of less than eight for an indicator was interpreted to mean that this indicator had been implemented to a high level and would not require additional attention for improvement. This could be either because the indicator had been well-implemented (rating of 5 or 6) or because the importance was rated lower than other indicators (4 or less). Field test results indicated that the resulting metric and the gap score measure seemed to capture differences in perception between senior managers from different organizations with some consistency.

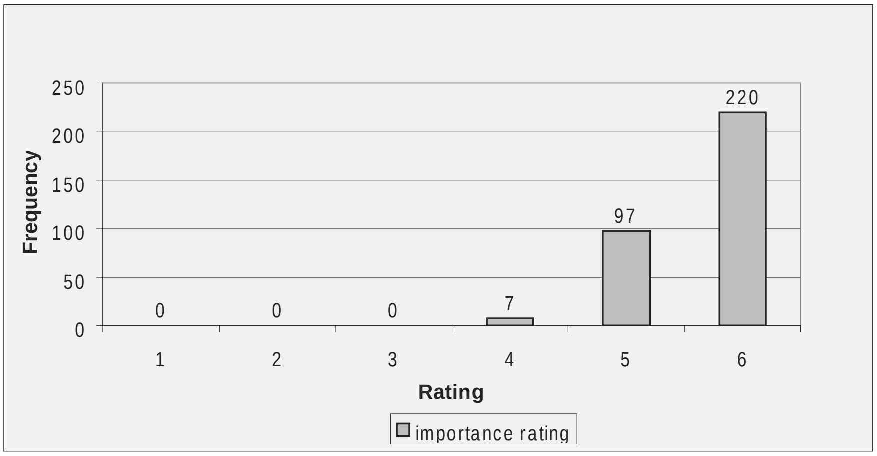
As summarized in Table 1, responses from organizations 1 and 2 did not result in an average gap score of less than 10 for any of the 26 indicators in the metric. Results of the pilot test on the individual indicators (numbered 1 to 26) for computing average gap scores are shown below. For example, organizations 1, 2, and 4 had average gap scores of 15 to 19 on 8, 10, and 6 indicators respectively, while none were recorded at the higher gap score for organization 3.

One could also interpret Table 1 as indicating that, compared to organization 3, organizations 1 and 2 are not performing well in the eyes of their senior managers. Also, senior managers in organization 4 have a less consistent view of their organization's performance. organization 3 seems to have reached a high level of HCTM

Table 1. Distribution of indicators by gap score range

Average Gap Score Range (high, low)	**Indicators**			
	Organization 1	**Organization 2**	**Organization 3**	**Organization 4**
6 - 10 (low)	none	none	1,3,10,11,12,13,14, 20,21,22,24	2,3,4,6,10,14,16,2 5
15 - 19 (high)	4,8,15,19,20,21,26	3,4,11,13,17,18,2 1,23,25,26	none	9,13,18,19,23,26

Figure 2. Topic importance rating



performance. Coincidentally, organization 3 was nationally recognized for its quality management system.

Phase III: National Survey Administration

The resulting survey design was influenced by the choice of a Web-based delivery system. The survey was presented in a logical sequence according to the overall model laid out for participants in the cover letter. Given that the target audiences for the survey were highly experienced and knowledgeable individuals, it was deemed unnecessary and probably not helpful in this study to scramble the sequence of the statements.

The Executive Committee of the Association of Canadian Academic Health Organizations (ACAHO) had enthusiastically endorsed this research and acknowledged its importance. Executive members, the CEOs of some of Canada's largest teaching hospitals, shared the view that HCTM approaches in the Canadian healthcare system could and needed to be strengthened. Their support and active promotion of this project presented an invaluable opportunity to attract participation from senior managers in Canada's teaching hospitals. Following the decision of the ACAHO's Executive Committee, which included the CEOs of six of Canada's largest healthcare organizations and ACAHO's Executive Director, to sponsor this survey, letters were sent to the CEOs of member organizations asking for their organization's participation.

The list of ACAHO member organizations was compared with the list of organizations ranked by a number of "technology intensive beds" from the Canadian Health Assotiation guide. This list indicated that ACAHO member organizations provided a national target audience of the majority of highly technology intensive organizations. It was felt that the senior managers of these organizations were best positioned to assess HCTM practices in Canada.

The Web-Survey Procedure

E-survey was chosen in this study because:

- Senior hospital administrators are expected to have access to Web-based technology.

- The spirit of the HCTM topic warrants an application that would better serve the customers; speed of data collection, convenience for the recipients, ease of data handling and online data analysis are among the key influencing factors on making this choice.

Interestingly, the most time-consuming aspect of the process was determining and iteratively correcting hundreds of e-mail addresses. The process of developing and posting the survey on the Web involved a number of additional steps. The e-survey content was developed as a MS Word document, including all associated notes and rating scales. The Web pages with forms to accept data input were designed with Dreamweaver 4.0 Web design software.

The decision to design the Web-survey tool rather than to use purchased software was based on the ability to design a more professional look and the need to create data files of submitted responses that could easily be exported directly into MS Excel program for convenient storage and rapid analysis. The e-survey was tested internally and

Table 2. Model of indicators

(a) Strategic Management	(b) Management of Change and Innovation	(c) Organizational Management
Technology Strategy 1. Tech and business strategies aligned 8. Technology linked to customer needs **Chief Technology Officer** 3. Executive member formulates technology strategy 4. Executive member administers/manages technology strategy **Knowledge Management** 5. Vision re technology through routine scanning 6. Environmental factors re technology strategy identified	**Customer Focus** 9. Line managers responsible for customer service 13. Performance management based on customer satisfaction **Change Management** 10. Senior management participates in change management 11. Corporate learning culture prevails 12. Senior management promotes organizational vision/direction 19. Organization has strategy to respond to change **Integration of Innovation Chain** 14. Communication system promotes technology 15. Innovation system integrates R&D/operations 20. Organizational design uses multi-functional teams **Human Resource Management** 17. HR development/planning key to technology capability 18. Accomplishments formally recognized 21. Skilled Employees seek innovative opportunities	**Effectiveness and Flexibility** 2. Key technology supported by infrastructure 16. Technology diffusion/transfer encouraged **Operations Management** 22. Organization uses sound project/process management 23. Technology life cycle systematically managed **Assessment and Evaluation** 24. Technology related risks evaluated 25. Technology assessment informs technology decisions 26. Performance due to technology management evaluated

externally before it was posted live on the British Columbia Institute of Technology (BCIT) server via the BCIT School of Health Sciences Web site. It was sent electronically with the cover letter allowing convenient links to the e-survey URL to hundreds of senior managers for whom e-mail addresses had been secured and verified.

Finally, a "thank-you" page was generated and sent automatically upon receipt of a survey response to the Web server. Participating organization and individuals were assured of the privacy of their identities and that any published reports would not identify and link organizations, individuals, or responses.

RESEARCH RESULTS

Owing to space limitation, issues of instrument validity, consistency, and reliability are only highlighted here prior to discussing the study results[4].

First, content validity was optimized through the iterative process in which the instrument was developed. Content analysis informed the metric development and expert opinions in pilot and field test stages refined the instrument.

Furthermore, validity and reliability of the metric was strengthened through:

Table 3. Gap score differences between clusters 1 and 2

Indicator	Difference	% Difference	Cluster 1	Cluster 2
7	0.61	6.54	9.33	8.73
22	0.80	7.46	10.73	9.93
10	1.24	13.15	9.43	8.19
9	1.53	13.53	11.31	9.77
25	1.89	14.26	13.25	11.36
17	2.00	15.85	12.62	10.62
8	2.34	16.03	**14.60**	12.26
2	2.15	16.17	13.30	11.15
21	2.33	17.72	13.15	10.81
20	2.36	18.50	12.76	10.39
23	2.58	18.57	13.89	11.31
15	2.59	18.82	13.76	11.16
24	2.20	19.00	11.58	9.38
6	2.49	20.66	12.05	9.56
19	3.27	21.48	**15.22**	11.95
11	2.54	21.77	11.67	9.13
18	3.34	24.38	13.70	10.36
14	2.83	24.50	11.55	8.72
16	2.94	25.28	11.63	8.69
1	3.93	28.98	13.56	9.63
13	4.02	28.98	13.87	9.85
12	2.69	29.89	9.00	**6.31**
5	3.98	31.61	12.59	8.61
26	4.86	31.52	**15.42**	10.55
4	5.71	48.15	11.87	**6.15**
3	7.29	56.42	12.92	**5.62**

- Supporting the instrument development process with a comprehensive literature review and content analysis (face/content validity);
- Refining the metric based on initial faculty suggestions and on comments and suggestions received from experts in the two-stage pilot and field testing (content validity);
- Clustering items on the instrument following expert panel review on its content as well as on comprehensibility, understandability, clarity, and structure of the instrument (construct validity);
- Factor analyzing responses during repeated testing of the instrument for key dimensions (construct validity); and
- Triangulating the results via a comparison of findings with a performance review of the same institutions using a completely different metric, which was conducted completely independently (concurrent validity).

Last, the Cronbach's alpha coefficient was also computed to determine internal consistency of the metric. It was applied to each of the dimensions resulting from the factor analysis. The reliability of the instrument was assessed for each of the indicators according to the factors established by the factor analysis. Reliability was high for all indicators with alpha ranging from 0.79 to 0.94.

Survey Responses

Thirty-three ACAHO member organizations were invited to participate in this project. The offices of the CEOs invested considerable energy in putting these lists together. In the end, approximately 850 individuals in 28 organizations received the survey; of these, 324 individuals responded, representing an average response rate of approximately 38% per participating institution with a range of 24% to 58%.

As noted, owing to privacy and confidentiality concerns, participating organizations were coded from 1 to 33. This is how the individual organization will be identified for the remainder of the discussion. For example, organization 1 submitted 36 responses, reflecting a 33% response rate relative to the number of senior managers who received the survey. Of these respondents, six executive members had patient care responsibilities (VP Nursing, VP Medicine, etc.) and seven did not (VP Finance, VP Support Services, etc.). Of the responding managers reporting to VPs, eight had patient care responsibilities and 15 did not. The respondents were first asked to rate the importance of the topic of the survey (1 indicates "not at all", 2 indicates "very little", 3 indicates "little", 4 indicates "some", 5 indicates "great", and 6 indicates "very great"). As indicated in Figure 2, 317/324 = 98% rated the topic's importance as great or very great. No respondent rated the importance at less than 4 (some) on the 1 to 6 rating scale.

One of the key questions to be answered by this research was the ability of the metric to differentiate between organizations relative to their HCTM approach. As no significant differences in gap scores between the categories of respondents were found, it indicated that all responses from the same organization could be grouped for the purpose of testing for significant differences between organizations. The 20 organizations that supplied five or more responses were included in this part of the analysis. Based on ANOVA and the resulting p-values, significant differences ($p < 0.001$) was found in the way in which senior managers across the 20 organizations perceived the state of HCTM in their organization. This suggests the need to further investigate the possibility that senior managers with direct patient care responsibilities responded differently from those that did not have direct patient care responsibilities. The same applies to the question of whether executive team members (VPs, CEOs) would respond differently from those senior managers reporting to VPs.

Table 4. Indicator distribution

Gap score quartiles	Cluster 1 indicators	Cluster 2 indicators
6.0 - 8.4		**3,4,12**,10,
8.5 - 11.0	12,10,22 1	6,14,11,24,6,1,9,13,22,18,20,26,17,21
11.1 - 13.5	9,14,24,16,11,4,6,5,20,3,21,25,2,1 2	,15,23,25,19,8
13.6 - 16.0	18,15,13,23,**8**,**19**,**26**	

Factor analysis was applied to identify variables that reflect some common underlying factor or dimension. Table 2 shows three underlying factors (or "latent variables") on which the 26 indicators "loaded": (a) strategic management; (b) management of change and innovation; and (c) organizational management.

Cluster analysis was used to group "like" hospital organizations based on their responses ("like" based on indicator gap scores) to optimize the comparison and identification of differing HCTM practices and capabilities. A practical three-cluster solution resulted out of the 20 organizations with more than five responses:

- Cluster 1: Organizations 1, 2, 3, 5, 6, 7, 9, 11, 13, 14, 17, 18, 19
- Cluster 2: Organizations 4, 8, 10, 12, 16
- Cluster 3: Organizations 15, 20

Testing for significant differences in-between clusters indicated that there was a statistically significant difference ($p < 0.01$) between the cluster mean gap scores for each of the three dimension variables. Since cluster 3 consisted of only two institutions with comparatively fewer responses, further comparative analysis was focused on the differences between clusters 1 and 2. The significant differences between the two clusters of organizations, based on senior managers' perception, are influenced by some variables more than others.

Table 3 lists the 26 variables in the survey in the order of the percent difference between cluster 1 and 2 mean gap scores. As illustrated, it shows that the average gap scores for each and every indicator were lower for cluster 2 organizations. Percent differences range from 6.54% to 56.62%. The two indicators 3 and 4 with the highest percent difference were also the ones with the highest variability on a national basis.

The data revealed what constituted the major difference among managerial responses between the two clusters of organizations. On the one hand, cluster 2 managers rated their organization very highly (low gap score) in the following three areas:

- Indicator 3: Executive member formulates technology strategy
- Indicator 4: Executive member administers/ manages technology strategy
- Indicator 12: Senior management promotes organizational vision/direction

On the other hand, cluster 1 managers rated their organization particularly low (high gap scores) in the areas:

- Indicator 8: Technology linked to customer needs
- Indicator 19: Organization has strategies to respond to change
- Indicator 26: Performance due to technology management evaluated

Table 4 portrays the distribution of indicators in the four quartiles of the gap score range.

Apparently, cluster 1 managers rated none of the indicators in the lowest quartile of gap scores compared to four indicators for cluster 2 managers. In contrast, seven indicators were rated in the highest quartile for cluster 1 as opposed to none by cluster 2 managers. It is interesting to note that indicators 8, 19, and 26 were also the ones that were deemed to need the most attention based on the overall national average. They also top the list of indicators needing attention in cluster 1 organizations. The largest differences in reported perception based on implementation ratings and gap scores between cluster 1 and cluster 2 managers related to indicators 3 and 4 of our HCTM model.

It can be concluded that the resulting metric is able to distinguish between differences in perception by senior managers about their organization's HCTM standard. Significant differences between clusters of organizations can be identified consistently across all three dimensions of the HCTM construct. In particular, the perceived presence of "Chief Technology Officer" roles in the organization seems to contribute strongly to the differences in mean gap scores between clusters. Other "strategic management" and "change management" variables seem to be perceived as being implemented to a greater extent in cluster 2 institutions. Improvements seem to be possible along all three dimensions.

The Hay Group Study

As indicated, the purpose of studying HCTM status is to assess its impact on health organizational performance in terms of alignment of IT goals with the strategic business goals and objectives of the organization. For a private sector organization, this may be expressed as competitive advantage, market share, profit, return on investment, or other such measure. What would be the metric that would indicate organizational "success or survival" of our study hospitals? What difference would it make that one cluster of hospitals seems to manage technology better than another cluster?

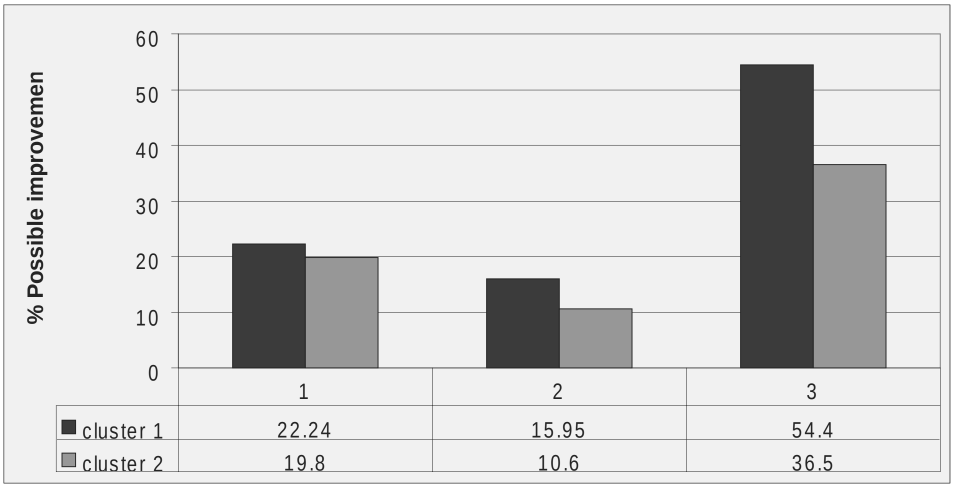
Fortunately, ACAHO provided independently generated information about operational efficiency and clinical efficiency of the same study hospitals (and others not included in the study) on a confidential basis. The Hay Group in Toronto compiled the data in an annual report entitled Benchmarking Comparison of Canadian Hospitals. The Hay study was commissioned by the hospitals in an effort "to improve the efficiency, effectiveness and quality of their care processes". The comparisons were based on CIHI hospital separation data, on accounts and statistics reported to Ministries of Health, and other data provided by the hospitals.

In order to determine if there was any association between the HCTM performance of the two clusters of hospitals identified in this study and other measures of performance, the Hay study outcomes for the same two clusters were analyzed. The following two summary measures, as defined in the report, were used:

- The measure for the overall clinical efficiency of a hospital is the percentage of inpatient days that can be reduced if a hospital were to achieve benchmark levels of performance. The smaller the percentage, the more efficient the hospital.
- The measure for the overall operational efficiency of a hospital is the potential reduction in operating cost to the 25th percentile performance, including direct care, administrative, and support functions. Subject to some caveats, the hospital with the smaller potential reduction can be considered more efficient with respect to the areas examined in these comparisons.

If an average gap score of 6 (over all responses and all indicators for each organization) was considered as the benchmark target, the potential HCTM improvement could be calculated and

Figure 3. Hay study comparison



expressed in terms of potential percent reduction. This measure was used in the comparison. As demonstrated in Figure 3, the Hay study comparison results provided independent validation of our findings with respect to organizational performance of the two clusters of hospitals studied and, at the same time, provided support of the concurrent validity of our HCTM metric.

CONCLUSION

Altogether, there has been recurring and increasing interests between both the academic and practitioner communities about the potential impact of HCTM on healthcare organizational performance. Like the management of other key strategic resources, such as health human resource and financial resource management, HCTM should and could become a more easily understood concept and its impact more conveniently assessed. In the past, the role of IT in healthcare has been poorly understood. Technology innovation, in particular, has sometimes been portrayed as a key contributor to escalating healthcare costs, as an expensive clinical tool that is primarily driven by some key decision maker's interest, or as a "hype" promoted by a vendor with few benefits but many empty promises. If it really is to support the goals and objectives of the healthcare system, IT in healthcare has to be better managed. This study contributes to a more systemic approach to understanding and even measuring the impact of IT in healthcare and has generated new knowledge about appropriate HCTM practices in teaching hospitals.

The study used a cross-sectional survey involving cluster sampling. The target sample was representative of senior administrators in the Canadian healthcare system as it comprises executive team members of the 40 largest Canadian teaching hospitals. Selected individuals from this group responded to a newly developed HCTM metric. Factor analysis and cluster analyses were the two key statistical inference methods applied to the resulting data set. Major clusters of respondents were those associated with individual organizations. These responses were analyzed to identify systemic differences in HCTM practices among different clustering organizations. For in-depth analysis of the data, the Analysis of Variance (ANOVA) technique was applied. It assumed that respondents from within each organization could be treated as one cluster. Unfortunately, field test analysis was inconclusive on this question due to limited sample size. Yet, the outcome of this study provided new thinking on HCTM, a theoretical contribution to the advancing fields of MOT and MMT. It also recognizes and bridges a multi-disciplinary science base with the broad modern concepts of healthcare, technology, and management.

This study, however, did not develop a new definition for the construct "technology". Rather, the perspective of technology as an *extension of human and organizational ability* was adopted from the literature. Based on an extensive literature content analysis, the study does propose a new definition and a theoretical model for the HCTM construct. The major critical dimensions of the HCTM metric include: (a) strategic management factors; (b) management of change and innovation factors; and (c) organizational management factors. Interestingly, two indicators (Indicators 3 and 4 of our metric) focus on the function of a chief technology officer (CTO) were found to be the largest differences in reported perception based on implementation ratings and gap scores between high performing and low performing teaching hospitals. Indicator 3 states: "A designated member of the executive team is responsible for the formulation of the organization's technology strategy" while Indicator 4 states: "A designated member of the executive team is responsible for the administration and management of the organization's technology strategy." This result strongly confirmed the message from the literature about the necessity of executive attention and leadership for HCTM. Along with providing leadership, coordination, and facilitation, the responsibilities of the CTO include such activities as gatekeeping, advocacy, funding, sponsorship, policy and procedure development, promotion, capacity building, and overseeing the technology management system. Not all hospitals have a CTO, in which case the responsibilities discussed here are either taken up by the CEO or, more likely, delegated to someone else such as the chief financial officer or chief operations officer. This person's job is to understand the strategic business issues, the customers, and the technology. She or he should be an effective leader and command the respect of his or her employees, managers, and peers. Our finding implies that a good CTO is key to differentiating among clustering organizations as to their HCTM status.

Another set of indicators that were shown to be critical for poor performing hospitals and presented the highest need for improvement on a national, system-wide basis is the set of Indicators 8, 19, and 26. Indicator 8 states: "Technology is linked to clearly identified customer needs and priorities." Indicator 19 states: "The organization has strategies to respond flexibly and rapidly to technological change." Finally, Indicator 26 states: "The organization's performance as a function of TM activities is routinely evaluated and benchmarked." A high gap score on Indicator 8 implies a lack of ongoing access to information about current and future needs and priorities of customers and staff. Indicator 19, similarly, addresses how quickly the organization responds to shifts in technological trends. A high gap score here implies that the organization ignores "market pull" strategies as part of its customer-relations policy, essentially going along with a "technology push." Product development times, innovation cycles, and overall cycle times for putting ideas and innovations into practice are longer than expected relative to industry norms. Lastly, a high gap score for Indicator 26 implies a lack of systematic performance reviews for different aspects of HCTM. These might include their impact on overall organizational goals, objectives, and customer service, including, for example, annual performance evaluations of the executive fulfilling the CTO role. Apparently, our results tell us that poor performing hospitals must pay particular attention to these factors.

Moreover, these results are very similar to the issues raised regarding patient safety and how it is to be enhanced given the recent dramatic increase in literature on this topic. TM and HCTM principles identified here are almost identical to the governance requirements in high reliability industries (aviation, nuclear power, etc.) which are also technologically intensive. Indeed, lack of thoughtful adoption and use of technology is a major factor identified in the patient safety literature. Thus, the lessons learned from this research could have a direct bearing on approaches to enhancing patient safety in healthcare settings.

In summary, our findings indicate that there are key differences in HCTM sophistication among Canadian teaching hospitals. Major differences occur in areas of strategic technology management, followed by change management, and, to a lesser extent, organizational management. The perceptions of senior managers are not significantly influenced by their area of responsibility relative to patient and non-patient care, or by their position in the reporting structure. Improvements are needed generally in all areas addressed by the HCTM metric. A further question relates to the impact of HCTM sophistication on organizational performance and success. This study was able to explore, to some extent, the relationship between clinical and operational efficiency and HCTM performance. It did not address the more strategic impact of HCTM performance on customer satisfaction and customer service levels relative to customer needs.

Future Research

This study built on earlier research in the areas of MOT and MMT. An outcome of this study is the enhancement and the expansion of the current research agenda for the MOT and MMT disciplines. Research in this area should be focused on development of more specific measurement items, refinement of the measures and applications of these measuring tools. For example, an expanded application of our HCTM metric is to go beyond teaching hospital settings.

Another stream of research could further explore the metric relative to particular health systems issues at national, provincial, and institutional levels of management. While knowledge could undoubtedly be expanded in each of these facets, some examples of priority research questions for HCTM metric refinement include:

- Who are the customers in healthcare, what are their needs, and what order of priority is attached to those needs?
- There has been mixed success with the HCTM as a *tactical* resource, what particular challenges arise out of *strategic* HCTM?
- While senior managers in teaching hospitals provided their perspectives on HCTM; how can IT contribute to clinical and operational efficiencies, quality and levels of care, and customer satisfaction?
- The study results are based on perceptions of senior managers. How can follow-up case studies be designed to compare perceptions with reality?
- Case studies would also be useful with respect to the real, rather than perceived, differences between some cluster 1 and cluster 2 hospitals. To what extent are CTO functions explicitly established, given the large difference in perception between the clusters on this point?
- The Web-based survey approach was very successful. Language and design decisions, however, did impact on response rate. Participants had to complete all questions. They had to choose a rating between 1 and 6. Some would have preferred an “N/A” or “don’t know” category. Translation into French might have generated more responses from Quebec.

Moreover, the metric could be used to probe into the strengths and weaknesses of an individual institution. A case study with full management and staff participation would allow further statistical validation of the instrument. The motivation for undertaking this study was that in Canada's healthcare organizations, technology was not managed well enough, neither tactically nor strategically. The CEOs of some of Canada's largest teaching hospitals expressed support for this notion when they agreed to promote and support this study among their peers. Of the 324 surveyed senior managers, a vast majority of 317 confirmed the importance of HCTM for Canadian healthcare managers by rating it as "great" to very great (between 5 and 6 on a 6-point scale), the remaining managers rated it as being of "some" importance (rating 4).

Finally, other key research questions emerging from this study are issues related to the definition and identification of customers, customer groups, and customer needs in healthcare. Organizations like teaching hospitals are faced with competing customer interests (provincial services, community services, research, and education services). Resource planning of any kind, human, technology, or financial, is difficult under circumstances where the primary customers and their needs are not well-defined. Lately, with the emergence of e-health, a whole new scenario of customer/consumer relationships is unfolding with significant implications for providers and agencies of healthcare services (Tan, 2005).

REFERENCES

Andersen, D. F., Belardo, S., & Dawes, S. S. (1994). Strategic information management: Conceptual frameworks for the public sector. *Public Productivity & Management Review, XVII*(4), 335-53.

Anderson, P. (1993). Toward exemplary research in the management of technology: An introductory essay. *Journal of Engineering and Technology Management, 10*(1, 2).

Armstrong, R, Brunelle, F., Angus, D., & Levac, G. (2001). The changing role of Canadian health care CEOs: Results of a national study. *Health Care Management Forum Supplement.*

Badawy, M. K. (1996). A new paradigm for understanding management technology: A research agenda for "technocologists". *International Journal of Technology Management, 12*(5, 6), 717-732, 721.

British Columbia Ministry of Health Planning. (2002). *A new era for patient centred health care; Building a sustainable, accountable structure for delivery of high-quality patient services.* BC Ministry of Health Planning publication.

Eisler, G. (2002). *Health care technology management (HCTM): An assessment of its application in Canadian teaching hospitals.* Unpublished doctoral dissertation, University of British Columbia, Canada.

Eisler, G., Sheps, S., Satuglu, S., & Tan, J. (2002). *Health technology management in Canadian teaching hospitals: An empirical investigation.* Paper presented at the Hospital of the Future Conference, Chicago, IL.

Geisler, E. (2000). *The metrics of science and technology.* Westport, CT: Quorum Books.

Geisler, E., & Heller, O. (1996). *Managing technology in healthcare*. Boston: Kluwer Academic Publishers.

Geisler, E., & Heller, O. (1998a). Management of medical technology (MMT): Research, education, and practice. *International Journal of Technology Management, 15*(3, 4, 5), 196-210.

Geisler, E., & Heller, O. (1998b). *Management of medical technology: Theory, practice, and cases.* Boston: Kluwer Academic Publishers.

Gray, B. H. (1992). Why nonprofits? Hospitals and the future of American health care [see comments]. *Frontiers of Health Services Management, 8*(4), 3-32; discussion 43.

Kaounides, L. C. (1999). Science, technology, and global competitive advantage. *International Studies of Management & Organization.*

Lewis, B. R. (1993). *The information resource management concept: Domain, measurement, and implementation status (technology planning).* Unpublished doctoral dissertation, Auburn University. UMI Dissertaton Services.

McGee, J., & Thomas, H. (1989). Technology and strategic management progress and future dire. *R & D Management, 19*(3).

Morton, M. S. (1991). *The corporation of the 1990s: Information technology and organizational transformation.* New York: Oxford University Press.

Neumann, L. C., et al. (1999). Achieving success: Assessing the role of and building a business case for technology in healthcare. *Frontiers of Health Services Management.*

Perrino, A. C., & Tipping, J. W. (1989). Global management of technology. *Research Technology Management, 32*(3), 12.

Ramanathan, K. (1990). Management of technology: Issues of management skill and effectiveness. *International Journal of Technology Management, 5*(4).

Sharif, N. (1994). Integrating business and technology strategies in developing countries. *Technological Forecasting and Social Change, 45*(2), 151-167.

Tan, J. (2005). *E-Health care information systems—An introduction for students and professionals.* San Francisco, CA: Jossey-Bass.

Tan, J., Wen, J., & Awad, N. (2005). Health care and services delivery systems as complex adaptive systems. *Communications of the ACM, 48*(5), 36-44.

Tapscott, D. (1996). *The digital economy: Promise and peril in the age of networked intelligence.* New York: McGraw-Hill.

Tapscott, D., & Caston, A. (1993). *Paradigm shift: The new promise of information technology.* New York: McGraw-Hill.

Ulhoi, J. P. (1996). Towards a theoretical and methodological corporate technology management framework. The strategic perspective. *International Journal of Technology Management, 12*(2), 199-209, 201.

van Wyk, R. J. (1988). Management of technology: New frameworks. *Technovation, 7,* 341-351.

World Health Organization. (2000). *The World Health Report 2000: Health systems: Improving performance.* Geneva, Switzerland: World Health Organization.

ENDNOTES

[1] The original version of this paper received the Best paper/presenter award in the Hospital of the Future Conference, Chicago, August 2002. Authors appreciate the feedback from reviewers.

[2] The Conference Board of Canada is an independent, not-for-profit research organization with a mission to help its members to "anticipate and respond to the increasingly changing global economy".

[3] Readers interested in the details of this methodology should also consult Lewis (1993).

[4] Interested readers may refer to the PhD dissertation linked to this study (Eisler, 2002).

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 1, Issue 1, edited by J. Tan, pp. 1-21, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 6.16
Applying Strategies to Overcome User Resistance in a Group of Clinical Managers to a Business Software Application:
A Case Study

Barbara Adams
Cyrus Medical Systems, USA

Eta S. Berner
University of Alabama at Birmingham, USA

Joni Rousse Wyatt
Norwood Clinic, Inc., USA

ABSTRACT

User resistance is a common occurrence when new information systems are implemented within health care organizations. Individuals responsible for overseeing implementation of these systems in the health care environment may encounter more resistance than trainers in other environments. It is important to be aware of methods to reduce resistance in end users. Proper training of end users is an important strategy for minimizing resistance. This article reviews the literature on the reasons for user resistance to health care information systems and the implications of this literature for designing training programs. The other principles for reducing resistance (communication, user involvement, strategic use of consultants) are illustrated with a case study involving training clinical managers on business applications. Individuals responsible for health care information system implementations should recognize that end user resistance can lead to system failure and should employ these best practices when embarking on new implementations.

INTRODUCTION

Traditionally, health care has lagged significantly behind other industries in the use of information technology (Parton & Glaser, 2002). The use of computers in health care has up until recently been primarily to automate the business and administrative functions. Today, a variety of pressures are forcing the health care industry to invest more money and effort into using information technology in clinical settings. New legislation to protect privacy and confidentiality of medical information encourages the development of electronic medical records (US Dept of Health and Human Services, 2002). Concerns over medical errors have led to an increased interest in clinical decision support systems and computer-based physician order entry (Bates et al., 1999; Leapfrog Group, 2000). These developments will lead to more need for direct use of computers by health care providers who are used to manual processes for the same tasks. In addition, many clinicians are now assuming managerial positions in health care where they will be expected to use traditional business applications as well (Merry, 1999). Not only are these new managers not used to automating some of these tasks, but as clinicians, they have not seen use of the computer as part of their professional role. As a chairman of a clinical department once said, “What do I have a secretary for?”

The reluctance to use the new systems may be perceived as resistance, or in fact there may be real resistance to the changes that information technology makes in the clinical work processes (Worthley, 2000). In either case, administrators, information technology personnel, or clinicians charged with promoting the use of information technology in the health care environment may encounter more resistance than is found in other environments, so added to the issues of training end users that are common across a variety of settings, health care project managers also need to be aware of methods to reduce resistance in end users (Kaplan, 1997).

User resistance is a common occurrence when new information systems are implemented within a health care organization. There is also a sizable amount of literature on health care system implementations to explain and give insight into some of the reasons behind this resistance and to suggest strategies for overcoming it (Ash et al., 2000; Jiang et al., 2000; Lauer et al., 2000; Lorenzi & Riley, 1995; Lorenzi et al., 1997, 2000; McNurlin & Sprague, 1998; Worthley, 2000). The principles advocated in these studies can be used to develop a variety of end user training programs in health care settings. We will discuss this literature and the implications for design of training programs and will illustrate the application of these principles with a case study that involved training clinical managers on business applications.

POSSIBLE EFFECTS OF USER RESISTANCE

When computers were first introduced into the health care setting, the technical issues were the most important to address, but as Lorenzi et al. (1997) discuss, now that many technical issues have been solved the managerial, organizational and “people” issues are equally, if not more, important. Worthley (2000) attributes the failure of many system implementations to the failure to properly address user resistance. He discusses five different forms that resistance can take: sabotage of computer equipment; employees being absent or late to work; “badmouthing” the system, not using the new system and continuing to use the old system; and data tampering (Worthley, 2000, p. 170).

REASONS FOR USER RESISTANCE

Kaplan (1997) discusses some of the causes of resistance to information systems. She mentions “user-centered” theories, which “consider resistance to be due to factors inherent in users, such as their lack of knowledge or their reluctance to change" (Kaplan, 1997). Other researchers have also identified resistance to change as a significant source of implementation problems (Ash et al., 2000; Brown & Coney, 1994; Jiang, Muhanna & Klein, 2000; Lorenzi & Riley, 2000; Yaghmaie, Jayasuriya & Rawstorne, 1998).

In addition to resistance to change, both Lorenzi and Riley (1995) and Worthley (2000) discuss some other reasons for user resistance:

- Fear of loss of prestige and status in the organization due to not knowing the new information systems (Lorenzi & Riley, 1995; Worthley, 2000)
- Pressure to develop new skills (Lorenzi & Riley, 1995)
- Pressure of higher performance expectations (Lorenzi & Riley, 1995)
- Fear of loss of social interaction with other workers (Worthley, 2000)
- Historical reasons, such as a previous bad experience with an information technology effort (Worthley, 2000)
- Benefits may not be clear to the user (Ash et al., 2000).

STRATEGIES FOR A SUCCESSFUL INFORMATION SYSTEMS IMPLEMENTATION

Based on this literature review, the following were recurrent themes as strategies to overcome user resistance:

- *Communication* - Several researchers name communication as one of the most important strategies for a successful systems implementation (Jiang et al., 2000; Krishnan, 1999; Lauer et al., 2000; Lorenzi & Riley, 1995; Worthley, 2000).
- *User involvement* – Ives and Olson (1984) found as a result of reviewing over 20 articles that "participation leads to increased user acceptance and use by encouraging realistic expectations, facilitating the user's system ownership, decreasing resistance to change, and committing users to the system."
- *Clarification of benefits* - A strategy discussed by Ash et al. (2000) "is to make sure the system provides immediate benefits to users." Ash et al. suggest that telling the user what the short and long term benefits will be for that individual user will motivate a person to use the system more than telling the user what the overall benefits of the system will be for the organization.
- *Role of consultants* - Consultants can be beneficial in information systems implementation by filling in the experience and knowledge gaps of their clients (Bauman, 2001). By being an outsider, they are not involved in office politics and their decisions are based on what is in the business' best interest, not the best interest politically (Bauman, 2001). Very often a credible outside consultant can get a user to adopt new practices that internal managers have tried unsuccessfully to implement.
- *Training* - Training has been one of the main topics researchers have emphasized as essential to successful information systems implementation (Ash et al., 2000; Jiang et al., 2000; Lauer et al., 2000; Lorenzi & Riley, 1995; Lorenzi et al., 1997, 2000; McNurlin & Sprague, 1998; Worthley, 2000). Both Lorenzi and Riley (1995) and Worthley (2000) advise the use of the "just-in-time" training concept when training users on a new information system. This means that training should occur just prior to implementation. They also suggest training the users in the order they are going to use the system. There is a danger of training users too early and then finding that the users have forgotten

much of what they learned and/or are not as familiar with the product when the actual implementation occurs (Lorenzi & Riley, 1995). This means that the implementation schedule should be carefully monitored and the training plans should be revised if there is significant delay.

Lorenzi and Riley (1995) suggest that training address both technical content and attitudes. "Any training needs to be a combination of educating people in how to use the system plus building their enthusiasm for doing so" (Lorenzi & Riley, 1995). Lorenzi and Riley (1995) and Worthley (2000) discuss the importance of training manuals and online help. Complex systems may require special training, such as training in stages (Lorenzi & Riley, 1995). Breaking up the training will allow the users to get comfortable with one part of the system and may help to build confidence in learning the more complex parts (Lorenzi & Riley, 1995). Researchers have mentioned that physicians may require specialized training (Ash et al., 2000). Physicians "will rarely commit to group training sessions of any length" (Lorenzi & Riley, 1995) and "frequently like to be trained by other physicians" (Ash et al., 2000). One method that Lorenzi et al. (1997) suggest when training physicians on a new system is to design a training program that adapts to their current work styles. One idea is to use "training aids that are prepared on 3x5 index cards because most physicians are accustomed to keeping pertinent information in this manner" (Lorenzi et al., 1997).

CASE STUDY

We applied these principles in designing a training program in the use of the Microsoft Excel application to clinical managers of an outpatient medical clinic. Although all of the principles discussed above in regard to major systems implementation are not applicable to the small case study, many of them are. In addition, many of the same reasons for resistance can be found with small business applications as with large informational technology implementations. Below we describe the background to the request for assistance, the application of the principles in the design of the training program, and the outcomes.

BACKGROUND

Southern Medical Clinic (not its real name) is a multi-specialty group practice started in 1926. Currently, the clinic has been home to several of the physicians for over 20 years. One particular physician has been there for over 40 years, over half of the entire life of Southern Medical. Likewise, the support staff boasts many that have been with the facility for 25, 30, and even 40 years. In health care today such statistics are rare. Southern Medical clinic still operates as a multi-specialty, physician owned clinic. They still use paper medical records, they still house a completely handwritten master patient index, they have a paper card catalogue of all the patient charts, and they even make paper copies of all insurance remittances. For such a facility to continue to thrive, it must not only let technology in the door, but also embrace what technology has to offer, yet it is also to be expected that with such a stable staff there may reluctance to do things differently from the way they have always been done. Southern Medical Clinic has overcome much resistance already. All of the practice management, accounting, HR, claims filing and payroll are electronic. Internet classes are held and spreadsheets have just replaced old handwritten forms. They will soon be scanning all remittance reports to a server, and one department is toying with the idea of electronic medical records. The board meetings now have physicians with pocket PCs and the preferred means of mass communication is e-mail.

The problem that formed the focus for the case study was the reluctance of the clinical department managers to learn to use Excel. The managers had been encouraged to use Excel for a variety of tasks, including documenting employee leave, which they currently monitored manually. They were provided with spreadsheets from the human resources department that they were expected to use. Although a few were interested in learning how to use the spreadsheets themselves, most of the managers were not extremely computer literate and appeared to prefer to calculate the leave manually and ask their secretaries to enter the data into the spreadsheet. Obviously such a practice has a greater opportunity for introducing error and is less efficient than directly entering and calculating the leave with the spreadsheet, yet the managers were not eager to change the system with which they were familiar. It was with this background that the clinic manager requested outside consultation and training for her staff.

APPLICATION OF PRINCIPLES FOR OVERCOMING RESISTANCE

The strategies in the literature review were applied to developing the training program. A strategy discussed by Bauman (2001) was to bring in outside consultants. The trainer did not have any affiliation with the health care organization and did not know any of the clinical managers in the training classes. When preparing the training program, an effort was made to make sure that the users would see the personal benefits of using the program as advocated by Ash et al. (2000). To accomplish this, the training program included examples of some current processes that the managers performed manually and the managers were told that the goal of the training was to work on automating them. Some of the benefits of using the software were emphasized and the managers were told how they would be able to use the software as soon as they went back to work. Another strategy that was incorporated into the training program was based on Lorenzi and Riley's (1995) suggestion to break up the training into stages. There were three training sessions scheduled and each session progressively covered more complex material. Lorenzi and Riley (1995) discussed how breaking up the training in stages will allow the users to get comfortable with one part of the system and help to build confidence in the users when they are faced with the more complex parts.

Prior to taking the Excel training the department managers were required to pass a proficiency test in basic computing and Microsoft Word. Out of the 18 health care managers in this organization who were eligible to take the Excel class, 12 enrolled and 11 completed the class. There were three sessions, each two hours long, completed over a three-week period, with multiple sessions given each week to accommodate small groups of the managers. The training groups were mixed as far as level of computer experience. An evaluation form was developed to get feedback on the teaching and to assess any resistance that the health care managers might have had towards learning and using Excel. The evaluation form was distributed at the end of the final training session. The instructor informed the managers that their employer would not see the individual evaluation forms.

Outcomes

The 11 health care managers completed the evaluation form about the Excel training class. On the whole they were very positive. Over 80% responded positively to nine of the 10 questions. The exception still was positive, in that eight of the managers felt the class was less difficult than they thought it would be and only one thought it was more difficult. What was particularly interesting was that all of the managers agreed that their interest in using Excel was increased (73% responded Agree and 27% responded Strongly Agree), and that they would use Excel in their work (55% Strongly Agree; 45% Agree). The respondents were very positive about the instructor, with all of them interested in taking another class with her. Approximately nine months after the training session, the actual extent of use of the software was assessed. Results showed over half of the trained managers were now using the software. Of those who were not, 18% did not have access to a computer. Twenty-seven percent of the original group of trainees were still resisting using the software.

DISCUSSION

Given the expectation of encountering resistance, it was surprising that during the class as well on the final evaluation this resistance did not appear at all. The training program was designed and developed in an effort to minimize user resistance and appeared to have done so. When designing the Excel training program, the one overall goal, from the perspective of the organization, was for the managers to produce a sick and vacation report in Excel. This is a report that all of the managers were required to do and all were performing this report manually. In the first class managers were told that the sick and vacation report was the overall goal and they responded positively. During the classes, there were conversations about reports and processes that they perform manually and discussions about how they could use Excel to automate some of those other processes as well. The basics of Excel were covered in the first class. In the second and third class, on the focus was working the leave report. Showing the managers how Excel could help automate a time consuming manual report helped to motivate them to use and learn Excel. By being an outsider to the organization, the instructor was not encumbered by any relationships with the users. The users knew that how they did in the training classes was not going to affect how they are evaluated or perceived in their jobs. Thus, using outside consultants to provide a different perspective and increase credibility, designing the program to illustrate personal benefits, and focusing the training on gradually building up confidence in use resulted in very positive evaluations from a group of users perceived to be resistant to learning new technologies.

Although it would be gratifying to think that the attention to evidence-based good practices in instructional design were solely responsible for the positive outcomes, there are other explanations. It is possible that the most resistant users did not take the class, since a third of the eligible managers did not take advantage of the training opportunity. In addition, sometimes users' lack of confidence using a computer system can be perceived as resistance to the system. It may have actually been the clinical managers' lack of confidence using Excel that was perceived as resistance. The training classes may have helped to build their confidence and they realized it was not as difficult as they thought it would be.

CONCLUSION

As health care systems begin to implement new information technology, end user resistance is likely to become increasingly common. There are a variety of reasons for the resistance, but there are also strategies that have been shown to minimize it. These strategies include communication, user involvement and training, with specific focus on increasing user confidence. These strategies can be used with large-scale implementations and can be incorporated within the training programs themselves. Individuals responsible for health care information system implementations should recognize that end user resistance can lead to system failure and should employ these best practices when embarking on new implementations.

REFERENCES

Ash, J.S., Anderson, J.G., Gorman, P.N., Zielstorff, R.D., Norcross, N., Petit, J., et al. (2000). Managing change: An Aanalysis of a hypothetical case. *Journal of the American Medical Informatics Association, 7,* 125-134.

Bates, D.W., Teich, J.M., Lee, J., Seger, D., Kuperman, G.J., Ma'Luf N., et al. (1999). Impact of computerized physician order entry on medication error prevention. *Journal of the American Medical Informatics Association, 6*(4), 313-321.

Bauman, C. (2001). Readying your organization for new technology. *IT Health Care Strategist, 3*(7).

Brown, S.H., & Coney, R.D. (1994). Changes in physicians' computer anxiety and attitudes related to clinical information system use. *Journal of the American Medical Informatics Association, 1,* 381-394.

Ives, B., & Olson, H. (1984). User involvement and MIS success: A review of research. *Management Science, 30,* 586-603.

Jiang, J.J., Muhanna, W.A., & Klein, G. (2000). User resistance and strategies for promoting acceptance across system types. *Information Management, 37,* 25-36.

Kaplan, B. (1997). Addressing organizational issues into the evaluation of medical systems. *Journal of the American Medical Informatics Association, 4,* 94-101.

Krishnan, R.S. (1999). Without nurses, inventory control impossible. *Hospital Materials Management,* 24, 10-14.

Lauer, T.W., Joshi, K., & Browdy, T. (2000). Use of the equity implementation model to review clinical system implementation efforts. *Journal of the American Medical Informatics Association, 7,* 91-102.

Leapfrog Group for Patient Safety. (2000). http://www.leapfroggroup.org/

Lorenzi, N.M., & Riley, R.T. (1995). *Organizational aspects of health informatics: Managing technological change*. New York: Springer-Verlag.

Lorenzi, N.M., & Riley, R.T. (2000). Managing change. *Journal of the American Medical Informatics Association, 7,* 116-124.

Lorenzi, N.M., Riley, R.T., Blyth, A.J.C., Southon, G., & Dixon, B.T. (1997). Antecedents of the people and organizational aspects of medical informatics: Review of the literature. *Journal of the American Medical Informatics Association, 4,* 79-93.

McNurlin, B.C., & Sprague, R.H., Jr. (1998). *Information xystems management in practice*. New Jersey: Prentice Hall.

Merry, M. (1999). Wanted: A new breed of physician drivers for healthcare's nitroglycerin trucks. *Frontiers of Health Services Management, 15*(4), 29-35.

Parton, C., & Glaser, J.P. (2002, July). Myths About IT spending. *Healthcare Informatics.* http://www.healthcare-informatics.com/issues/2002/07_02/myths.htm

US Department of Health and Human Services. (2002). *Administrative simplification*. http://aspe.hhs.gov/admnsimp/

Worthley, J.A. (2000). *Managing information in healthcare: Concepts and cases.* Chicago, IL: Health Administration Press.

Yaghmaie, F., Jayasuriya, R., & Rawstorne, P. (1998). Computer experience and computer attitude: A model to predict the use of computerised information systems. *Medinfo' 98, Proceedings of the Ninth World Congress on Medical Informatics, 9*(2), 895-899.

APPENDIX: LESSON PLAN FOR EXCEL TRAINING SESSION 1

Introduction

- Define what Microsoft Excel is and what it is used for.
- Excel is a spreadsheet application that allows you to perform quick and accurate calculations on data that are entered into a worksheet. Using Excel also helps to avoid errors and present data in a professional format.
- Discuss how Excel can be beneficial to them in their daily jobs.

Objectives for Session 1

- Learn how to open and create an Excel workbook.
- Be able to identify the components of the Excel screen.
- Learn how to use the menus and toolbars to perform commands in Excel.
- Be able to enter text and values in a worksheet.
- Learn how to close and save an Excel file.

Assignment

- In class, we will create a calorie counter worksheet. This will involve the students entering text and data. They will have a brief introduction to simple formulas. We will use the auto sum function and average function in this assignment. Next session will cover formulas in more detail.

LESSON PLAN FOR EXCEL TRAINING SESSION 2

Introduction

- Question and Answer section from last session
- Review of what we covered in the last session
- Discussion of formulas in Excel

Objectives for Session 2

- Learn how to perform calculations in Excel by the using shortcut formulas and by entering formulas.
- Be able to edit and delete data in Excel.
- Be able to cut, copy, and paste in Excel.
- Learn how to move cells around in Excel using the drag and drop method.

Assignment

- In class, we will begin work on a sick and vacation report that will allow them to apply skills learned from the previous lessons. The report will include entering text, values, and formulas. The objective is for them to use the worksheet in their jobs to keep up with the sick and vacation hours for each of the employees in their department.

LESSON PLAN FOR EXCEL TRAINING SESSION 3

Introduction

- Question and Answer section from last session
- Review of what we covered in the last session
- Discuss number formatting in Excel and how that can change the way values are displayed.

Objectives for Session 3

- Learn how to apply number formats.
- Learn how to format cells using the formatting toolbar.
- Be able to delete and insert columns and rows.
- Be able to change the width of a column and the height of a row.
- Learn how to use the undo function to undo your mistakes and the redo function in case you change your mind and want to redo the action.
- Be able to move between worksheets.
- Learn how to add, delete, and rename worksheets in a workbook.
- Learn how to get help from the office assistant and other resources in Excel.

Assignment

- We will continue to work on the sick and vacation report.

(See Worksheet on following page.)

Excel Training Class **Evaluation Form**

Please circle the response for each that best reflects your opinion. Please write any additional comments or suggestions on the back of this form.

1. The instructor covered too much material in each session.

   Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

2. The sequence of course content facilitated learning about the subject matter.

   Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

3. The Excel class was less difficult than I thought it would be.

   Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

4. The instructor was friendly and easy to talk to.
   ***Student Comment: Helped above class subject for me*.**
   Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

5. The instructor delivered the course content clearly.

   Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

6. This class met your objective for what you wanted to accomplish in this training.

   Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

7. I learned more about Excel than I expected.

   Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

8. This course increased your interest in using Excel.

   Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

9. I will use Excel in my work.

   Strongly Agree   Agree   Neutral   Disagree   Strongly Disagree

10. If you had the opportunity, would you take another class from the instructor?
    Yes   No

This work was previously published in Journal of Organizational and Ender User Computing, Vol. 16, No. 4, edited by Mahmood, pp. 37-54, copyright 2004 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Section VII
Critical Issues

This section addresses conceptual and theoretical issues related to the field of medical informatics, which include maintaining high ethical standards and ensuring that sensitive medical information remains private. Within these chapters, the reader is presented with analysis of the most current and relevant conceptual inquires within this growing field of study. Particular chapters also address mining biomedical literature, standardization of health and medical informatics, and fuzzy logic in medicine. Overall, contributions within this section ask unique, often theoretical questions related to the study of medical informatics and, more often than not, conclude that solutions are both numerous and contradictory.

Chapter 7.1
"Do No Harm": Can Healthcare Live Up to It?

Nat Natarajan
Tennessee Technological University, USA

Amanda H. Hoffmeister
Cookeville Regional Medical Center, USA

ABSTRACT

The healthcare sector is a very important one in many countries and faces numerous quality and performance problems of great significance to all citizens who live there. Although there are many performance parallels between healthcare and other sectors, there are also numerous special issues and terminology, as well. Recent publications on medical errors have put the spotlight on the issue of safety in healthcare. There are opportunities for the healthcare sector to learn from other industries where many techniques and practices for preventing errors have already proved their worth. It is important and beneficial to encourage healthcare specialists to learn from other sectors the concepts, best practices, and tools for preventing errors and improving safety. This chapter addresses the key issues and challenges relating to the management and transfer of such knowledge and places them in the context of quality and knowledge management.

INTRODUCTION

Public spending on healthcare in Germany, France, Canada, UK, US, Australia, and Japan is at least 5% of their respective Gross Domestic Products (GDPs). In the US, UK, Australia, and Canada major health reforms have been proposed (*The Economist*, 2003). The total spending (public and private) in the healthcare sector in the US accounts for roughly 14% of its national income. In the US, the federal and state governments are the largest payers for healthcare services. The last decade has witnessed increasing attention to the healthcare-related issues such as widening coverage for access to healthcare, cost containment, quality of care, and regulation. These issues have been debated in legislative, academic, and professional forums. Issues concerning safety and quality have recently been in the public limelight. Recent publications have increased public awareness of safety or lack thereof in healthcare systems. They are the Institute of Medicine's (IOM) study titled *To Err is Human: Building A Safer Health System* (Institute of Medicine, 2000) and the follow-up report that was triggered by it, *Report of the Quality Interagency Coordination* (QuIC) *Task Force to the President* (Quality Interagency Coordination Task Force, 2000). The second report of the IOM, *Crossing the Quality Chasm: A New Health System for the 21st Century* (Institute of Medicine, 2001) goes beyond safety and identifies other areas where the need for improvement is urgent. Safety is viewed as one of the dimensions of healthcare performance. Effectiveness, patient-centeredness, timeliness, efficiency, and equity are the other dimensions. One of the ten recommended principles to guide the design of health systems is that safety should be system property (Institute of Medicine, 2001). The patient safety issue is not confined to the US. In the UK, a report published in June 2000 estimated 840,000 incidents and errors occur in the National Health Service (NHS) every year (BBC News, 2001).

QUALITY MANAGEMENT IN HEALTHCARE

The traditional approach to quality management in healthcare has relied on licensure, certification and accreditation, and the use of chart review methods. Over the years, a number of organizations have been involved in the development and deployment of these structural quality assurance mechanisms. Most notable are the Joint Commission on Accreditation of Healthcare Organizations (JCAHO), Commission on Professional and Hospital Activities (CPHA), Regional Medical Programs (RMPs), Experimental Medical Care Review Organizations (EMCROs), Professional Standards Review Organizations (PSROs), and Peer Review Organizations (PROs). In one sense, these organizations form the backbone of the regulatory structure of the healthcare industry. Thus, the traditional quality management was externally driven. The positive aspect of this system is that it provides safeguards to the public in terms of standards of healthcare and minimal competence of the healthcare professionals and ensures at least minimal participation of those professionals in quality management activities (Williamson, 1991). Its shortcomings are that it uses the negative incentive of punishment in the cases of non-compliance, and the underlying premise seems to be that sanctions are needed to ensure quality. It does not promote learning. It also involves a lot of paperwork, absorbing time and effort while achieving only modest improvements.

In the last 15 years or so, there have been many instances where the healthcare organizations have used the principles that underlie the industrial quality control model (Chesanow, 1997). These are the principles of total quality management (TQM) and continuous quality improvements (CQI) that have been successfully implemented in the manufacturing sector.

Briefly, these principles can be summarized as follows (Berwick, Godfrey & Roessner, 1990a): (1) productive work is accomplished through processes; (2) sound customer-supplier relationships are absolutely necessary for sound quality management; (3) the main source of quality defects is problems in the process; (4) poor quality is costly; (5) understanding the variability of processes is a key to improving quality; (6) the modern approach to quality is thoroughly grounded in scientific and statistical thinking; (7) total employee involvement is critical; (8) quality management employs three basic interrelated activities: quality planning, quality control, and quality improvement.

There have been many case studies of applications of CQI in healthcare (Carey & Lloyd, 1995). However, the concepts and tools of CQI did not always find acceptance among healthcare administrators and providers. Skepticism such as: “How do we define and measure quality, which is a more subtle concept in healthcare?”, “Isn’t quality mainly a matter of the physician making the correct decision?”, “Where is the uniform product in medical care when every patient is different?” were expressed.

It is difficult to generalize about the effects of TQM in healthcare from isolated examples. Fortunately, the National Demonstration Project (NDP) on Quality Improvement in Healthcare provides a collection of experimental projects whose purpose was to study if the TQM model will work in the healthcare setting (Berwick, Godfrey & Roessner, 1990a). The NDP brought together 21 experts in quality management to work with a leadership team in 21 healthcare organizations represented by health maintenance organizations (HMOs), hospitals, and group practices in the US. The experts (who were from major US companies, consulting firms, and universities) were to offer their expertise in transferring quality management concepts and tools to these organizations that were willing to try them out. The participating organizations were to report on the results eight months later. Even though not all of the 21 teams completed the projects and reported back, the experiences of all of them contributed to the following important lessons learned (Berwick, Godfrey & Roessner, 1990b): (1) quality improvement tools can work in healthcare; (2) cross-functional teams are valuable in improving healthcare processes; (3) data useful for quality improvement abound in healthcare; (4) costs of poor quality are high, and savings are within reach; (5) involving doctors is difficult; (6) training needs arise early; (7) non-clinical processes draw early attention; (8) healthcare organizations may need a broader definition of quality; (9) in healthcare, as in other industries, the fate of quality improvement is first of all in the hands of leaders.

Despite these important insights that were gleaned from these projects, as acknowledged by the initiators of the project, there were two major gaps that were associated with these projects (Berwick, Godfrey & Roessner, 1990b). First, only a few project teams addressed core clinical processes such as physician decision-making, diagnostic strategies, and medical treatments. This seems to be in agreement with other research findings on the application of TQM in healthcare (Shortell, Levin, O’Brien & Hughes, 1995). Most teams worked on business and service support processes such as appointment waiting times, Medicare billing, patient discharge processes, and the hiring and training of nurses. Here the problems were similar to the quality problems in other industries. Interestingly enough, not even one of the teams measured success in terms of improved health status of the patient! Another gap was that the projects did not try to change the organizational cultures. Clearly, both these gaps have to be addressed in any effort to reduce or eliminate medical errors.

ERRORS IN HEALTHCARE

It is instructive to consider some findings about the occurrence and the impact of errors in healthcare. These stylized facts bring home in a compelling way the significance and the enormity of the tasks that lie ahead in the prevention of these errors. According to the IOM report (Institute of Medicine, 2000):

- Preventable adverse events (see the following below) are a leading cause of death in the United States. When extrapolated to the more than 33.6 million admissions to US hospitals in 1997, the results of two studies (involving large samples of hospital admissions, one in New York and the other in Colorado and Utah) imply that at least 44,000 and perhaps as many as 98,000 Americans die in hospitals each year as a result of medical errors. Deaths due to preventable adverse events exceed the deaths attributable to motor vehicle accidents (43,458), breast cancer (42,297) or AIDS (15,516).

- The rate of healthcare errors is far higher than the error rate in other industries. In one study of intensive care units (ICU), the correct action was taken 99.0% of the time, translating to 1.7 errors per day. One out of five of these errors was serious and/or potentially fatal. If performance levels even substantially better than those found in the ICU (for example, 99.9%, a 10-fold reduction in errors) were applied to the airline and banking industries, it would still equate to two dangerous landings per day at O’Hare International Airport and 32,000 checks deducted from the wrong account per hour (Leape, 1994)! In these industries, such error rates would not be tolerated.
- Errors occur not only in hospitals but in other healthcare settings, such as physicians’ offices, nursing homes, pharmacies, urgent-care centers, and care delivered in the home. Unfortunately, very little data exist on the extent of the problem outside of hospitals. The IOM report indicated, however, that many errors are likely to occur outside the hospital. For example, in a recent investigation of pharmacists, the Massachusetts State Board of Registration in Pharmacy estimated that 2.4 million prescriptions are filled improperly each year in the state (Agency for Healthcare Research and Quality, 2003b).

Errors in healthcare do not lend themselves to commonly agreed-upon definition and classification. This poses a challenge in the design and implementation of measures to prevent them. For instance, according to IOM, an **error** is defined as the failure of a planned action to be completed as intended or the use of a wrong plan to achieve an aim. However, QuIC (Quality Interagency Coordination Task Force, 2000), in order to address all the relevant issues, has expanded the IOM definition as follows: An **error** is defined as the failure of a planned action to be completed as intended or the use of a wrong plan to achieve an aim. Errors can include problems in practice, products, procedures, and systems.

Medical error is defined through adverse events. An **adverse event** is an injury caused by medical management rather than the underlying condition of the patient. An adverse event attributable to error is a “**preventable adverse event**.” **Negligent adverse events** represent a subset of preventable adverse events that satisfy legal criteria used in determining negligence, i.e., whether the care provided failed to meet the standard of care reasonably expected of an average physician qualified to take care of the patient in question (Institute of Medicine, 2000).

Medical error is an adverse event or near miss that is preventable with the current state of medical knowledge. Note that consideration of errors is broadened beyond preventable adverse events that lead to actual patient harm to include “near misses,” sometimes known as “close calls.” A **“near miss”** is an event or situation that could have resulted in an accident, injury, or illness, but did not, either by chance or through timely intervention. The broader definition of error allows learning from close calls as well as from incidents leading to actual harm (Quality Interagency Coordination Task Force, 2000). Such learning has proved to be very valuable in other industries, such as aviation and nuclear power. It is also worth noting that not all adverse errors are preventable. **Unpreventable adverse event** is defined as an adverse event resulting from a complication that cannot be prevented given the current state of knowledge.

The implication of these definitions and classifications is that a single approach to error reduction and prevention is not fruitful. Different types of errors call for different approaches.

Error Prevention in Other Industries

Error prevention strategies in other industries have been part of an overall strategy for process improvement. The important features of these strategies can be summarized as follows:

1. *They have emphasized the control of the variation in the process.* Sets of tools that help to monitor, identify, and eliminate the root causes of process variations were utilized. How the responses of management can reduce or amplify the variation in the process was understood by considering the common and special causes of variation. The important lesson was that only management could and must do something about common causes of variation. Control charts were used to monitor process stability and signal the presence of special causes of variation. It must be pointed out that in Complex Adaptive Systems (CAS) like healthcare one has to carefully distinguish between meaningful variation resulting from innovation and variation which leads to waste (Institute of Medicine, Appendix B, 2001).
2. *Processes were simplified; standards were developed, documented, and adhered to.* One of the principles of lean manufacturing is to simplify the process and make it transparent to the workers (Womack, Jones & Roos, 1990). Simple and visual signals like *kanban* cards and a system of lights are used in the pull system of Just-in-Time (Schonberger, 1986). Errors can be easily detected and prevented in a visual factory. Standardization of work methods — which implies reduction of variation — is considered an important aspect of *kaizen* or continuous improvement. Once a process is improved using tools like the Plan Do Check/Study Act (PDC/SA) cycle, it is documented for future use. Standardization encompassed all aspects of the process. For instance, a system like the 5S encompassed housekeeping and promoted safety in the work place (Shaw & Gillard, 1996). Food production and service is an industry where safety is of paramount importance. As part of its efforts to enhance the safety of food supply system in the US, the Food and Drug Administration (FDA) has mandated Hazard Analysis and Critical Control Point (HACCP) systems as a prevention standard**.** HACCP systems represent a systematic approach to the identification and control of the biological, chemical, and physical hazards that are reasonably likely to occur in a particular food in a particular production process.
3. *Inexpensive but effective solutions were used to design errors out.* Poka-yoke or error proofing is a powerful, simple, and often inexpensive technique for eliminating mistakes from occurring in the first place or to provide early warning that a mistake is about to occur. The basic ideas behind mistake-proofing were first developed and applied by Shigeo Shingo under the rubric of Zero Quality Control (Shingo, 1986). Shingo felt that defects have to be prevented by preventing the mistakes that cause them. And this has to be designed into the elements of the process, i.e., into the equipment, procedures, the operating environment, and training of the employees. Of late, the principles of mistake-proofing have also been applied to service processes. An application in healthcare is the use of indented trays in surgical procedures to prevent surgeons from leaving any particular instruments inside the patient!
   It is important to note that, in services, often the customer is an active participant in the

service creation and delivery processes. Customer mistakes can occur at the stages of: (1) preparing for the service encounter; (2) during the encounter; and (3) post-encounter resolution (Chase & Stewart, 1995). This has important implication for application in healthcare — mistake-proofing efforts must be directed not only at the equipment and employees of the providers but customers of healthcare as well. For example, if you are a patient in a hospital, consider asking all healthcare workers who have direct contact with you whether they have washed their hands (Agency for Healthcare Research and Quality, 2003a). There is great potential for the application of mistake-proofing in healthcare.

4. *Automation, information, and simulation technologies were used to reduce and prevent defects.* It is well known that computer-based automation in manufacturing has led to increased process consistency enabling much tighter tolerances to be met. For instance, robots can weld and paint more uniformly than humans. Information technology has been used in service industries to improve quality and customer service, e.g., by Ritz Carlton hotel chain (Bounds, 1996). Simulation and modeling has also proved to be very valuable in providing insights about the functioning of processes. Flight simulators are routinely used in aviation for training and education. Healthcare organizations can learn from the military as well. The healthcare profession can use computer simulation and military-based teamwork training as a means of training. Human factors engineering and ergonomics is incorporated into the design of processes and systems in industries such as nuclear power. By designing jobs, machines and operations that take into account human limitations, human errors can be prevented or reduced (Rooney, Heuvel & Lorenzo, 2002).

5. *Trained and empowered employees implemented error prevention methods. They often worked in a multi-disciplinary team structure.* There was a cultural shift. Process improvement was no longer the responsibility of management alone. Employees were now responsible for not only doing the job but also improving it. The leadership in the organization provided the support. Employees were encouraged to report on errors and devise solutions. In some of the companies that implemented Just-in-Time (JIT), the workers were given the authority to stop the line to prevent defects from being passed down stream. Once the line was stopped, problem-solving teams were formed on the spot to address the problem (Suzaki, 1987). The aviation reporting system, which the IOM and others have suggested as a model that healthcare should emulate, depends on the collection of as much information as possible about close calls as well as errors that actually resulted in harm. In the aviation industry, the identity of those who report and those who are involved in the incident are protected. This encourages people to report errors and makes the information available quickly.

6. *Organizations have benchmarked their processes to identify and close performance gaps in critical processes.* The best-in-class processes — which were not necessarily in the same industry — were benchmarked. For instance, the activities of pit crews during pit stops in the motor racing circuit have been studied to benchmark worker changes on assembly lines (Walleck, O’Halloran & Leader, 1991). Granite Rock, a building materials supplier, has benchmarked its concrete mix delivery process with Domino’s pizza delivery process (Bounds, 1996)!

7. *Process improvement strategies focused on multiple dimensions of process performance*. The Six Sigma approach that was pioneered by Motorola in the 1980s and now being utilized by many manufacturing and service companies aims at *simultaneous* reduction of defect rates, cost, and cycle time in the process (Bounds, 1996). Six Sigma has been applied to manufacturing and non-manufacturing (service and business) processes (Chassin, 1998). Linking costs to quality and cycle times is a relatively new concept in healthcare (Castañeda-Méndez, 1996).

What is Different About Healthcare?

There are some characteristics of the healthcare industry that distinguish it from other industries. Although the managerial processes in the healthcare industry are similar to those of other industries, the prevalent norms, culture, practices, and the regulatory framework can promote or hinder efforts to improve safety. These characteristics also influence the extent to which the best practices for error prevention in other industries are relevant and transferable to the healthcare sector.

1. Healthcare systems are very complex systems in terms of its constituent elements (e.g., patients, physicians and other healthcare professionals, purchasers of healthcare, payers, insurers, regulators, accrediting and licensing bodies, healthcare administrators, and professional groups), the web of relationships between them, and the knowledge, skills, and technologies that are utilized (Institute of Medicine, 2000). These groups have their own definition of errors and quality in healthcare. The interests of some of these groups are also in conflict.
2. It is difficult in healthcare systems to establish precise linkages between the inputs and the resulting outcomes (Garvin, 1990). Therefore, it is not always possible to study and model the delivery of clinical care as a "process" as it is done in manufacturing. The outcomes in terms of patients' physiological and psychological states lag — sometimes in years — the treatments they receive. The outcomes are also dependent on the cooperation and the compliance of patients themselves. In this sense, healthcare systems are quite different from other industries such as airlines, nuclear power, and food supply where safety is a critical issue. Hence, it becomes quite difficult to implement the principle that quality gurus like Dr. Deming have emphasized — quality can be improved only by improving the process. Lack of process knowledge could also be one of the factors that lead to litigation being viewed as the only recourse in the case of adverse outcomes. In fact, many legal cases in healthcare involve scrutiny of the underlying clinical processes in the courtroom.
3. In healthcare, there is an information asymmetry between the care provider and the patient. The professional judgment of the caregiver determines the nature of the clinical care that is provided. Thus, the patient often cannot judge the quality or the safety of clinical care that he or she is receiving. In this respect, healthcare may not be very different from other professional services (e.g., accounting, legal, education) where such asymmetry exists and professional judgment influences the outcomes. Nevertheless, it underscores the difficulties in developing and monitoring quality and safety standards for clinical practices. This makes the healthcare market one in which the consumers do not have the requisite information about quality, safety, and utility to make informed choices. Thus consumers cannot adequately

exert their power in the market place — as it happens in other industries — to bring about changes in healthcare providers.

4. In many healthcare institutions, there are dual lines of authority — one involving the medical staff and the other the administrative staff. This complicates decision-making concerning design and implementation of safety improvement projects. In other industries, the managerial core has control over the technical core.
5. Healthcare organizations are concerned about litigation in the context of tracking and reporting medical errors. This inhibits the relevant information from being shared within and between healthcare organizations. Legal constraints on access to and sharing of information relating to patients also prevent the dissemination of the information, which could be useful in preventing errors.
6. In other industries like the automotive, large customers routinely exert pressure on their suppliers for continuous improvements in quality and costs. This is just beginning to happen in healthcare as large group purchasers want the healthcare organizations to improve their performance, but this has not become standard industry practice so far.
7. Mechanisms and institutions for research and dissemination of research on safety in healthcare are still in a developmental stage (Institute of Medicine, 2000; Quality Interagency Coordination Task Force, 2000). The difficulties are compounded by the fact that advances in clinical knowledge, medications, and technologies in healthcare are taking place very rapidly.
8. There are powerful subcultures in healthcare organizations based on occupation and specialization, e.g., physicians, nurses, and pharmacists. Their interests and functional orientations do not facilitate a systems approach to the promotion of safety (Zabada, Rivers & Munchus, 1998).

Prospects for Improving Safety in Healthcare

In the 1990s, many healthcare facilities and health plans placed an increasing emphasis on improving healthcare quality. Since 1999, it has become a requirement for managed care plans that contract directly with Medicare (Scott, 1998). However, these programs have not sufficiently addressed the problem of medical errors.

There are a number of factors that reduce the effectiveness of existing programs to prevent medical errors (Institute of Medicine, 2000). Performance measurement and improvement programs within healthcare organizations do not directly address the problem of medical errors. Programs that have been specifically developed to prevent medical errors often operate in isolation. Efforts by external organizations to monitor errors also face limitations. For example, JCAHO has experienced significant difficulty in securing hospitals' participation in its "sentinel events" reporting system because of concerns about legal vulnerabilities or punitive actions. A number of different programs exist to detect adverse health events, although no one system is designed to detect the full scope of medical errors. Systems designed to hold organizations or individuals accountable for bad outcomes are commonly limited by underreporting of adverse events.

Clearly, the quality improvement programs within healthcare organizations could be enhanced or adapted to address errors. However, there are serious obstacles such as: (1) a lack of awareness that a problem exists; (2) a traditional medical culture of individual responsibility and blame; (3) the lack of protection from legal discovery and liability. This causes errors to be concealed; (4) the primitive state of medical information systems, which hampers efficient and timely information collection and analysis;

(5) inadequate allocation of resources for quality improvement and error prevention throughout the healthcare system. No insurer pays hospitals for safety initiatives, no matter how beneficial they are. And in most hospitals, revenues are declining as private and government insurers try to cut reimbursements; (6) inadequate knowledge about the frequency, causes, and impact of errors, as well as effective methods for error prevention; (7) a lack of understanding of systems-based approaches to error reduction (such as those used in aviation safety or manufacturing); (8) there are even greater barriers to error reduction in non-hospital settings, where the general absence of organized surveillance systems and lack of adequate personnel hinder local data collection, feedback, and improvement.

The IOM report and the action agenda developed by QuIC will go a long way in addressing these barriers. The QuIC report identified the actions that have already been taken and will be taken to reduce medical errors. They included, among others, plans of Department of Defense (DoD) to implement a new reporting system in its 500 hospitals and clinics serving approximately 8 million patients. This confidential reporting system will be modeled on the system in operation at the Department of Veterans Affairs and will be used to provide healthcare professionals and facilities with the information necessary to protect patient safety. DoD providers will inform affected patients or their families when serious medical errors occur.

The Food and Drug Administration (FDA) will develop new standards to help prevent medical errors caused by proprietary drug names that sound similar or packaging that looks similar, making it easy for healthcare providers to confuse medications. The agency will also develop new labeling standards that highlight common drug-drug interactions and dosage errors related to medications.

The Department of Veterans Affairs is considered as one of the nation’s leaders in patient safety, having instituted patient safety programs in all of its healthcare facilities serving 3.8 million patients nationwide. The Veterans Administration (VA) healthcare system will form an innovative alliance with NASA to develop a medical error reporting system similar to the system NASA has operated successfully for the Federal Aviation Administration (FAA) since 1975. Aviation errors are reported by pilots, air-traffic controllers, mechanics, and all others involved in air transportation, to the Aviation Safety Reporting System (ASRS).

In 2001, the Department of Defense (DoD), investing $64 million, began the implementation of a new computerized medical record, including an automated order entry system for pharmaceuticals, that makes all relevant clinical information on a patient available when and where it is needed. It will be phased in at all DoD facilities over three years.

The QuIC member agencies, including DoD, Veterans Administration (VA), and the Agency for Healthcare Research and Quality (AHRQ), began a collaborative project with the QuIC Task Force and the Institute for Healthcare Improvement (IHI) to reduce errors in “high hazard areas,” such as emergency rooms, operating rooms, intensive care units, and labor and delivery units.

Information Technology

The federal government has played a pivotal role in the application of information technology to healthcare. It has some of the earliest research on computerized patient records, studies evaluating the impact of computer reminder systems on laboratory testing errors, and research on the effect of computers on drug ordering. VA and DoD are recognized national leaders in the implementation of electronic medical records and decision-support tools.

QuIC can have an impact through its participants’ activities in the area of electronic records and order entry. Most healthcare providers currently work with handwritten patient notes, which are often difficult to read, not readily available, incomplete, and prone to alteration, destruction, and loss. Structured, electronic order entry systems that require complete data entry remove ambiguities that arise from incomplete information or illegible writing. Real-time decision support is a powerful technological tool for error proofing. Decision-support systems can intercept errors, such as interactions between incompatible medications and the prescription of drugs to which the patient's electronic medical record notes an allergy. Bar-coding of medications and use of robotics in dispensing medications can ensure that the appropriate medication is provided to the right patient at the right time.

Information technology can also play a very important role in preventing errors in the delivery of clinical care itself. Electronic medical records and interactive decision-support tools have the potential to allow healthcare providers timely knowledge of a patient's health history and improve clinical care. Often physicians are overwhelmed with the changes in the knowledge base and find it hard to keep up with the literature — now they can turn to information technology for help. New England Medical Center has done work in the areas of Decision Support Systems (DSS) and Expert Systems (ES) that help physicians in making accurate diagnosis in emergencies and in selecting the best course of treatments for individuals and groups. Choices are developed by ES in the form of decision trees which incorporate individual risk preferences and special medical conditions (Grossman, 1994).

STEPS IN THE RIGHT DIRECTION

- One hospital in the Department of Veterans Affairs uses hand-held, wireless computer technology and bar-coding, which has cut overall hospital medication error rates by 70% (Agency for Healthcare Research and Quality, 2003b). This system is to be implemented in all VA hospitals. VA has an exemplary patient safety program, and the DoD is developing one that is modeled after that of VA.
- The specialty of anesthesia has reduced its error rate by nearly seven-fold, from 25 to 50 per million to 5.4 per million, by using standardized guidelines, protocols and simple mistake-proofing devices, redesigning and standardizing equipment (Agency for Healthcare Research and Quality, 2003b).
- In the UK, a missed diagnosis that almost resulted in the death of a three-year-old spurred her parents to set up a charity to develop a computer system to prevent such mistakes. The online system, called ISABEL, is free to all doctors. The rapid system uses pattern recognition software to search for information in pediatric textbooks, once it is given the symptoms. It even builds in other doctors' experiences of making mistakes (News Telegraph, 2002).
- The VA/Stanford simulation center for crisis management in healthcare at Palo Alto, California, uses a simulator that can execute all the physiological functions and mimic many complex and realistic clinical crisis scenarios. Instructors can alter these scenarios. By simulating an actual patient during surgery as much as possible, doctors can get quick feedback on if they injected the wrong drug, made incorrect interpretations and other errors. In the future, such simulators are likely to be part of medical education (Institute for Healthcare Improvement, 2000).
- SSM Healthcare became the first winner of the Baldrige award in the healthcare sector (National Institute of Standards and Technology, 2003). Sister Mary Jean Ryan, the CEO of SSM for last 17 years, said she once failed to report her own error in medicating a patient, so SSM has created a "blame-free"

zone for reporting not only errors but near-misses. "Half the reported incidents lead directly to system improvements," she said (Broder, 2002).

- General Motors (GM), which has 1.25 million people in its health plans, is the largest private purchaser of healthcare in the US. It is working with partners in the healthcare industry to teach them principles of lean organization such as standardized work; workplace organization and visual controls; error proofing; employee process control; planned maintenance; and reduction of variation (Shapiro, 2000). Some hospitals have tried to be proactive in adapting such best practices from manufacturing by implementing patient-focused care teams analogous to manufacturing cells (Taylor, 1994).
- "The Medical Center of Ocean County has embraced an approach that has put the facility in Brick, New Jersey, on the cutting edge of medication system developments" (Darves, 2002). The new system involves using pharmacy technicians both on nursing units and in the emergency department and implementing Just-in-Time (JIT) delivery of medications to replace the traditional 24-hour cart system. Patient safety has improved, time savings have accrued, and nurse and patient satisfaction has increased.
- In an important development in the dissemination of comparative data, the federal centers of Medicare and Medicaid have developed a massive database, including in it every one of the 17,000 Medicare and Medicaid-certified nursing homes in every state in the US. This data is now available online (Medicare, 2003) and also on a toll-free telephone number, 1-800-MEDICARE (Medicare, 2003). For the first time, objective comparisons on several quality measures and inspection results can be made. It also provides individual nursing homes a database to benchmark, learn, and improve quality of care.

Other than the initiatives spawned by the above-mentioned federal agencies and taskforces, a number of other non-governmental programs have also been launched. For instance, the Leapfrog group was formed by the Business Roundtable in 2000. Leapfrog is a coalition of more than 135 public and private healthcare organizations and a voluntary program which has as part of its mission "… to help save lives and reduce preventable medical mistakes by mobilizing employer purchasing power to initiate breakthrough improvements in the safety of healthcare and by giving consumers information to make more informed hospital choices" (Leapfrog Group, 2003). It has focused on three patient safety practices that it believes have the potential to save lives by reducing preventable medical mistakes in hospitals. They are: Computer Physician Order Entry, Evidence-Based Hospital Referral, and Intensive Care Unit Physician Staffing. The group gathers data from hospitals on their status with regard to these practices and makes the information available to the consumers and communities.

The American Society for Quality (ASQ) is partnering with the National Patient Safety Foundation (NPSF) to offer solutions for reducing errors and increasing patient safety. Successful applications of methods such as Six Sigma will be shared in the events organized by them. ASQ is also influencing public policy in this area. It has presented a paper to the congressional health policy staff advocating the wider adoption of proven quality methods in an effort to reduce medication errors (American Society for Quality, 2002).

In the UK, where the National Health Service (NHS) is the major provider of healthcare, the National Public Safety Agency (NPSA) has been created. It is a special health authority created in July 2001 to coordinate the efforts of the entire country to report, and more importantly to learn from, adverse events occurring in NHS-funded care. The NPSA will collect and analyze incident and other patient safety information and provide timely and relevant feedback to healthcare organizations, clinicians, and other healthcare professionals, and patients (National Patient Safety Association, 2003).

While these success stories and developments contribute to the application of knowledge management systems to improve patient safety, a word of caution may be in order. The success stories may indicate what has worked in specific instances, but given the nature of the industry not too much is going to be heard about failures. The knowledge about what does not work may be lacking.

In one sense, the task of transferring best practices is made easier in the case of healthcare. In the 1980s, when the manufacturing sector in the US faced a crisis with respect to product quality and competitiveness, not many role models and forums for sharing best practices were available. The Baldrige Award was set up to address that gap. But today, there is no shortage of knowledge or the mechanisms to disseminate that knowledge. The Baldrige framework has been extended to the healthcare sector. Award winners such as SSM Healthcare can serve as role models. There are many other state and local quality award programs that provide avenues for sharing best practices. Within healthcare itself, organizations like the National Coalition for Healthcare, Institute for Healthcare Improvement, Leapfrog Group, and The National Patient Safety Foundation disseminate the best practices.

These forums and mechanisms are indeed necessary, but they are not sufficient to ensure the actual transfer of these best practices. For instance, it has been noted that the National Safety Partnership publicized a list of 16 proven and accepted best practices in medication safety. Many of them have been known for years but most are not used in a majority of hospitals. The case of the healthcare industry in Japan also reinforces this point. Despite having a manufacturing industry known for its world class quality management system, the healthcare system in Japan did not undergo a similar quality revolution (Wocher, 1997). The reasons for this are rooted in the institutional characteristics of the Japanese healthcare sector. According to Wocher, the Japanese healthcare sector lags far behind US in terms of efficiency, overall management, and the application TQM and CQI ideas (Wocher, 1997).

All this leads one to conclude that there has to be a "felt need" for industry-wide, effective prevention of errors, and improving patient safety. At the level of an individual healthcare organization, this "felt need" will be generated by a mixture of factors. They are: (1) the organizational leadership that is willing and able to overcome the internal obstacles; (2) economic incentives from the market place that affect the bottom-line; (3) external compulsions by the consumers, purchasers, payers, accreditors, and regulators; and (4) the leadership role and the role models of the federal agencies. Earlier discussion of the above issues suggests that, of late, there are encouraging signs that some traction for improving patient safety has been created by these factors. And that portends well for patient safety and knowledge management for patient safety as well.

REFERENCES

Agency for Healthcare Research and Quality. (2003a). *20 tips to help prevent medical errors.* Patient Fact Sheet. AHRQ Publication No. 00-PO38. Rockville, MD: Agency for Healthcare Research and Quality. Retrieved from the World Wide Web: http://www.ahrq.gov/consumer/20tips.htm.

Agency for Healthcare Research and Quality. (2003b). *Medical errors: The scope of the problem*. Fact sheet, AHRQ Publication No. 00-P037.

Rockville, MD. Retrieved from the World Wide Web: http://www.ahrq.gov/qual/errback.htm.

American Society for Quality. (2002, September). ASQ lends a hand at reducing healthcare errors. *Quality Progress, 35*(9), 29.

BBC News. (2001, April). NHS aims to improve patient safety. *Health*. Retrieved from the World Wide Web: http://news.bbc.co.uk/l/hi/health/1281329.stm.

Berwick, D. M., Godfrey, A. B. & Roessner, J. (1990a). Symptoms of stress in healthcare systems. *Curing Healthcare: New Strategies for Quality Improvement* (pp. 1-17). San Francisco, CA: Jossey-Bass.

Berwick, D. M., Godfrey, A. B. & Roessner, J. (1990b). Ten key lessons for quality improvement. *Curing Healthcare: New Strategies for Quality Improvement* (pp. 144-157). San Francisco, CA: Jossey-Bass.

Bounds, G. M. (ed.) (1996). *Cases in Quality*. Homewood, IL: Richard D. Irwin.

Broder, D. S. (2002, April). A beacon for better healthcare. *The Washington Post.*

Carey, R. G. & Lloyd, R. C. (1995). *Measuring Quality Improvement in Healthcare: A Guide to Process Control Applications.* New York: Quality Resources.

Castañeda-Méndez, K. (1996). *Value-based Cost Management for Healthcare: Linking Costs to Quality and Delivery*. New York: Quality Resources.

Chase, R. B. & Stewart, D. M. (1995). *Mistake-proofing: Designing Errors out.* Portland, OR: Productivity Press.

Chassin, M. (1998). Is healthcare ready for Six Sigma quality? *Milbank Quarterly,* 76(4), 565-1.

Chesanow, N. (1997, August). Making doctors' lives easier — and patients happier. *Medical Economics,* 1, 118.

Darves, B. (2002, December). Just-in-time drug delivery pays off. *The Newsletter of Quality Issues in Healthcare*. Retrieved from the World Wide Web: www.qualityindicator.com.

The Economist. (2003, March). Regime change for doctors. *The Economist Global Agenda*. Retrieved from the World Wide Web: http://www.economist.com/agenda/displayStory.cfm?story_id=1626257.

Food and Drug Administration. (2001). Retrieved from the World Wide Web: http://www.fda.gov/ola/2001/foodsafety1010.html.

Garvin, D. A. (1990). Afterword: Reflections on the future. In *Curing Healthcare: New Strategies for Quality Improvement*, (pp. 159-165). San Francisco, CA: Jossey-Bass.

Grossman, J. H. (1994). Plugged-in medicine. *Technology Review,* 97(1), 22-29.

Hoerl, R. W. (1998, June). Six Sigma and the future of the quality profession. *Quality Progress,* 1(6), 35-42.

Institute for Healthcare Improvement. (2000, February). Reducing medical errors and improving patient safety: Success stories from the front lines of medicine. *Accelerating Change Today.* Retrieved from the World Wide Web: www.ihi.org.

Institute of Medicine. (2000). *To Err is Human: Building a Safer Healthcare System*. L. Kohn, J. Corrigan & M. Donaldson (eds.). Washington, DC: National Academy Press. The on-line versions of the report can be found at the website of National Academy Press. Retrieved from the World Wide Web: www.nap.edu.

Institute of Medicine. (2001). *Crossing the quality chasm: A new health system for the 21st century.*

Washington, DC: National Academy Press. The on-line version of the reports can be found at the website of National Academy Press. Retrieved from the World Wide Web: www.nap.edu.

Leape, L. L. (1994). Error in medicine. *Journal of the American Medical Association*, 272, 1851-57.

Leapfrog Group. (2003). Retrieved from the World Wide Web: http://www.leapfrog group.org/about.htm.

Medicare. (2003). Retrieved from the World Wide Web: www.medicare.gov.

National Institute of Standards and Technology. (2003). Baldrige National Quality Program. *SSM Healthcare*. Retrieved from the World Wide Web: www. quality.nist.gov.

National Patient Safety Association. (2003) Retrieved from the World Wide Web: http://www.npsa.nhs.uk.

News Telegraph. (2002, June). *Parents who did something positive*. Retrieved from the World Wide Web: www.news.telegraph.co.uk.

Quality Interagency Coordination Task Force (QuIC) Report to the President. (2000, February). Doing what counts for patient safety: federal actions to reduce medical errors and their impact. *Retrieved from the World Wide Web:* http://www.quic.gov/report/.

Rooney, J. J., Heuvel, L. V., & Lorenzo, D. K. (2002, September). Reduce human error. *Quality Progress*, 35(9), 27-36.

Schonberger, R. (1986). *World Class Manufacturing: Lessons of Simplicity Applied*. New York: Free Press.

Scott, J. S. (1998, October). Questing for quality: QISMC-ly cutting the QAPI. *Healthcare Financial Management*, 52(10), 32.

Shapiro, J. P. (2000, July). Taking the mistakes out of medicine. *U. S. News & World Report*, 129(i3), 50.

Shaw, A. & Gillard, J. (1996). Partnering the Honda way. In G. Bounds (Ed.), *Cases in Quality*, (pp. 26-56). Homewood, IL: Richard D. Irwin.

Shingo, S. (1986). *Zero Quality Control: Source Inspections and the Poka-Yoke System*. Portland, OR: Productivity Press.

Shortell, S., Levin, D., O'Brien, J. & Hughes, E. (1995). Assessing the evidence on CQI: Is the glass half empty or half full? *Journal of the Foundation of the American College of Healthcare Executives*, 40, 4-24.

Suzaki, K. (1987). *The New Manufacturing Challenge: Technique for Continuous Improvement*. New York: Free Press.

Taylor, S. (1994, January). The medical center takes proactive approach to healthcare reform. *Industrial Engineering*, *26*(1), 20-23.

Walleck, S., O'Halloran, D. & Leader, C. (1991). Benchmarking world-class performance. *McKinsey Quarterly*, 1, 7.

Williamson, J. W. (1991). Medical quality management system in perspectives. Chapter 2 in J. B. Crouch (Ed.), *Healthcare Quality Management for the 21st Century*. The American College of Physician Executives, Tampa, FL: American College of Medical Quality.

Wocher, J. C. (1997). TQM/CQI efforts in Japanese hospitals—why not? In V. A. Kazandijian (Ed.), *Effectiveness of CQI in Healthcare: Stories from a Global Perspective*, (pp. 49-108). Milwaukee, WI: Quality Press.

Womack, J. P., Jones, D. & Roos, D. (1990). *The Machine that Changed the World*. New York: R. A. Rawson.

Zabada, C., Rivers, P. A. & Munchus, G. (February 1998). Obstacles to the application of total quality management in healthcare organizations. *Total Quality Management*, *9*(1), 57-59.

This work was previously published in Creating Knowledge-Based Healthcare Organizations, edited by N. Wickramasinghe, J.N.D. Gupta, and S. Sharma, pp. 286-303, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.2
HIPAA:
Privacy and Security in Health Care Networks

Pooja Deshmukh
Washington State University, USA

David Croasdell
University of Nevada, Reno, USA

ABSTRACT

This chapter explores privacy and security issues in health care. It describes the difference between privacy and security in the context of health care, identifies sources of concern for individuals who use information technologies for health-related purposes, and presents technology-based solutions for privacy and security in health care networks. The purpose of the chapter is to provide an investigation of the sources of concern for regulations and technologies in the health care industry. The discussion is based on the Health Insurance Portability and Accountability Act (HIPAA) and its eight guiding principles. The chapter explores the implications of legal and regulatory environments driving HIPAA regulations, the need for privacy and security in health care networks, and information technologies used in the health care industry. Related ethical issues, current technologies for providing secure solutions that comply with the regulations, and products emerging in the market are also examined.

INTRODUCTION

Data communication infrastructures are changing how health information and health care is provided and received. People using tools such *as the Internet for health-related purposes — patients, health care professionals, administrators, and researchers, those creating or selling health products and services, and other stakeholders — must join together to create a safe environment and enhance the value of the Internet for meeting health care needs. Because health information, products, and services have the potential to both improve health and do harm, organizations and individuals that provide health information via the Internet have obligations to be trustworthy, provide high quality content, protect users' privacy, and adhere to standards of best practices for services in health care. People using telecommunications infrastructures in health care share a responsibility to help assure the value and integrity of the information by exercising judgment in using health care sites, products, and services.* Internet Health Coalition (2000)

The Health Insurance Portability and Accountability Act (HIPAA) has brought about significant changes in the procedures and practices within the health care industry. As newer information technologies are implemented in health care organizations, the challenge becomes to increase network connectivity and enable access to key information without compromising its confidentiality, integrity, or availability. With the advent of HIPAA regulations, health care organizations are required by law to have procedures in place to protect the privacy of patient information. This chapter addresses issues related to privacy and security of patient information in health care networks. It provides a background on HIPAA regulations, drivers for the need for privacy and security in health care organizations, the role of technology-based solutions, and the products available to the industry. The chapter includes a discussion of the ethical issues driving the design and implementation of information in support of HIPAA guidelines. The increased use of information technology in health care promises greater functionality and decreasing costs. While these factors point towards continued development of more robust applications, careful selection and implementation is necessary to ensure the security and privacy of patient information.

The evolution of networking technologies has enabled businesses to provide enhanced services, greater access to information, and higher levels of availability for both the service providers and the customers. While many industries have easily adopted internetworking technologies, others have been unable to do so because of the inherent complexities of their specific businesses. The health care industry is a prime example. Health care is a document-intensive industry that has faced significant challenges in migrating to the near "paperless" environments that many industries strive to achieve utilizing networking technologies (Cisco Systems, 2002). Furthermore, health care organizations work with highly sensitive data such as patients' personal health information. As such, health care organizations must be keenly aware of the privacy concerns and security risks of converting to electronic infrastructures.

BACKGROUND

Anyone seeking health-related information, products, or services has a right to expect that organizations and individuals who provide such information follow a set of guiding principles. If confidences are not kept, individuals will be less forthcoming with information, which in turn may impact the care they receive. Health information includes information for staying in good physical condition as well as for preventing and managing disease. It may also include information for making decisions about health products and health services and may be in the form of data, text, audio, and/or video. In addition, enhanced health information may be available through programming and interactivity (Internet Health Coalition, 2000). Managing health information in a technological world implies the persistent storage and potential dissemination of health-related data using data communication networks. Such environments have the potential to compromise both the security and privacy of the records maintained on these networks.

Although privacy and security are terms that are often used interchangeably, there is an inherent difference between the two concepts. The Merriam-Webster online dictionary (2003) defines *security* as:

The quality or state of being secure: as ***a****: freedom from danger:* ***SAFETY b****: freedom from fear or anxiety* ***PROTECTION*** *(1): measures taken to guard against espionage or sabotage, crime, attack, or escape.*

Whereas *privacy* is defined as:

The quality or state of being apart from company or observation: SECLUSION b: freedom from unauthorized intrusion <one's right to privacy>.

By combining the two terms, one can gain a complete picture of protecting information in context of the health care industry. Both security and privacy must be considered in order to adequately address the issue of safeguarding patient information (Fleisher, 2001). The Health Insurance Portability and Accountability Act (HIPAA), effective as of April 14, 2000, attempts to address privacy and security issues in the context of health-related activities. Accordingly, HIPAA defines *security* as:

The regulations which address the protection of data resident on provider computers or networks, as well as the protection of data while it is being transmitted to third parties. Primarily, security addresses the technical components related to the collection, protection, and dissemination of data. (HIPAA Standard164.530)

Whereas *privacy* is defined as:

The regulations, which address the protection of patient information in any format and by any user. Privacy necessitates providing an individual's health related information and disclosure of how and where that information is being used. (HIPAA standard 142.308) (Fleisher, 2001)

More simply stated, the security regulations address technical components of health care information, which are monitored by a Security Officer. Privacy regulations are more operational in nature and are managed by a Privacy Officer (Wilson, 2002).

Organizations are becoming increasingly dependent on data communication networks for their daily business communications, database information retrieval, and distributed data processing. The rise of the Internet and wireless communications has provided opportunities to connect to computers anywhere in the world. However, this capability has also increased the potential vulnerability of organizational assets. As a result, not only do organizations need to prevent their assets from threats such as fraud and theft but they also need to be concerned with the potential loss of consumer confidence from a publicly known security break-in. Other security concerns include potential losses from disruption of an application or natural disasters (Fitzgerald & Dennis, 2000).

SECURITY ISSUES IN THE HEALTH CARE INDUSTRY

In today's uncertain political and economic environment, many factors are driving the need for secure networks in the health care industry. A significant business challenge is to increase network connectivity and to enable access to key information assets without compromising the confidentiality, integrity, or availability of those assets. Some believe that using new technologies can assist with the process of securing patient information. According to Oracle Corporation, incorporating technology in hospitals will make the process of providing access to patient information more secure since all the data will be stored in one place (Couzin, 2001). This will reduce the risk associated with patient information "floating" around the hospital where it could potentially be accessed by multiple doctors, nurses, technicians, and administrators. In some of the solutions implemented by Oracle, doctors no longer need to spend valuable time filling out forms, tracking patient charts, or waiting to pick up X-ray or MRI results as these tasks are now automated through the use of technology (Couzin, 2001). While this has greatly increased the ease of use, it has also increased the vulnerability and security of the underlying content.

Conversely, the health care industry's steady move towards the computer-based patient record and the overall trend of delivering health care information using information technology in lieu of paper has raised anxiety regarding the security of that information. Issues related to the security of an enterprise network could create fear and uncertainty for health care executives. Securely sharing patient information over distributed regional networks that link multiple hospitals, clinics, and doctors' offices has become an issue of key importance in health care organizations (KBeta Security Web, 2001). For an organization to begin building a highly secure network, it must first understand the issue of security and why it has become such a priority in health care today.

To this end, network security has never been more important. For example, e-vandalism is occurring unnoticed in many of today's information-centric companies. Clandestine hackers and vandals not only steal a company's confidential information, but also damage its reputation in the process. The loyalty that companies have worked so hard to build could disappear if customers and business partners believe that their personal data is at risk (McMillan, 2002). The concern is thus magnified for health care enterprises trying to deploy network solutions.

Another factor driving the need for greater attention to security concerns is the move to sharing information with remote physicians. With the growth of outpatient care sites health care providers can no longer be content with just one local area network. In the past, doctors did not consider having the capacity to receive patient data in their homes. The ability of physicians to practice medicine remotely is becoming a competitive differentiation in the marketplace (Sarasohn-Kane, 2003). This practice is beneficial not only to the doctor in terms of flexibility and convenience, but also to patients who could be diagnosed and receive treatment virtually across time and physical space. Health systems and plans want to attract and retain the best physicians in the community by touting the ability to provide information anywhere and anytime a doctor needs it. The confidentiality of such information becomes paramount in the ability to conduct business.

These issues have also gained impetus as network intrusions by computer viruses have become more prevalent in recent years. The threat of new and more virulent computer worms and viruses has heightened the level of consumer awareness and concern about the use or misuse of personal health records. An IBM survey conducted in 1999 revealed that 33% of Americans would trust banks to handle their personal information properly, but only 23% placed the same faith in health care providers (Sarasohn-Kane, 2003). In a September 2000 Gallup poll, 77% of the respondents said the privacy of their personal health information was very important. Eighty-four percent were concerned that this information might be made available to others without their consent. Only 7% of the respondents said they were willing to store or transmit personal health information on the Internet and only 8% felt a Web site could be trusted with such information (Hunt, 1999).

Confidence in security measures is even more important given the push towards consumer-driven health care. Consumers' desire to access information about their own health has resulted in increased use of interactive medical networks such as the Internet and intranets. In part, the changes are due to the improving technological capability along with managed care organizations' desire for patient empowerment by asking patients to assume more self-management. Consumers' use of these networks has initiated the development of personal health information management software that may be a precursor to an electronic medical record owned by the patient. As a result, health care information is becoming more portable. Previously, the hospital literally owned patient records. In contrast today, multiple owners outside the hospital may be contending for it. The implementation of technical solutions has provided greater portability and convenience in the health care arena. At the same time, the demand for security is becoming even more imperative (Sarasohn-Kane, 2003). Copies of electronic files containing vast amounts of confidential information can easily be sent electronically over a network connection without any indication that the information was stolen.

Consumers want to control where their health information goes. Part of the enterprise's risk management analysis is to realize that we live in a litigious society. As such, it is prudent on the part of businesses to take every reasonable step to ensure the confidentiality of health care information. More trust in the health care system will help ensure better health outcomes through the use of technology.

PRIVACY ISSUES IN THE HEALTH CARE INDUSTRY

Health care providers maintain and share a vast amount of sensitive patient information for a variety of reasons. Such records are kept and shared for diagnosis and treatment of the patients, payment of health care services rendered, public health reporting, research, and even for marketing and use by media. While traditional paper-based systems have vulnerabilities, they also place some natural limits on the ability of information collectors to share and disseminate information. It is sometimes a challenge to locate paper records. In order to disseminate the information, someone must physically remove the records from the premises either by carrying, copying, mailing, or faxing the documents. These limitations create a double-edged sword. Although such systems may protect information from being disseminated for improper reasons, they may also obstruct the flow of information being shared for legitimate, health care-related purposes (Choy, Hudson, Pritts, & Goldman, 2001).

Health information can be easily located, collected, and organized with the migration of the health care industry toward electronic data collection, storage, and transmission. One major drawback is that sensitive and personal patient information can be sent to any number of places thousands of miles away with the click of a mouse button. Thus, some consumers may be afraid to take advantage of the technology because of privacy and confidentiality concerns. According to the Ethics Survey of Consumer Attitudes conducted by the Cyber Dialogue and the Institute for the Future for the California Health Care Foundation and the Internet Health Care Coalition in January 2000, more than 75% of the people surveyed are concerned about Web sites sharing information without their consent, thus impacting their willingness to use the Internet for health- related services (Goldman & Hudson, 1999).

Consumers are increasingly worried about the loss of their privacy, and have heightened concerns when it comes to their health information (Brewin, 2003). They worry that their health information may be used or disclosed inappropriately and leave them vulnerable to unwanted exposure, stigma, and discrimination, possibly leading to economic losses. Patients fear that their personal information will be used to deny them health insurance, employment, credit, and housing. With the increase in the use of technology and the ease with which information can be transmitted, there has undoubtedly been a considerable increase in the access of health care information. People who access such data without appropriate authorization are motivated either by profit or at times just plain curiosity (Goldman & Hudson, 1999). As a result, consumers sometimes take drastic steps to keep their health information private. According to one survey, almost one out of six U.S. adults has taken extreme steps to maintain the privacy of his or her medical information. Patients withhold information from their doctors, provide inaccurate or incomplete information, and doctor-hop to avoid a consolidated medical record. They go as far as paying out-of-pocket for care that is covered by their insurance, or even avoiding care altogether (Goldman & Hudson, 1999).

Such privacy-protection behavior, which consumers/patients do both offline and online, can result in a significant cost to their health. A study released by the Pew Internet and American Life Project found that 89% of Internet users who seek health information online are worried that others will find out about their activities and are worried that the Internet companies will give this information away. Eighty-five percent fear that insurance companies might change their coverage after finding out what online information consumers had accessed (Choy et al., 2001). By concealing information, patients risk undetected and untreated conditions. At the same time, the doctor's ability to diagnose and treat patients is jeopardized without access to complete and accurate information. Further, future treatment may be compromised if the doctor misrepresents patient information so as to encourage disclosure. This in turn can have a detrimental effect on the community, as without full patient participation upfront, the information collected will be unreliable for users downstream. Ultimately, health care initiatives that depend on complete and accurate information may be undermined (Goldman & Hudson, 1999).

LEGAL AND REGULATORY ENVIRONMENT

Regulatory factors are driving the current trend toward security and privacy standards in the transmission of health care information over enterprise networks. State and federal legislation, professional and standards organizations, and internal organizational risk management departments are driving the need for security measures. Many states, for example, regulate the use of electronic signatures and medical records. The Joint Commission on Accreditation of Healthcare Organizations (JCAHO) addresses security and confidentiality issues in the Information Management section of its accreditation manual. Overlaying all these factors, however, is the greatest of all regulatory drivers: a recent federal law called HIPAA.

The Health Insurance Portability and Accountability Act, signed into law in the United States in August 1996, mandates the adoption of national uniform standards for the electronic transmission of health and patient information. The intent of HIPAA is "administrative simplification" and protection of patient privacy. HIPAA requires that the health care industry promote a national, uniform security standard for the secure electronic transmission of patient-identifiable information.

HIPAA is a turning point for the health care industry because it requires that the industry develop a set of national standards that will help bring the much-needed data-standard unity to health care transactions and provide assurance that confidential patient information will be safe or safer than paper-based patient records.

Although HIPAA does not mandate the collection or electronic transmission of health information, it requires that standards be adopted for any electronic transmission of specified transactions. To ensure protection of privacy, the law provides for confidentiality protections for information processed in accordance with the new standards. It requires organizations to focus on *Electronic Data Interchange (EDI) transactions for health plan enrollment, eligibility, claims payment, premium payment, coordination of benefits, and referral/authorization*. HIPAA also mandates p*rotecting confidentiality of individually identifiable patient information in an automated system*. It requires organizations to be able to demonstrate sound practices that protect patient confidentiality and security.

HIPAA security requirements are broad, covering any organization that generates or otherwise handles electronic patient records and other e-medical data. HIPAA requires the health care organizations to implement encryption, user authentication, and other security measures to safeguard the integrity, confidentiality, and availability of electronic data by mid-2003. Entities affected by this law include virtually all private and government hospitals, outpatient centers, nursing centers, Health Maintenance Organizations (HMO), Preferred Provider Organizations (PPO), insurance companies, firms providing clinical information systems for medical labs, providers of pathology, radiology, patient billing, and pharmacy records, medical software application providers, and even related Web portal companies (Yozons Technology, 2003).

Penalties for noncompliance to the law can be severe. The civil penalty for violating transaction standards is up to $100 per person per violation and up to $25,000 per person per violation of a single standard for a calendar year. The penalty for knowing misuse of individually identifiable health information can reach $250,000 and/or imprisonment for up to ten years. HIPAA has hit the nation's $1.3 trillion health care industry quickly by becoming the de facto security guideline for federal privacy standards regarding health care information. The privacy standards, which govern electronically Protected Health Information (PHI), went into effect as of April 14, 2003 and could create a legal nightmare for the health care industry, requiring a massive training effort and costing millions of dollars. There is also concern that litigation over a failure to adhere to HIPAA security standards may dampen the use of technologies such as wireless LAN systems in hospitals (Brewin, 2003).

Given the mandated HIPAA compliance, many organizations have been working overtime to ensure their organizations are aligned correctly. In order to examine the implications of HIPAA in the workplace, interviews were conducted with professionals responsible for information systems and telecommunication services in regional medical centers. The combination of HIPAA and the hospitals' endeavor to become HIPAA compliant has resulted in additional responsibilities for individuals such as the Privacy and Security Officers for their respective medical centers. On the whole, these professionals consider HIPAA to be a double-edged sword. In their opinion, to a large degree most hospitals and health care organizations have always been very sensitive to privacy and confidentiality of patient information. HIPAA has simply formalized some of those issues and ensured that the standards are being applied equally. The primary gap in the protection of patient information was seen at pharmacies as they previously shared patient information with various vendors. If a patient were to buy prescription drugs at the pharmacy, he or she might receive advertisements with information on drugs related to their condition. With the advent of HIPAA, pharmacies need to be monitored and are not allowed to share patient information with vendors without the consent of the patient. In addition, the electronic transaction of transmitting patient information to bill insurance companies needs to be supported by software that is HIPAA compliant.

IMPLICATIONS OF HIPAA IMPLEMENTATION

Overall HIPAA is believed to be a very positive thing, as patients find comfort in knowing that there are standards in place to safeguard their personal health information. Still, some patients have negative perceptions regarding how HIPAA affects the privacy and security of personal information. Most hospitals are very explicit when stating what happens with patient information and with whom such information is shared. However, such an expression, alluding to a more open process, has raised concerns among some patients who have only recently become aware of such standard practices of health care organizations. They believe that HIPAA has allowed health care organizations to share more information than they previously could. In effect, HIPAA has heightened peoples' awareness of issues related to information privacy and security. For many, these concern are issues that were previously unknown or of little interest. As such, efforts are being made to educate such individuals in order to make them feel more comfortable about the privacy of their information.

The security piece of the regulation mandates the implementation of security measures within health care organizations. These security measures were previously hard to implement because there was not much return on investment. With HIPAA, it has become easier to justify these requirements from a business standpoint. The security measures may contain a layered approach to securing the network. A plan needs to be in place to ensure that every single layer of the network has been "hardened" to make it secure. In addition, some medical centers have assigned different access capacities to their various staff members depending either on their location in the hospital or the privileges assigned to them. Further constraints have been implemented in order to manage and protect the IT resources. For instance, users may need to follow a specific format for their passwords in order to ensure that they are not easily decipherable. For systems containing clinical information used to make emergency decisions, some hospitals have implemented a "break the glass" procedure, in which, for example, if a password is not working in emergency situations, physicians are still able to get to the information. Extra audits and alerts are put in place so that if someone "breaks the glass," network administrators are automatically notified and upper management can be apprised of the related circumstances.

The processes associated with accessing records have also undergone changes as illustrated in the following scenario. If a physician examines one patient and consults with another physician on the case, the second physician may be unable to access the patient's records since he or she is not the "physician of record." In such a situation, the second physician could override the access blocks by agreeing to have their name appear in the audit report. If the second physician agrees, administrators can monitor records for inappropriate activities and follow-up with the physician to address access issues as needed.

There are obvious cost implications on the implementation of HIPAA. Organizations now need to maintain a fair-sized HIPAA contingency budget every year. There are costs such as traveling to understand more about HIPAA, man-hours, and employee education.

SOLUTIONS

Health care services and professionals are working toward providing environments that safeguard health-related information. In part, actions have been encouraged and set in motion as a result of HIPAA compliance efforts. Some solutions are enabled by information and communication technologies while others rely on standards of practice that have been advanced by the health care community. The eHealth code of Ethics, initiated in 2000, helps ensure that people understand the potential risks of managing their own health care and the health of those in their care. The eight guiding principles of the code work to ensure candor, honesty, quality, professionalism, responsible partnering, accountability, informed consent, and privacy of patient information (Internet Health Coalition, 2000) (Table 1).

In a similar vein, the Health on the Net (HON) Foundation (2002) Code of Conduct for medical and health Web sites addresses the reliability and credibility of information on the Internet. Specifically, it addresses the authority of the information provided, data confidentiality and privacy, proper attribution of sources, transparency of financial sponsorship, and the importance of clearly separating advertising from editorial content (Health on the Net Foundation, 2002).

ROLE OF TECHNOLOGY-BASED SOLUTIONS

In addition to guiding principles for behavior, health care enterprises are faced with a number of technical and operational challenges. Among these are the needs to operate more efficiently, to expand the scope of the enterprise, to provide greater access to information from a variety of locations and platforms including mobile/wireless, and to greatly improve the security and privacy of information. These challenges can, at times, seem contradictory. The responses to these challenges necessitate many different initiatives, including security planning, creation of wide area networks, and adoption of wireless/mobile platforms. In addition, the health care industry's ongoing, massive consolidation has resulted in the emergence

Table 1. Eight guiding principles of eHealth code of ethics (Internet Health Coalition, 2000)

e-Health Code of Ethics: 8 Guiding Principles	
Candor	▪ Disclose vested financial interests ▪ Disclose key information for consumer decisions
Honesty	▪ Present information truthfully ▪ No misleading claims
Quality	▪ Accurate, clear, current, evidence-based ▪ Readable, culturally competent, accessible ▪ Citations, links, editorial board and policies
Informed Consent	▪ Privacy policy and risks ▪ Data collection and sharing ▪ Consequences of refusal to consent
Privacy	▪ Prevent unauthorized access or personal identification of aggregate data ▪ Let users review and update personal data
Professionalism	▪ Abide by professional codes of ethics ▪ Disclose potential conflicts of interest ▪ Obey applicable laws and regulations ▪ Point out limits of online practice
Responsible partnering	▪ Choose trustworthy partners, affiliates, and links ▪ Maintain editorial independence from sponsors ▪ Tell users when they are leaving the site
Accountability	▪ Provide management contact info ▪ Encourage user feedback ▪ Respond promptly and fairly to complaint

of so-called Integrated Delivery Systems (IDS). These systems are designed for large, regional providers that need to share clinical and other information among numerous hospitals, clinics, home-care agencies, and other facilities. With the advent of multiple clinics and hospitals sharing data, a health care organization must contend with factors such as leased telecommunications lines and external circuits and services, rather than local services inside a building (KBeta Security Web, 2001).

From unauthorized users to disgruntled employees to cyber-terrorists, the threat to health care information systems cannot be taken lightly. Poorly written software, the demand for convenience over security, and overworked, undertrained IT health care professionals present substantial information systems vulnerabilities (Beaver, 2003). The HIPAA security rule is about information security best practices. Technology for secure networks includes tools such as firewalls, encryption, user authentication, and software that detects and reports network vulnerabilities and unauthorized activity (KBeta Security Web, 2001).

MANAGEMENT ISSUES

It is imperative to realize that technology is not the silver bullet that people have come to believe in and rely on. Simply relying on the technical solutions is not sufficient to ensure the security of the information. While cutting-edge network technology might be available to make networks secure, technology is only an enabler (KBeta Security Web, 2001). Other issues that need to be considered include ongoing information risk assessment, information security audits, disaster recovery, and business continuity plans (Beaver, 2003). Information security is a business issue as well. Organizational and cultural issues are paramount in making the technology fulfill its potential. The key is to impart a *culture of security and confidentiality* to an organization. As a corporate cultural issue, security and confidentiality integrate through diverse areas of technology, organization, and regulation. Security is an integrated approach in which an organization needs to have its entire management team involved in the decision-making processes. These processes should include key decision makers from multiple and varied departments such as legal, human resources, and operations.

Given the intersecting technical, organizational, and regulatory factors, it should be understood that security of enterprise networks is both an ethical as well as a cultural issue, requiring constant, iterative education and awareness. In order for information security initiatives to be effective, it is critical to not forget the end user. In fact, the human factor can be the weakest link in an information security program (Beaver, 2003). Organizations must reinforce employee awareness through an ongoing program of education, reward, and recognition. Individual user accountability is a critical component of network security. A system cannot allow, for example, several providers to use the same terminal simply by using the same password and logon for the sake of convenience.

Mobile computing applications and the use of wireless technologies in health care has seen a great deal of growth and expansion capabilities. From a security standpoint, the social changes are probably the hardest to implement. It is easy to have the firewalls and technologies in place, but it is harder as well as more important to manage the social aspect of such a change. If people leave their PC screen on and the others walking by can see the information displayed on it, then security doesn’t mean anything. Similarly, if people stick their passwords on post-it notes on their screens, then the security measures become meaningless. Maybe security will change with biotechnology, and people won’t have to remember multiple passwords while they simply remember to bring their thumb to work.

The information owners must determine security risks, impact, and the severity of a potential compromise. Additionally, the information owners should be responsible for determining a balance between the costs and benefits of security for their particular organization. Organizational risk is an aggregate factor and must be determined collectively by all of the information owners within and throughout the organization. Securing an organization's information assets is ultimately an upper management responsibility and must be managed from the top down from a business perspective. Health care managers must understand the business impact of information risks and the implications involved if systems are not secured. To protect themselves from legal liabilities, health care organizations need to show due diligence in attempting to implement best practices in this regard.

IMPLICATIONS AND CONCLUSIONS

In the end, it is necessary to understand that there is no such thing as 100% security. However, it is vital that reasonable measures be in place to reduce the chances of unauthorized access of confidential information. With HIPAA privacy regulation compliance mandatory since April 14, 2003, health care providers need to ensure that their systems meet the federal health privacy policies. Although the law allows for incidental disclosures of information, providers covered by the rule are expected to put reasonable safeguards in place to protect their patients' information. This means that sign-in sheets may be used in a doctor's reception areas but patients should no longer be asked to write down their conditions because other patients could see the sheet. In an emergency room, the large white boards listing patient names and conditions should be moved to areas out of public view. In hospitals, patient charts should be turned to face the wall so people walking by cannot read them. New computer software allows doctors' offices to identify patients by full name or just by initials, just in case others might catch a glance of a PC screen. Most hospitals have new policies regarding the release of information regarding a patient's condition. Such information was once routinely provided to family, friends, clergy, and reporters who called. Under the new rules, hospitals must give patients a chance to opt out of any hospital directory. No information, even that a person is a patient in the hospital, may be released without the patient's consent. Even if a patient agrees to being included on a general patient listing, hospitals may release only limited information without specific patient authorization and only if a caller asks about a patient by name (Meckler, 2003).

Technologically, the continued growth and acceptance of the Internet, widespread growth of wireless devices with greater functionality and decreasing costs of technology solutions all point towards continued development of more robust software applications. These developments may improve adoption of technology, but careful selection and implementation are necessary to ensure the security and privacy of patient information. Eventually, organizational policies, technical solutions, and regulatory guidance should improve the acceptance of e-technology and increase its value to the health care organizations. With greater security of patient information, health care organizations can build patient trust by protecting confidential patient information. This trust between the patient and the provider in turn will lead to improvement in the overall quality of health care.

REFERENCES

Beaver, K. (2003). Information security issues that healthcare management must understand.

Journal of Healthcare Information Management, April, *17*(1), 46-49.

Brewin, R. (2003,). New HIPAA security rules could open door to litigation. *Computerworld*, February 20. Retrieved April 23, 2003 from: *http://www.computerworld.com/governmenttopics/government/policy/story/0,10801,78684,00.html*

Choy, A., Hudson, Z., Pritts, J., & Goldman, J. (2001). Exposed online: Why the new federal health privacy regulation doesn't offer much protection to Internet users. Report of the Pew Internet & American Life Project. Health Privacy Project, Institute for Health Care Research and Policy, Georgetown University, Washington, DC. Retrieved April 30, 2003 from: *http://www.healthprivacy.org/usr_doc/PIP_HPP_HealthPriv_report.pdf*

Cisco Systems (2002). Network security solutions for health care: Making HIPAA SAFE. August 21. Retrieved May 17, 2003 from: *http://www.cisco.com/warp/public/cc/so/neso/sqso/hipaa_wp.htm*

Couzin, J. (2001, March 26). A Hospital for the Digital Age. *IDG.net.* March 26. Retrieved April 10, 2003 from: *http://www.idg.net/go.cgi?id=477347*

Fitzgerald, J,. & Dennis, A. (2000). *Business data communication and networking* (6th ed.). New York: John Wiley & Sons.

Fleisher, S.M. (2001). Health care data security: HIPAA draft security rules and California law. *California Health Care Symposium 2001.* May 10. Retrieved April 09, 2003 from: *http://www.ehcca.com/presentations/casymposium/fleisher.pdf*

Goldman, J., & Hudson, Z. (1999). Health privacy 101 exposed: A health privacy primer for consumers. Retrieved April 23, 2003 from: *http://www.healthprivacy.org/info-url_nocat2302/ info-url_nocat.htm*

Health On the Net (HON) Foundation (2002). HON code of conduct (HONcode) for medical and health Web sites. Retrieved January 15, 2004 from: *http://www.hon.ch/HONcode/*

Hunt, K. (1999). Confidentiality of medical records: National survey . The Princeton Survey Research Associates for the California HealthCare Foundation. Retrieved April 23, 2003 from: *http://www.chcf.org/press/view.cfm?itemID=12267 - 41k*

Internet Health Coalition (2000). eHealth code of conduct. Retrieved January 15, 2004 from: *http://www.ihealthcoalition.org/ethics/ehcode.html*

KBeta Security Web (2001). White Paper: Security and Health Care Enterprise Networks: Balancing Technology and Culture. Retrieved April 25, 2003 from: *http://www.kbeta.com/SecurityTips/hipaa/CiscoHipaa WhitePaperori.htm*

McMillan, M.H. (2002). Cognitive security management: Meeting the challenges of eHealth, *2002* Annual HIMSS Conference and Exhibition: Session 48.

Meckler, L. (2003). New patient privacy rules to take effect: Health data handling facing adjustments. Associated Press. Retrieved June 22, 2003 from: *http://www.boston.com/dailyglobe2/104/nation/*

Merriam-Webster Online (2003). The language center. Retrieved April 19, 2003 from: *http://www.m-w.com/cgi-bin/dictionary*

Sarasohn-Kahn, J. (2003). Public perceptions and health care privacy. *CSO Magazine* [online]. Retrieved April 28, 2003 from: *http://www.cio.com/sponsors/041502cisco/public.html*

Wilson, M., (2002). Security and privacy— Two sides of the same coin. *HIMSS Proceedings: Educational Sessions.*

Yozons Technology (2003). Healthcare, biotech and pharmaceuticals. Retrieved April 23, 2003 from: *http://www.yozons.com/pub/uses/medical.jsp*

This work was previously published in Information Ethics: Privacy and Intellectual Property, edited by L. Freeman & A.G. Pease, pp. 219-238, copyright 2005 by Information Science Publishing (an imprint of IGI Global).

Chapter 7.3
Medical Ethical and Policy Issues Arising from RIA

Jimmie L. Joseph
University of Texas at El Paso, USA

David P. Cook
Old Dominion University, USA

INTRODUCTION

New technologies can lead to social upheaval and ethical dilemmas which are unrecognized at the time of their introduction. Medical care technology has advanced rapidly over the course of the past two decades and has frequently been accompanied by unforeseen consequences for individuals, the medical profession and government budgets, with concomitant implications for society and public policy (Magner, 1992; Marti-Ibanez, 1962).

Advances in information technology (IT) during the last decade and a half are now impacting the medical profession, and the delivery of medical advances, in ways that will impact public policy debates for the foreseeable future. The World Wide Web (Web) makes information that was once the eminent domain of medical professionals available to average citizens who are increasingly demanding medical treatments from the leading edge of medical technology. For example, CenterWatch (www.centerwatch.com) provides a wealth of information concerning clinical trials and offers a conduit by which patients can become involved in such studies. The availability of such information has also led to patients suffering from life-threatening diseases not part of such clinical trials to request special access to potentially life-saving therapies. As a result, the Web is increasing the complexity of answering public policy questions surrounding what medical technologies to make available to the public, who will be eligible to receive new medical treatments, and at what cost.

BACKGROUND

In medicine, it has traditionally been the medical practitioner who has possessed the greater breadth and depth of information and knowledge in the provider-patient relationship (Magner 1992; Porter 1992; Robinson 1931). This condition, where one party has more information in a transaction than another party or parties (relative to the transaction), is known as information asymmetry. The existence of information asymmetry, in many transactions, has created the need to develop public policies to protect the interests of parties with less information (Akerlof, 1970; Hellwig 2001; Nayyar, 1990).

The risk of false advertising, the presence of disclaimers for contests, the existence of lemon laws for cars, and insider trading laws all attest to the need for societies to establish guidelines to handle the ethical issues which can arise from information asymmetry. Public policy has been designed to protect medical services consumers from the potential deleterious effects of information asymmetry (Akerlof, 1970; Hellwig, 2001; Milgrom & Roberts 2001). Specifically, it has dictated that doctors and other medical professionals be licensed in the nation, state or province in which they practice (Davies & Beach, 2000; Digby, 1997; Fournier & McInnes, 1997). Laws governing the practice of medicine, and what is acceptable and legal in the course of treatment and medical experiments have been developed. Oversight boards have also been established to ensure healthcare providers offer care in accordance with standards of care (Fournier & McInnes, 1997). Such precautions are intended to increase patients' confidence in the care being provided, even under the yoke of information asymmetry. The precautions do not reduce or eliminate the presence of information asymmetry.

A commonly held belief is that information parity, and information access that moves parties in transactions toward parity, is the cornerstone for the elimination of the ethical quandaries introduced by information asymmetry (Akerlof, 1970; Diamond, 1984; Hellwig, 2001). Until recently, medical information has presented two problems for the lay person: accessibility and comprehensibility (Ghalioungui, 1963; Marti-Ibanez, 1962; Magner, 1992; Porter, 1992; Robinson, 1931).

Accessibility refers to the opportunity to find, as well as the ease and convenience associated with locating, information. Medical journals are expensive, and prior to online libraries, many university libraries prided themselves on the expansiveness of their collections of journals. For a patient with a medical problem, or a family problem, a primary constraint was the ability to locate and access the salient medical information dealing with the problem. If the patient did not live in a town with a medical school, he or she would have to depend on the local library's often sparse repository of medically-related offerings to search for relevant information. Generally, assuming the patient could identify an appropriate source of information, it was not locally available (Digby, 1997; Matthews, 2000). This would necessitate relying upon the availability of interlibrary loan or a trip to the library that held the sought-after text or journal. More likely than not, however, the affected patient would not have the ability to reference such medical resources. Most commonly, the patient's knowledge of medical treatment options was limited to that offered by a medical care provider, word-of-mouth, or from popular press publications. Historically, lack of information accessibility has been a powerful protector of information asymmetry.

Even if patients could locate and access relevant medical information they were still faced with the problem of comprehensibility. Medical journals are written for an audience with a minimum understanding of biology, chemistry, physiology, anatomy, biochemistry, and pharmacology. There is a requisite base of knowledge for understanding medical texts and journals, and much medical literature is written for an audience with specialized knowledge of a particular field of medicine. For the lay audience, such literature is virtually indecipherable.

With the practically insurmountable twin problems of accessibility and comprehensibility, patients received most of their medical information through the filter of the physician. Few physicians could be considered members of the *avant garde*. Consequently, patients were more than likely to learn only of medical therapies with which their physicians were familiar (Marti-Ibanez, 1962; Porter, 1992; Robinson, 1931). Thus, information on progressive treatments not approved for use, or uncovered by medical insurance, would not be widely disseminated to the general public through physicians. Further, licensed physicians familiar with such treatments would face numerous legal hurdles to providing information regarding experimental or unapproved medical therapies.

DEVELOPMENTS THAT HAVE REDUCED INFORMATION ASYMMETRY

In the twentieth century, information on many medical and scientific advances was made available to a large cross-section of the population via print and electronic mass media. However, the availability of such media outlets did not significantly alter the level of information asymmetry between the patient and the medical practitioner. First, national oversight of media outlets limited their ability to report on medical treatments and breakthroughs not licensed in that particular country (Barry & Raworth, 2002). Perhaps, just as importantly, mass media tend to report on items of broad interest to their constituents. Esoteric medical discoveries of interest to a minute segment of the population were less likely to garner much media attention. This dynamic has shifted radically since Tim Berners-Lee's creation of the World Wide Web in the 1990s. With this innovation, the barriers to information symmetry in medical information have eroded, albeit imperceptibly at first.

The Web facilitates global access to medical information without regard to national laws or availability of treatment. The filter on medical information provided by physicians (i.e., national governments, regional health boards and media censors) was removed, and patients could access information at a rate nearly commensurate with their physicians'. Patients are made aware of new and developing medical treatments rapidly via such online outlets as CenterWatch and ClinicalTrials.gov (www.clinicaltrials.gov). ClinicalTrials.gov is a service of the U.S. National Institutes of Health. Numerous Web sites are also dedicated to alternative medical therapies for a variety of illnesses. Yahoo! Health provides a starting point for identifying potential alternative medicine resources (dir.yahoo.com/Health/alternative_medicine). Individuals are thus free to decide which medical innovations interest them, and then seek information on them.

The information availability barrier to information symmetry has been significantly reduced, leaving the difficulty with comprehensibility as a challenge to information parity. The Internet and Web, however, allow motivated individuals to locate multiple presentations of information on the same medical innovation in an effort to identify an intelligible (to them) presentation of the information. E-mail, online chats, text messaging, personal computer (PC) based video conferencing and inexpensive long distance through voice over IP (VOIP) facilitate the ability of communities of consumers sharing similar medical conditions to pool their intellectual capital in an effort to interpret the extant information on a medical innovation and to reduce that information to the lowest common denominator of intelligibility. Wikis constitute a powerful tool by which all kinds of, including healthcare, information is being disseminated. A plethora of healthcare topics ranging from the flu to HIV are available. One of the advantages of wikis is the fact they are developed and edited by online communities of users. Consequently, even conceptually difficult and technically rigorous knowledge bases can be interpreted and presented in a non-technical manner understandable to a broad readership. As a result, the second barrier to information symmetry continues to be reduced.

Technology has created a reduction in asymmetry (RIA) between the medical professional and the patient. Where the patient in the office of a medical professional once considered the practitioner before him to be the expert on the treatment of a condition, it is now feasible for the patient to have more information on the available treatments than the attending physician. The rapid advance of medical innovations and the volumes of information associated with each make it virtually impossible for a physician to be an expert on the entire spectrum of possible treatment options. Further, many medical therapies are experimental and will not find their way into mainstream medical care or even be approved for use for years. While a physician may have a passing familiarity with such therapies, it is unlikely he or she will have a strong motivation to become expert on them. A patient with a particular malady, however, not only has great incentive to research that condition and all treatments, including those that may currently be unapproved or categorized as experimental, but the Web also provides a means by which he or she can do so.

Online communities for particular maladies offer their sufferers a means of support in coping with their illnesses. Arguably, such communities provide a variety of informational and emotional support. ProHealth's ArthritisSupport.com (www.arthritisSupport.com) provides an example of such an online community. It provides a full range of information and support for sufferers of both osteoarthritis and rheumatoid arthritis in the form of online chats, message boards, coping tips, clinical trial and drug information, physician selection, as well as information related to disability claims. The existence of such communities demonstrates the power of the Web to bring patients with similar interests together for their mutual benefit in terms of support and a reduction in information asymmetry.

NEW RESEARCH ISSUES

The supposition of information asymmetry research (e.g., in economics and finance) is that the efficiency of markets can be improved by diminishing the level of asymmetry in transactions (Akerlof, 1970; Milgrom & Roberts, 2001; Payne, 2003). Further, research and public policy is predicated on the commonly held belief that information parity, and information access that moves the parties towards parity, is the cornerstone for the elimination of the ethical quandaries introduced by information asymmetry (Akerlof, 1970; Cowles, 1976; Diamond, 1984; Hellwig, 2001). However, movement toward information symmetry in the patient-provider dyad creates a number of issues that have implications for public policy makers, a sample of which is discussed here.

A significant issue with patients having unfettered access to medical information is that the information available online is not limited by geopolitical boundaries. The Web provides information on pharmaceuticals, treatments, innovations, technologies and facilities that may not be approved for use in the patient's home country or region. Large numbers of patients who desire and demand locally unavailable treatments place pressure on policy makers to examine the possibility of making such treatments available, perhaps prematurely. In some cases, pressure may mount to make a treatment available even if the long-term effects and efficacy of the treatment are uncertain.

Policy makers are placed in the position of determining whether to relax standards by which treatment options are evaluated prior to their release for public consumption. Already, we have seen mechanisms put into place to make investigational new drugs (IND) available to certain patients on an emergency, individual, or compassionate basis. The Regulatory Affairs Professionals Society (www.raps.org) is an organization dedicated to the health product regulatory profession with members from 50 countries involving industry, government, research, clinical, and academic organizations. Such organizations are intended to serve as a resource for information and education for those involved in the regulation of the healthcare sector.

Budgetary constraints frequently dictate the availability of medical therapies, particularly those that are new and expensive (Barry & Raworth, 2002; Schwartz et al., 1999). In nations with privatized healthcare, insurance companies may refuse to cover such treatments in favor of less expensive alternatives, or require the insured to pay a higher co-payment. In societies where healthcare has been nationalized, budgetary constraints are particularly troublesome since decision makers must allocate scarce medical resources either regionally or nationally. Consequently, even approved medical treatments may not be available in a particular area, or may be rationed (Barry & Raworth, 2002; Schwartz, Morrison et al., 1999).

Thus, if exogenous information enters the system and creates demand for a particular form of care, patients may seek a mix of healthcare alternatives unanticipated by medical boards. Conversely, citizens may not seek care in the volume expected by medical boards at particular locations. Consequently, resources may be over-allocated in one area, and under allocated in others. In situations where RIA leads patients to treatments which may be available in another country, the unanticipated current year expenditure to cover the extra-national treatment must come from existing budgets. Thus, funds budgeted for the current year for intra-national treatments are unavailable, and some must do without or a deficit will result. Additionally, allowing those on a long waiting list (or those excluded from the waiting list altogether) to jump ahead in the queue for treatment elsewhere, encourages those with low priority in the queue to find alternative treatment avenues.

Another serious issue may develop with patients' easy access to medical professionals in other countries. Scenarios exist where a locally unavailable treatment is available elsewhere and discussed with a patient. Patients may become well informed of the side-effects, efficacy, and benefits of a wide range of treatments not available to them; nonetheless, they will seek these treatments. When they learn of the treatments' unavailability they will demand the treatment be made available to them, or they may seek illegal means of acquiring the therapy. Online sources of certain medical services and therapies are a likely means by which to circuitously avail oneself of desired medical treatment or pharmaceuticals. E-pharmacies, for example, represent a possible conduit for the improper dispensing or illegal sale of drugs in a region or country. The challenge, here, resides in how to allow the online healthcare provider to continue to offer valuable services while ensuring the source is not abused.

The rapid rise of online communities, healthcare wikis, and other Web-based medical information sources leads to a number of important issues that need to be addressed. One such issue revolves around the accuracy of the information provided to patients via such conduits. How can we assure the veracity of the information? This is a particularly relevant question when considering user-developed informational sources such as wikis. In a wiki, the information's accuracy is assessed by the users themselves. Consequently, it is possible that conceptually difficult and technically rigorous information may be incorrectly interpreted and consolidated by users. Thus, the simplification of information may pose certain dangers to its potential consumers. Further, concerns exist with respect to online articles targeted toward specific drugs or medical therapies. How can a user of such information be assured of the information's quality? In traditional print media, we can rely, to some extent, upon an external review process. This quality check, however, is less certain in online arenas.

Another potential danger of the ubiquitous availability of medical information is information overload. What are the potential deleterious effects of information overload? It may be possible that this information overload may contribute to patients' increased confusion regarding medical conditions and available treatment modalities as opposed to alleviating such confusion. Many patients arguably make decisions concerning their healthcare, at least in part, based on the information they find online. Consequently, what is the legal culpability of the purveyors of such information when patients make decisions that are detrimental rather than beneficial to their medical care on the basis of said information?

Another relevant issue that must be addressed is the degree to which governments are responsible for controlling and disseminating available information to their citizenries. In what types of information campaigns should governments engage? The answer to this question may lie to some degree upon national cultures as well as the nature of the government in question. With the wide availability of information sources that link online users to mainstream medical modalities as well as those considered to be alternative medical modalities, some governments provide advice with respect to using such sources of information. For example, the National Center for Complementary and Alternative Medicine (www.nccam.nih.gov) in the United States provides links to its Web site visitors that explain complementary and alternative medical therapies in an effort to better educate medical consumers. The Australian government also maintains an internet presence to better educate its populace and to provide links to medical information through its Department of Health and Ageing (www.health.gov.au) as does the UK through the Health and Well-Being section of its DirectGov Web site (www.direct.gov.uk/HealthAndWellBeing/fs/en).

Since many governments attempt to provide their citizens access to relevant health-related information, two important issues arise: (1) at what point does government responsibility end and (2) are less educated or technically aware citizens exposed to increased information asymmetry. The first issue does not arise strictly from RIA, but does pose an interesting policy issue. The response to this issue will clearly vary depending upon the nature of country's society (e.g., socialist versus capitalist). The second issue also poses interesting policy implications that are directly related to information asymmetry.

If we assume that lower income and/or educational levels contribute to less access to Internet-based information technologies for individuals, then is it reasonable to assume that these individuals are subject to traditional levels of information asymmetry? Consequently, we can expect RIA gaps to vary widely across the citizens within and between countries. As a policy issue within a country, then, the problem that must be addressed is how to halt the growth of, and reduce, RIA gaps. How can a society reduce the educational disparity that results from differing socio-economic statuses? And, assuming such a disparity can be appreciably reduced, how can that society grant equal access to the internet-based information technologies that have contributed so much to RIA?

FUTURE TRENDS

Web technologies not only provide a means by which to increase information parity, but also represent a vehicle by which the scientific process may be compromised. Individuals participating in clinical trials can easily set up chat areas, news groups and email mailing lists to contact others in the same or similar studies. This allows participants to compare symptoms and progress, and may skew or completely thwart a medical study. Individuals in severe or chronic pain and those suffering from life threatening illness have great motivation to ensure they are getting the most efficacious treatment available. From the standpoint of public policy, however, this is problematic. While no one would argue that an individual does not have the right to seek the best treatment available for his or her disorder, altering the results of a clinical trial reduces the data available to determine what is best for an entire class of patients (Cowles, 1976; Digby, 1997; Matthews, 2000; Spilker & Cramer, 1992).

As Web access continues on its march toward ubiquity, the world's better-informed citizenry will, through their behaviors and desire for increasingly sophisticated medical therapies, place increased burdens on the health systems of their respective countries. Policy makers will continue to see citizens insinuating themselves into the process by which societies allocate their scarce medical resources. Debates about what medical technologies to make available, to whom, and at what cost which were once held in the relatively isolated environs of governmental and insurance provider offices are increasingly being influenced by the special interests of a diverse group of Web-savvy citizens. We can expect to see continued growth in the number and sophistication of the forums by which such special interests are expressed. Consequently, not only will policy makers have to continue to make difficult decisions about healthcare delivery, but will also have to do so under the unforgiving lamp of public scrutiny. This will place a burden on policy makers to provide mechanisms by which public concerns can be effectively factored into the decision-making process.

We are already seeing pressure mount to make medical therapies such as complementary and alternative medical treatments and promising treatments not yet authorized by the FDA available to citizens of the United States through authorized health care practitioners. This pressure has resulted in the introduction of such proposed legislation as the Access to Medical Treatment Act to the U.S. House of Representatives. Such proposed legislation is strongly supported by special interest groups such as the American Association for Health Freedom (www.healthfreedom.net). Another example of the interest in allowing better access to experimental drugs is demonstrated by the Abigail Alliance (www.abigail-alliance.org). This Web site is dedicated to providing patients with life-threatening illnesses access to potentially life-saving treatments. Further, foundations such as the LiveStrong Lance Armstrong Foundation (www.livestrong.org) are highly visible sources of support and information for sufferers of life-threatening diseases. The call to action is clear. Public policy decision makers must be prepared to deal with the cries of their citizenry. Dynamically assessing an increasingly complex information environment with ever-better informed medical consumers is necessary. It is not sufficient to react to demands for better access to medical care; policy makers must proactively develop coherent policies and information campaigns to facilitate the flow of relevant medical information and care.

CONCLUSION

This then is a microcosm of the problem facing public policy decision makers and society: Which is more important to society—individual medical treatment or the societal distribution of scarce medical resources? Is it in the public interest to allocate limited medical resources to those with the best information, or should medical professionals be trusted as better qualified to determine appropriate treatments and to prioritize patient access to medical technology?

RIA between patients and medical practitioners raises public policy issues which will affect the delivery of medical care for decades to come. It is unrealistic to think that persons or families facing grave illnesses will not seek all available information that may improve their lives or the lives of their loved ones. Whether it is paid by the patient, private health insurance or public funds from national insurance or nationalized healthcare, patients want the best and most efficacious care available. For most of these areas, RIA and increasing information parity reduces ethical, social and public policy issues. For the practice of medicine, however, RIA places physicians and governments in uncharted ethical and policy territory. An awareness of the issues that can be raised by RIA in this unusual circumstance can help the authors of public policy to prepare for social changes, and not be caught unawares.

FUTURE RESEARCH DIRECTIONS

IT facilitates communication and can de-layer the communications path between medical researchers and patients affected with a specific condition. Future research could examine the use of medical information bulletin boards, chat room and virtual communities by patients to exchange treatment results. As a means of determining motivation levels of patients, the research could survey patients to see what research they have done on their disease or condition.

Future surveys could explore medical researchers and their experiences with control groups contaminated by the sharing of treatment results. Further, examination of physicians who recommended/submitted patients to participate in medical research can be conducted. The physicians could track the frequency with which patients ask questions indicating advanced knowledge of the study. The information could be combined with study information on the types of diseases being treated and ages of the patients in the medical study.

Future research could investigate the degree to which non-hospital treatment centers assist in the dissemination of medical information. Many of these centers are staffed by nurses, not physicians, and may act as an additional layer of filtering for information to the patient.

A study of the correlation between Internet access speed and the questions posed to physicians could help in understanding the impact of broadband access on the RIA. Further, a correlation with income could disclose if a digital divide is forming in medical care. If information concerning medical studies and treatment options is readily available to those with computers and Internet access, but the same information is less readily available to non-computerized patients, medical studies may be weighted towards the wealthy and technically savvy.

Future research could explore the creation of a clearinghouse for researchers to report tainted studies. In this way, the medical and IT communities can determine the degree of medical research contamination. Ethics researchers can explore and debate the future of the physician-patient interaction, and how this will change as physicians lose their monopoly on diagnosis and treatment options. Public policy researchers can explore the legal options which may be necessary to ensure that medical treatments are equitably available. Public policy researchers can also examine the necessity of maintain laws which limit the discussions physicians can have with their patients, in the face of internationalization of medical technology information.

Future research could also explore other areas in which RIA produces greater ethical issues for one of the parties. As information technology becomes pervasive, it is likely that other fields will discover that RIA does not always lead to more efficient markets. This understanding will allow researchers to identify situations in which RIA leads to greater moral dilemmas, not a reduction in ethical considerations.

REFERENCES

Akerlof, G. A. (1970). The market for 'lemons:' Quality uncertainty and the market mechanism. *Quarterly Journal of Economics, 84*(3), 788-500.

Barry, C., & Raworth, K. (2002). Access to medicines and the rhetoric of responsibility. *Ethics & International Affairs, 16*(2), 57(15).

Cowles, J. (1976). *Informed consent*. New York: Coward, McCann & Geoghegan, Inc.

Davies, C., & Beach, A. (2000). *Interpreting professional self-regulation*. London: Routledge Taylor & Francis Group.

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *The Review of Economic Studies, 51*(3), 393-414.

Digby, A. (1997). *The patient's view. Western medicine. I*. Oxford University Press.

Fournier, G. M., & McInnes, M. M. (1997). Medical board regulation of physician licensure: Is excessive malpractice sanctioned? *Journal of Regulatory Economics, 12*(2), 113-126.

Ghalioungui, P. (1963). *Magic and medical science in ancient Egypt*. New York: Barnes & Noble, Inc.

Hellwig, M. F. (2001). Risk aversion and incentive compatibility with ex post information asymmetry. *Economic Theory, 18,* 415-438.

Magner, L. N. (1992). *A history of medicine*. New York: Marcel Dekker, Inc.

Marti-Ibanez, F. (1962). *The epic of medicine*. New York: Clarkson N. Potter, Inc.

Matthews, J. N. S. (2000). *An introduction to randomized controlled clinical trials*. London: Arnold, A member of the Hodder Headline Group.

Milgrom, P., & Roberts, J. (2001). Informational asymmetries, strategic behavior, and industrial organization. *Game Theory and Industrial Organization, 77*(2), 184-193.

Nayyar, P. R. (1990). Information asymmetries: A source of competitive advantage for diversified service firms. *Strategic Management Journal, 11,* 513-819.

Payne, R. (2003). Informed trade in spot foreign exchange markets: An empirical investigation. *Journal of International Economics, 61,* 307-329.

Porter, R. (1992). *The patient in England, c. 1660-c. 1800. Medicine in Society*. Cambridge: Cambridge University Press.

Robinson, V. (1931). *The story of medicine*. New York: Tudor Publishing Co.

Schwartz, L., Morrison, J., et al. (1999). Rationing decisions: From diversity to consensus. *Healthcare Analysis, 7*(2), 195-205.

Spilker, B., & Cramer, H. A. (1992). *Patient recruitment in clinical trials*. New York: Raven Press.

FURTHER READING

Aguolu, I. E. (1997). Accessibility of information: A myth for developing countries? *New Library World, 98*(1132), 25.

Annonymous. (2003a). *Approvals of FDA-regulated products.* Retrieved March 30, 2004, from http://www.fda.gov/opacom/7approvl.html

Anonymous. (1998, Modified:09/01/1998). *What is insider trading*? Retrieved October 13, 2003, from http://www.sec.gov/divisions/enforce/insider.htm

Anonymous. (2003, 03/25/2002). *Complaints about broadcast advertising.* Retrieved October 13, 2003, from http://ftp.fcc.gov/cgb/consumerfacts/advertising.html

Anonymous. (2003). *For business: Advertising guidance*. Retrieved October 13, 2003, from http://www.ftc.gov/bcp/guides/guides.htm

Anonymous. (2003b, Modified:06/27/2003). *The laws that govern the securities industry.* Retrieved October 13, 2003, from http://www.sec.gov/about/laws.shtml

Barber, B. (1980). *Informed consent in medical therapy and research*. New Brunswick, NJ: Rutgers University Press.

Borgman, C. L. (2000). *From Guttenberg to the global information infrastructure: Access to information in the networked world*. Cambridge, MA: MIT Press.

Felch, W. C., & Scanlon, D. M. Scanlon (1997). Bridging the gap between research and practice: The role of continuing medical education. *JAMA, The Journal of the American Medical Association, 277*(2), 155(2).

Friedman, L. M., Furberg, C. D., et al. (1998). *Fundamentals of clinical trials*. New York: Springer-Verlag.

Gert, H. J. (2002). Avoiding surprises: A model for informing patients. *The Hastings Center Report, 32*(5), 23(11).

Harris, M., & Raviv, A. (1979). Optimal incentive contracts with imperfect information. *Journal of Economic Theory, 20,* 231-259.

Hassinger, E. W. (1982). *Rural health organization*. Ames: Iowa State University Press.

Holmstrom, B. R. (1985). *The provision of services in a market economy. Managing the service economy: Prospects and problems*. Cambridge, UK: Cambridge University Press.

Joseph, J. L., & Cook, D. P. (2005). Information imbalance in medical decision making: Upsetting the balance. In L. A. Freeman & A. G. Peace (Eds.), *Information ethics: Privacy and intellectual property.*. Hershey, PA: Information Science Publishing.

Majno, G. (1975). *The healing hand: Man and wound in the ancient world*. Cambridge, MA: Harvard University Press.

Marks, F. R., & Cathcart, D. (1974). Discipline within the legal profession: Is it self-regulation? *University of Illinois Law Review, 193*(Spring), 236.

Marti-Ibanez, F. (1958). Books in the physician's life. In F. Marti-Ibanez (Ed.), *Centaur: Essays on the history of medical ideas* (pp. 3-28). New York: MD Publicantions.

Marti-Ibanez, F. (1958). Centaur: Essays on the histroy of medical ideas. New York: MD Publications.

Marti-Ibanez, F. (1961). *A prelude to medical history*. New York: MD Publications.

Morgan, J. M. (2003). New device indications: Practice and cost implications in Europe. *Cadiac Electrophysiology Review, 7*(1), 49-53.

Osler, S. W. (1921). *The evolution of modern medicine*. New Haven, CT: Yale University Press.

Pennachio, D. L. (2003). CME: How to get yours? Requirements are multiplying, but so are the opportunities to meet them. *Medical Economics, 80*(16), *21*(3).

Shaneyfelt, T. M. (2001). Building bridges to quality. *JAMA, The Journal of the American Medical Association, 286*(20), 2600(2).

Shavell, S. (1979). Risk sharing and incentives in the principal and agent relationship. *Bell Journal of Economics, 10*, 55-73.

Summers, A. (1997). Nurses and ancillaries in the Christian era. In I. Loudon (Ed.) *Western Medicine* (pp. 192-205). Oxford: Oxford University Press.

KEY TERMS

E-Pharmacy: A term that refers to the existence of an online pharmacy that offers medical care providers the ability to prescribe medications for patients online. These medications can then be delivered to patients without ever requiring them to leave the confines of their homes. In some, cases prescriptions can be filled without a physician ever having physical contact with the patient.

Ethical Issue: A situation where the decision maker is required to weigh values and employ judgment in reaching a decision. Frequently, this is a situation which is outside of, or different from, those with which the decision maker has previous experience.

Information Asymmetry: A situation in which one party has more information in a transaction than another party or parties, relative to the transaction.

Information Parity: As situation in which both parties in a transaction have equal information related to the transaction.

Information Accessibility: The opportunity to find, as well as the ease and convenience associated with locating, information. Often, this is related to the physical location of the individual seeking the information and the physical location of the information in a book or journal.

Information Comprehensibility: The ability to comprehend and utilize information once located. Comprehensibility, if the information is syntactically and grammatically correct, is often dependant on the preparation and training of the individual accessing the material.

Locally Unavailable Treatment: Medical treatment which may not be available in a nation or region for economic, regulatory or technical reasons. The treatment may also be limited by law or administrative actions.

Reduction in Information Asymmetry (RIA): Reducing the disparity in information between the parties in a transaction, relative to the transaction.

This work was previously published in Handbook of Research on Public Information Technology, edited by G. Garson and M. Khosrow-Pour, pp. 324-332, copyright 2008 by Information Science Publishing (an imprint of IGI Global).

Chapter 7.4
Access Control for Healthcare

Yifeng Shen
Monash University, Australia

INTRODUCTION

Thanks to the rapid development in the field of information technology, healthcare providers rely more and more on information systems to deliver professional and administrative services. There are high demands for those information systems that provide timely and accurate patient medical information. High-quality healthcare services depend on the ability of the healthcare provider to readily access the information such as a patient's test results and treatment notes. Failure to access this information may delay diagnosis, resulting in improper treatment and rising costs (Rind et al., 1997).

Compared to paper-based patient data, computer-based patient data has more complex security requirements as more technologies are involved. One of the key drivers to systematically enhance the protection of private health information within healthcare providers is compliance with the healthcare information system security standard framework and related legislation. Security standards and legislation of the healthcare information system are critical for ensuring the confidentiality and integrity of private health information (Amatayakul, 1999). Privacy determines who should have access, what constitutes the patient's rights to confidentiality, and what constitutes inappropriate access to health records. Security is embodied in standards and technology that ensure the confidentiality of healthcare information and enable health data integrity policies to be carried out.

Based on the investigation of security standard and legislation, we can analyze and create basic security requirements for the healthcare information system. To meet the security requirements, it is necessary to deploy an appropriate access control policy and system within the organization. As discussed elsewhere (Sandhu, Coyne, Feinstein, & Youman, 1996), role-based access control (RBAC) is a promising technology for managing and enforcing security in a large-scale distributed system. In the healthcare industry, RBAC has already been adopted by the Health Level Seven (HL7) organization as a key access control standard (Blobel & Marshall, 2005).

HL7 was established in 1987 to develop standards for the electronic interchange of clinical, financial, and administrative information among independent healthcare-oriented computer systems. In June of 1994, HL7 was designated by the American National Standard Institute (ANSI) as an ANSI-accredited standards developer. HL7, in its draft Security Service Framework (Kratz et al., 2005) categorizes healthcare information security exposures in the following manner:

- **Disclosure:** Exposure, interception, inference intrusion
- **Deception:** Masquerade, falsification, repudiation
- **Disruption:** Incapacitation, corruption, obstruction
- **Usurpation:** Misappropriation

Although RBAC has been introduced to the latest version of HL7 (version 3) for strengthening the security features, it only includes those basic functions. Due to the complexity of the healthcare process, RBAC with only basic functions may not be sufficient. More context constraints need to be processed in addition to traditional RBAC operations.

The major contributions we have made in this article are:

- Illustrating the detailed design of a flexible and securer RBAC model for a healthcare information system based on HL7 standard;
- Introducing the basic elements of HL7 v3 and RBAC, which are necessary for us to realize our proposed model; and
- Analyzing the potential weakness of current HL7 standard and the basic RBAC model in terms of security and flexibility.

The rest of the article is organized as follows. The next section provides a general introduction and basic analysis of HL7 version 3. We then explain the RBAC concept model and describe our major work, and finish with our conclusion and future work.

HL7 VERSION 3

What is HL7?

Health Level Seven is one of several American National Standards Institute-accredited Standards Developing Organizations (SDOs) operating in the healthcare arena. Most SDOs produce standards (sometimes called specifications or protocols) for a particular healthcare domain such as pharmacy, medical devices, imaging, or insurance (claims processing) transactions. HL7's domain is clinical and administrative data (HL7, 2005).

HL7 is also a non-profit volunteer organization. Its members are the providers, vendors, payers, consultants, and government groups who have an interest in the development and advancement of clinical and administrative standards for healthcare services. In its achievements so far, HL7 has already produced HL7 Version 2 (HL7 v2) specifications (HL7, 2005), which are in wide use as a messaging standard that enables disparate healthcare applications to exchange key sets of clinical and administrative data. However, the newer specification HL7 Version 3 (HL7 v3), still under development, pertains to all aspects of clinical and administrative data in health services. Unlike its older version, HL7 v3 specifications are completely based upon the extensible markup language (XML) standards, and so have potential to win an instant acceptance by developers and vendors alike.

The target system during our research is based on HL7 v3, so only HL7 v3 will be described in this article.

The lack of data and process standards between both vendor systems and the many healthcare provider organizations present a significant barrier to design application interfaces. With HL7 v3, vendors and providers will finally have a messaging standard that can provide solutions to all of their existing problems.

HL7 v3 is based on a reference information model (RIM). Although RIM is not stabilized yet, once it is stabilized, it will be the most definitive standard to date for healthcare services.

The following section will highlight some key components of RIM.

Reference Information Model

RIM is the cornerstone of the HL7 Version 3 development process. An object model created as part of the Version 3 methodology, RIM is a large pictorial representation of the clinical data (domains) and identifies the lifecycle of events that a message or groups of related messages will carry. It is a shared model between all the domains and as such is the model from which all domains create their messages. RIM comprises six main classes (Beeler et al., 2005):

1. **Act:** Represents the actions that are executed and must be documented as health care is managed and provided.
2. **Participation:** Expresses the context for an act in terms such as who performed it, for whom it was done, where it was done, and so forth.
3. **Entity:** Represents the physical things and beings that are of interest to and take part in health care.
4. **Role:** Establishes the roles that entities play as they participate in health care acts.
5. **ActRelationship:** Represents the binding of one act to another, such as the relationship between an order for an observation and the observation event as it occurs.
6. **RoleLink:** Represents relationships between individual roles.

Three of these classes—Act, Entity, and Role—are further represented by a set of specialized classes or sub-types.

RIM defines all the information from which the data content of HL7 messages are drawn. It follows object-oriented modeling techniques, where the information is organized into classes that have attributes and that maintain associations with other classes. RIM also forms a shared view of the information domain used across all HL7 messages, independent of message structure.

HL7 v3 Security

The focus of HL7 security needs analysis on how systems communicate information using HL7 message. It is expected that healthcare application systems that implement HL 7 v3 will be required to have significantly more functionalities to protect the confidentiality of patient information and to authenticate requests for services than has been common in the past. The new functions may include, but are not limited to, limiting the right to view or transfer selected data to users with specific kinds of authorization, and auditing access to patient data, electronic signature, and authentication of users based on technologies more advanced than passwords. Version 3 will seek out and reference standards such as X.500 (Weider, Reynolds, & Heker, 1992) and RFC 1510 to support conveying the necessary information from one healthcare application system to another, so that these systems may perform the authorization and authentication functions. Version 3 will also seek out and adopt industry security standards that support conveying the necessary information from one healthcare application system to another, so that these systems may perform the confidentiality functions.

To meet the security goals, the HL7 Secure Transaction Special Group has created a security service framework for HL7 (Kratz et al., 2005). According to the scope of the framework, HL7 must address the following security services: authentication, authorization and access control, integrity (system and data), confidentiality, accountability, availability, and non-repudiation. The HL7 security service framework uses case scenarios to illustrate all the services mentioned above. All those case scenarios can help the readers to understand those services in a very direct way. However case scenarios are not detailed enough to be an implementation guide for the security services.

In this article we are going to design a flexible model for one key security service—access control. This model will extend the case scenarios to a very detailed level which can be directly used as an implementation guide for HL7 v3.

Figure 1. (

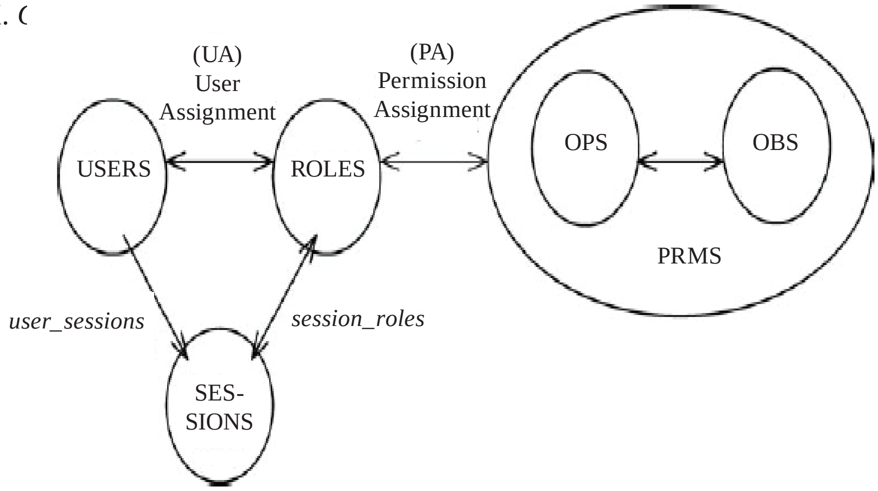
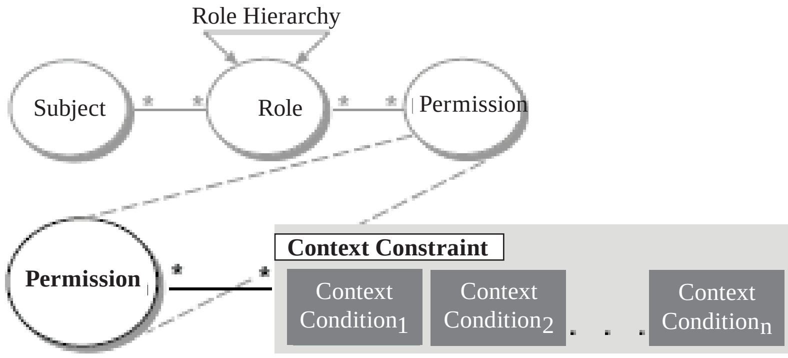


Figure 2. RBAC permissions with context constraints



ROLE-BASED ACCESS CONTROL

RBAC has became very popular in both research and industry. RBAC models have been shown to be "policy-neutral" in the sense that by using role hierarchies and constraints (Chandramouli, 2003), a wide range of security policies can be expressed. Security administration is also greatly simplified by the use of roles to organize access privileges. A basic RBAC model will be covered in this section, as well as an advanced model with context constraints.

Basic RBAC Model

The basic components of the RBAC model are *user*, *role*, and *permission* (Chen & Sandhu, 1996). The user is the individual who needs access to the system. Membership to the roles is granted to the user based on his or her obligations and responsibilities within the organization. All the operations that the user can perform should be based on the user's role.

Role means a set of functional responsibilities within an organization. The administrator defines roles, a combination of obligation and authority in organization, and assigns them to users. The user-role relationship represents the collection of users and roles.

Permission is the way for the role to access more than one resource.

As shown in Figure 1, the basic RBAC model also includes *user assignment* (UA) and *permission assignment* (PA) (INCITS359, 2003).

The user assignment relationship represents which user is assigned to perform what kind of role in the organization. The administrator decides the user assignment relationship. When a user logs on, the system UA is referenced to decide which role it is assigned to. According to the object that the role wants to access, the permission can be assigned to the role referenced by the permission assignment relationship.

The set of permissions (PRMS) is composed of the assignments between operations (OPS) and objects (OBS).

UA and PA can provide great flexibility and granularity of assignment of permissions to roles and users to roles (INCITS359, 2003). The basic RBAC model has clearly illustrated the concept about how role-based access control works within an organization. However it may not be dynamic enough when the business process becomes very complex. Thus the idea of context constraints is introduced to make the RBAC model more useful.

RBAC Model with Context Constraints

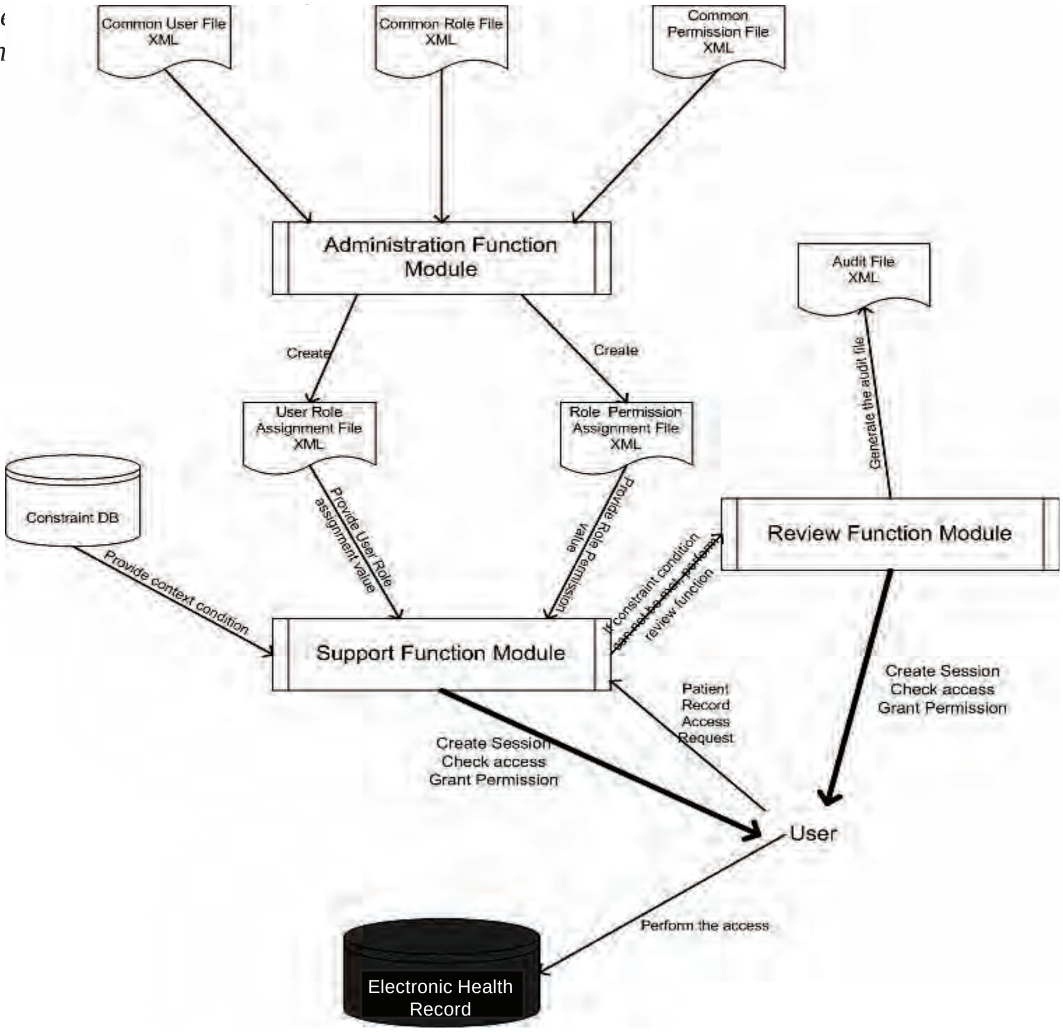
Traditional RBAC supports the definition of arbitrary constraints on the different parts of a RBAC model (Sandhu et al., 1996). With the increasing interest in RBAC in general and constraint-based RBAC in particular, research for other types of RBAC constraints has gained more attention (Bertino, Bonatt, & Ferrari, 2001). In this section we describe the context constraints in an RBAC environment.

A context constraint is an abstract concept. It specifies that certain context attributes must meet certain conditions in order to permit a specific operation. As authorization decisions are based on the permissions a particular subject/role possesses, context constraints are associated with RBAC permissions (see Figure 2).

The context constraint is defined through the terms context attribute, context function, and context condition (Strembeck & Neumann, 2004):

- A *context attribute* represents a certain property of the environment whose actual value might change dynamically (like time, date, or session-data, for example) or which varies for different instances of the same abstract entity (e.g., location, ownership, birthday, or nationality). Thus, context attributes are a means to make context information explicit.
- A *context function* is a mechanism to obtain the current value of a specific context attribute (i.e., to explicitly capture context information). For example, a function *date*() could be defined to return the current date. Of course, a context function can also receive one or more input parameters. For example, a function *age*(*subject*) may take the subject name out of the *subject*, *operation*, *object_* triple to acquire the age of the subject, which initiated the current access request, for example, the age can be read from some database.
- A *context condition* is a predicate that consists of an operator and two or more operands. The first operand always represents a certain context attribute, while the other operands may be either context attributes or constant values. All variables must be ground before evaluation. Therefore, each context attribute is replaced with a constant value by using the corresponding context function prior to the evaluation of the respective condition.
- A *context constraint* is a clause containing one or more context conditions. It is satisfied if and only if (iff) all context conditions hold. Otherwise it returns false.

Figur
health



A context constraint can be used to define conditional permissions. Based on the terms listed above, the conditional permission is a permission associated with one or more context constraints, and grants access iff each corresponding context constraint evaluates as "true."

As we can see, a context constraint can help the organization provide more flexible and securer control for the RBAC model.

Design a RBAC Model with Context Constraints for the Healthcare Information System Based on HL 7 Version 3

The access control model we are going to describe is a method to control access on a healthcare information system. It is developed to enhance the security and flexibility of traditional access control systems. The resource to be accessed in this article is limited to a patient's electronic health record (EHR).

The primary purpose of the EHR is to provide a documented record of care that supports present and future care by the same or other clinicians. This documentation provides a means of communication among clinicians contributing to the patient's care. The primary beneficiaries are the patient and the clinician(s) (ISO/TC-215 Technical Report, 2003).

System design will include two major phases:

1. **Components design:** Describes all the necessary elements that make up the system.
2. **Data flow design:** Describes all the processes that make the whole system work.

Components Design

As described in the previous section, the RBAC system must include the basic elements such as user, role, permission, user-role assignment, and role-permission assignment. All those elements will be associated with real values in our system design. Figure 3 illustrates the overall structure of the system.

- **User:** Anybody with authenticated identity can act as the user in the system. For example, after Tom successfully logs into the hospital's computer system with his user ID 19245678, he becomes the user of our system.
- **Role:** The set of roles can be retrieved from those functional roles that already exist in the current healthcare information system such as physician, pharmacist, registered nurse, and so forth. As the number of roles is limited in our system, we can store all the role information by simply using an XML file instead of a database. This file is named "Common Role File.xml."
- **Permission:** The scope of the permissions in our design will focus on those system operations (create, read, update, delete, execute, etc.). Similar to role information, we use another XML file with the name "Common Permission File.xml" to represent all the permission information.

As shown in Figure 3, the user, role, and permission file can be used as the basic input for the whole system. To generate user-role assignment and role-permission assignment relationship, we introduce the *administration function* module. This module is designed to create and maintain the user role assignment file and role permission assignment file.

The rules of user role assignment and role permission assignment are referenced from the security section of HL7 v3 standard (HL7 Security Technical Committee, 2005).

In addition to the administration function module, we also designed another two function modules: *support function* module and *review function* module. The support function module provides the core function of the system. It receives the access request from the user and makes judgment based on the input from different sources to decide whether the access can be granted. The detailed process will be described in the next section.

The review function module is an extension of the support function module. It is used for exceptional scenarios, such as emergent circumstances that do not satisfy the constraint condition. Every time the review function module is initiated, an audit file will be created to record all the necessary information of the exceptional case. In our system, the audit file is saved in XML format with the name " Audit File.xml."

The ultimate object the users want to access is the EHR. The existing database that stores all the EHRs can be used directly by our system.

Another database included in this system is the constraint database, which stores all the context attributes and context conditions. The context attributes and context conditions can be used as input for the support function module during the access control decision process.

In summary, the components can be categorized into three types based on the design:

- **Type 1 – Basic elements:** Role, User, Permission, User Role Assignment, Role Permission Assignment, Audit File. All these basic

elements are represented in XML format.
- **Type 2 – System functional modules:** Administration Function Module, Support Function Module, Review Function Module. All these modules provide the core functions of the system and are represented in real program.
- **Type 3 – Database:** EHR database and constraint database.

Data Flow Design

After all the components are defined, we will design the proper data flows to make the system work. The data flow design is based on the three functional modules previously discussed. Thus, we introduce three kinds of data flow in this article:

- data flow for administration function module,
- data flow for support function module, and
- data flow for review function module.

Data flow for administration function module

1. Administration function module reads the common user file, common role file, and common permission file. Those files contain all the user information, role information, and permission information respectively.
2. Based on the pre-defined user-role relationship/role-permission relationship, the administration function module creates a user role assignment file and a role permission assignment file in XML format.

Data flow for support function module

1. The user sends an access request to the support function module.
2. The support function module requests and receives role information about the user from the user assignment file.
3. The support function module requests and receives the permission which is assigned to the role. This can be retrieved from the permission assignment file.
4. Get context attributes and context condition information from the constraints database.
5. The support function module performs the "check access" function then grants the access permission to the user.
6. The user can retrieve the information from the EHR database.

Data flow for review function module

1. Sometimes the context condition cannot be met, however all the other conditions (permission assignment, user assignment) can be met and the user really wants to access the resource because of emergency. In this case, all authentication information will be forwarded to the review function module.
2. The review function module records all the necessary information and generates an audit file, then grants conditional access permission to the user.
3. The user can retrieve the information from the EHR database.

All the steps listed above for the data flow just give a brief description. More detailed steps are necessary when it comes to the system implementation phase.

CONCLUSION AND FUTURE WORK

Clinical information sharing between different healthcare information systems is the key factor to improve the quality of service. Health Level Seven is the data exchange standard for clinical information. In this article we first introduce the basic concept of Health Level Seven and role-based access control. Then we illustrate how to design a flexible role-based access control model for a healthcare information system based on Health Level Seven version 3. The design utilizes the existing access control feature of Health Level Seven version 3 and integrates context constraints to make the system more secure and more flexible.

In the future, the major work will be the development of those function modules and applying this model to a real healthcare information system to see how the security access control can be improved.

REFERENCES

Amatayakul, M. (1999). *Chapter three, section 5.0: Standards, processes and organizations: Setting standards in health care information*. Computer-Based Patient Record Institute.

Beeler, G., Case, J., Curry, J., Hueber, A., Mckenzie, L., Schadow, G., et al. (2005, July 31). *HL7 reference information model 2006*.

Bertino, E., Bonatt, P.A., & Ferrari., E. (2001). TRBAC: A temporal role-based access control model. *ACM Transactions on Information and System Security, 4*(3).

Blobel, B., & Marshall, G. (2005). Role based access control (RBAC) role engineering overview. *HL7® Version 3 Standard*.

Chandramouli, R. (2003, July 27-30). Specification and validation of enterprise access control data for conformance to model and policy constraints. *Proceedings of the World Multiconference on Systems, Cybernetics and Informatics,* Orlando, FL.

Chen, F., & Sandhu, R.S. (1996). Constraints for role based access control. *Proceedings of the ACM RBAC Workshop.*

HL7. (2005). *About_HL7.* Retrieved April 2006 from http://www.h17.org

HL7 Security Technical Committee. (2005). *Role based access control (RBAC) healthcare permission catalog version 2*.

INCITS359. (2003). *Role based access control*. American National Standard Institute.

ISO/TC-215 Technical Report. (2003). *Electronic health record definition, scope, and context.*

Kratz, M., Humenn, P., Tucker, M., Nolte, M., Wagner, S., Wilson, W. et al. (2005). *HL 7 Security framework*. Retrieved April 2006 from http://www.hl7.org/library/committees/secure/HL7_Sec.html

Rind, D.M., Kohane, I.S., Szolovits, P., Safran, C., Chueh, H.C., & Barnett, G.O. (1997). Maintaining the confidentiality of medical records shared over the Internet and the World Wide Web. *Annals of Internal Medicine*, *127*(2), 138-141.

Sandhu, R.S., Coyne, E.J., Feinstein, H.L., & Youman, C.E. (1996). Role-based access control models. *IEEE Computer*, *29*(2), 38-47.

Strembeck, M., & Neumann, G. (2004). An integrated approach to engineer and enforce context constraints in RBAC environments. *ACM Transactions on Information and System Security (TISSEC)*, *7*(3), 392-427.

Weider, C., Reynolds, J., & Heker, S. (1992). *Technical overview of directory services using the X.500 protocol*. Retrieved May 5, 2006, from http://www.ietf.org/rfc/rfc1309.txt

KEY TERMS

Electronic Health Record (HER): A longitudinal electronic record of patient health information generated by one or more encounters in any care delivery setting.

Extensible Markup Language (XML): A W3C initiative that allows information and services to be encoded with meaningful structure and semantics that computers and humans can understand.

Health Level Seven (HL7): One of several American National Standards Institute (ANSI)-accredited Standards Developing Organizations (SDOs) operating in the healthcare arena.

Permission Assignment (PA): Assigns permission to an authorized role.

Reference Information Module (RIM): The cornerstone of the HL7 Version 3 development process and an essential part of the HL7 v3 development methodology. RIM expresses the data content needed in a specific clinical or administrative context, and provides an explicit representation of the semantic and lexical connections that exist between the information carried in the fields of HL7 messages.

Role-Based Access Control (RBAC): A system of controlling which users have access to resources based on the role of the user.

User Assignment (UA): Assigns a role to a user.

This work was previously published in Encyclopedia of Information Ethics and Security, edited by M. Quigley, pp. 7-14, copyright 2007 by Information Science Publishing (an imprint of IGI Global).

Chapter 7.5
Securing Mobile Data Computing in Healthcare

Willy Susilo
University of Wollongong, Australia

Khin Than Win
University of Wollongong, Australia

ABSTRACT

Access to mobile data and messages is essential in healthcare environment as patients and healthcare providers are mobile. This is inline with the need of ubiquitous computing in everyday life. Mobile and wireless devices can assist in ensuring patient's safety by providing easy availability of the data at the point of care. Portability and accessibility of these devices enhances use of them in healthcare environment. However, data integrity and confidentiality of information in them need to be ensured to provide safe, effective and efficient healthcare. Mobile healthcare involves conducting healthcare related activities through using mobile devices such as a smart phone, Personal digital assistant (PDA), wireless enabled computer, iPod and so on. Mobile computing is suitable for healthcare as healthcare providers are mobile. These would be suitable for conducting patient's healthcare activities in emergencies, ward rounds, homecare, chronic disease management, conducting clinical trials, and so on. There are various projects using mobile devices to enhance patient's care. With the advancement of medical informatics, telemedicine and information technology, mobile data devices play an enormous role in healthcare system. In this chapter, we outline the need of mobile devices in healthcare, usage of these devices, underlying technology and applications, importance of security of these devices, securing mobile data communication

in healthcare through different security models and case examples of applications that we have developed, in particular (1) iPathology tool on iPod, (2) securing healthcare information using Pocket PC 2003, and (3) securing information on handheld devices. There were several incidents in the past due to the insecurity of mobile devices that can leak information to anyone who does not have access to the information. In this chapter, we will illustrate several techniques that we have developed to protect these malicious activities and how these are applicable for securing mobile data computing in healthcare.

INTRODUCTION

In recent years, a wide range of handheld devices have been introduced in the market. Pocket size handheld organizers, called personal digital assistants (PDAs), have evolved from offering simple services such as calendar, address book, and e-mail, to working as a powerful mobile network client that can be used to access corporate databases and be used as part of an e-commerce platform. The healthcare industry is not an exception with healthcare providers using various handheld devices for accessing information and for telemedicine purposes. These days, the mobile devices are also equipped with communication abilities, such as GSM/GPRS ability, WiFi 802.11a/b/g Wireless LAN or Bluetooth. Handheld devices are perfect for carrying information, such as multimedia information, on the fly. Users can enjoy the information while they are on their way to work, etc.

Securing such devices includes providing protection for wireless communication, stored data and against viruses and malicious codes, together with making provisions for preventing theft of the device. The more we become dependent on handheld devices, the more threatening the attack will be. Today the U.S. military is already relying extensively on such devices. These devices can be used for GPS, viewing maps of areas, interacting with other troop units and as (remote) sensors (such as those used against bio-warfare), etc. Moreover, the U.S. military is spending large amounts of money to expand the capabilities of such devices. Other applications in critical infrastructures can be expected. For example, today many firefighters and ambulance services rely on paper maps. In regions in the world that expand rapidly, there is a significant advantage to switch to handheld devices where maps are updated more frequently. Since several manuals are primarily available via the Internet, handheld devices will be used by technicians to repair machinery in nuclear power plants, airplanes, etc.

There are several major differences between a desktop computer and a handheld device including limited input methods, small screen size, lower range processor, and limited memory for the handheld device. These differences affect the design of applications that run on these devices. Palm OS and Pocket PC handheld devices do not have a keyboard. Although a third party keyboard, such as Palm Portable keyboard or Targus Stowaway keyboard, can be used, most users enter text to the device with a *stylus* which is then recognised by a character recognition program such as Palm *Graffiti* or *Microsoft Transcriber*. Other methods of input include using the on-screen keyboard, or typing text on a desktop PC that is connected to the PDA. Communication with the device is achieved by using the HotSync/Activesync that synchronizes the data with its copy that is stored on a desktop computer, using a communication port such as the serial port, the IrDA (Infrared Data Association) port, or the USB port.

Palm 1000 was introduced by 3Com in 1996. Devices running Palm OS have dominated the handheld market and currently having around 70% of the market share (Hansmann, Merk, Nickous, & Stober, 2001). Palm OS is also adopted by companies such as HandSpring and IBM with applications ranging from Palm OS native applications including Address Book and Date Book, to other applications that are available as freeware, shareware or commercial products from third-party providers, including fax software, calculators, and software development kits for developers. The second major operating system for handheld computers is Windows CE (or pocket PC) which takes almost the rest of the market share. The most recent version is known as Windows Mobile 2005, with its support to GSM/GPRS mobile phone connection, together with its ability to connect to Wireless LAN 802.11a/b/g and Bluetooth.

Our Contribution

In this chapter, first we review the background on handheld devices, together with their known weaknesses. Then, we proceed with the use of handheld devices in healthcare environment. Finally, we provide some case studies that we have done in the effort to securing such devices.

Background of Handheld Devices

In this section, we review the necessary background and knowledge in handheld devices.

PALM OS PDA

Storage on a Palm PDA includes RAM and ROM that reside on memory cards. A memory card can store 256 MB and is divided into a dynamic heap and multiple storage heaps that are used for providing runtime environment, and storing data and applications. Newer models of Palm PDA, starting from Palm m505, m515, and Tungsten, are equipped with SD (Secure Digital) expansion slots that could be used for Bluetooth or Wireless LAN 802.11b devices.

Palm OS and built-in applications are stored in the ROM while user's data and applications are stored in RAM which is always powered to allow persistent storage. Applications and users' data can be backed-up on a desktop PC, or to the backup expansion pack (for the new models).

Development Systems

There are several development systems for building programs for Palm OS with C and C++ being the most commonly used languages for Palm Computing platform. Other available languages are Basic, Java, and Assembly.

A newly created application can be tested on a Palm OS Simulator or a Palm OS Emulator before it is installed on a Palm device. Palm OS Simulator is only available on the Macintosh OS and requires the application to be linked with a Simulator library that allows it to run as an independent application. Palm OS Emulators are available for Windows, Mac OS and Unix systems. The Emulator imitates the Palm OS device and allows an application to run without requiring additional library.

Third Party Program Installation

There are a number of ways of downloading programs to a Palm PDA. Firstly, *HotSync* program can be used to download new applications during the synchronization between the Palm and a desktop PC. A second method uses a third party program such as *Pilot Install* (envi.com, online). Applications can also be downloaded using IrDA port or a Bluetooth connection. IrDA drivers for desktop PCs can send a file to another IrDA point and so an application can be sent to a Palm PDA. An application can also be attached to a mail. This requires connection to users' mailbox on their PC via a synchronization program and so in terms of difficulty is similar to downloading the program via the HotSync program. Downloading a program is also possible through a browser (either on-line or off-line browser), such as AvantGo.

Communication and Networking

Palm OS supports two types of data communication: *serial/USB* and *infrared*. The serial/USB communication is used to synchronize the Palm via a HotSync or be used to transmit or remove data through a serial bus. The Palm is equipped with an infrared transceiver and can *beam* data (such as contact details) to another Palm using appropriate communication protocols. This ability has been used to enable a Palm to remotely control another device or to print to an infrared printer.

A TCP/IP connection can be established when a Palm connects to the Internet via a modem or a cradle with a software that supports the PPP protocol. Palm OS provides two libraries for TCP/IP networks which are known as Palm OS Net Library and Internet Library.

Wireless Portal

Many handheld devices can only access Web pages that have specially formatted data to download. Webclipping and WAP are the two commonly used technologies that meet the limitations of small displays that are not capable of displaying regular multimedia Web sites. Although Web browsers such as ProxyWeb and PalmScape, are available for Palm, the two technologies above are usually preferred due to the size limitation of the device.

Palm.Net offers mobile Internet access for Palm device with ROM version 3.5 and above. This service is available in North America and allows users to connect to the Web through a connectivity gateway, operated by Palm.

WINDOWS CE DEVICES

Windows CE was introduced in 1996 to provide a common operating system for a range of devices including handheld computers, industrial controllers, embedded appliances and wireless communication devices. Windows CE has the same API as Windows 2000 and allows the programmers to use a subset of Win32 API development tools and languages including Visual Basic and Visual C++ targeted for CE devices. There are also special versions of the compiler, known as the Embedded Visual C++ and Embedded Visual Basic, that makes it easier to create programs.

Windows CE is not an on the shelf OS. Rather, a device manufacturer needs to configure the Windows CE operating system for the processor platform of his/her choice. The operating system is usually built into the device's ROM, although sometimes it is put in flash memory.

Windows CE devices support serial and USB communication for one-to-one communication, and network connection which can use one or many devices and more sophisticated protocols such as HTTP and file exchange protocols like remote access service (RAS). To support infrared communication based on IrDA protocol, the IrSock API has been added to the operating system.

The latest version of Windows CE is known as the Windows Mobile 2005 Operating System, which is equipped with GSM/GPRS mobile phone connection together with 802.11a/b/g Wireless LAN and Bluetooth support.

VULNERABILITIES OF PALM OS DEVICES

In the following, we illustrate the known vulnerabilities of handheld devices. Without losing generality, we only illustrate the vulnerabilities against Palm OS devices. Three types of attacks have been reported on Palm.

Eavesdropping Palm Infrared Connection

The Palm OS Infrared is vulnerable to data stealing attack. Two main reported cases include constructing a program called "NotSync" which enables a Palm to *steal* the password from another Palm via its IrDA interface (@stake, online), and an eavesdropping attack (Nutheesing, online).

Denial of Service Attacks via HotSync

Palm HotSync manager is used to synchronize data between a Palm and a desktop PC. The HotSync program has a network mode that allows remote synchronization via a modem, hence adding more flexibility to Palm.

It is shown (Knight, online) that setting HotSync in the Network mode can lead to a denial of service attack by sending a large amount of data followed by a new line which results in the Palm to crash.

Palm Virus and Trojan Horse

Similar to other operating systems, Palm OS is vulnerable to viruses and Trojan Horses. The first Palm OS virus, known as the "Liberty crack" (Harris, online), is written by one of the co-authors of the Liberty program. It is a destructive Palm application, disguised as a free version of the Liberty application, that deletes all of the programs stored on a Palm and then reboots the device. A more comprehensive survey of Palm OS weaknesses can be found elsewhere (Kingpin, 2001).

PROTECTION OF HANDHELD DEVICES VIA PERSONAL FIREWALL

We have created a personal firewall for Pocket PC 2003 handheld device. Our personal firewall will allow an authorized information to pass through the device, but an unauthorized information will not be able to enter the device. The technique that we used is by rewriting the Ethernet driver for Pocket PC 2003 device. We will illustrate our work in the following section.

Background

In such a pervasive connected environment, network security is critical to protect devices and the information they contain. Central to network security is the firewall, which acts as a barrier between an untrusted network and the trusted network (or device). Due to high-profile hacker attacks and proliferation of viruses, firewalls have become a fundamental security tool. At this stage, no security system can ensure with absolute certainty that all of the information will be protected all of the time. Firewalls are one of the most effective security tools that a network administrator can deploy to limit vulnerability.

A firewall is hardware or software that monitors the transmission of packets of digital information that attempt to pass through the perimeter of a network. A firewall can be described as a system for enforcing *access control policy* between two networks. A firewall can deny unauthenticated requests or requests with potential threats, while permitting authenticated requests, thus protecting the internal network. In most cases, firewalls are used to prevent outsiders from accessing an internal network. However, firewalls can also be used to guard one highly sensitive part of a private network against its other parts. Such sensitive parts are for payroll, payment processing, R&D systems, etc.

There are essentially three types of firewalls: (1) packet-filtering firewalls, (2) stateful firewalls, and (3) application-level firewalls. A packet-filtering firewall drops packets based on their source or destination address or ports. For example, it blocks all packets from a site that we do not trust, or blocks all packets to an internal machine that should be inaccessible from an external network. Linux's *ipfwadm* and *ipchains* are examples of this type of firewall. A stateful firewall works at the session layer of the OSI model, or the TCP layer

Figure 1. Relation between the modules

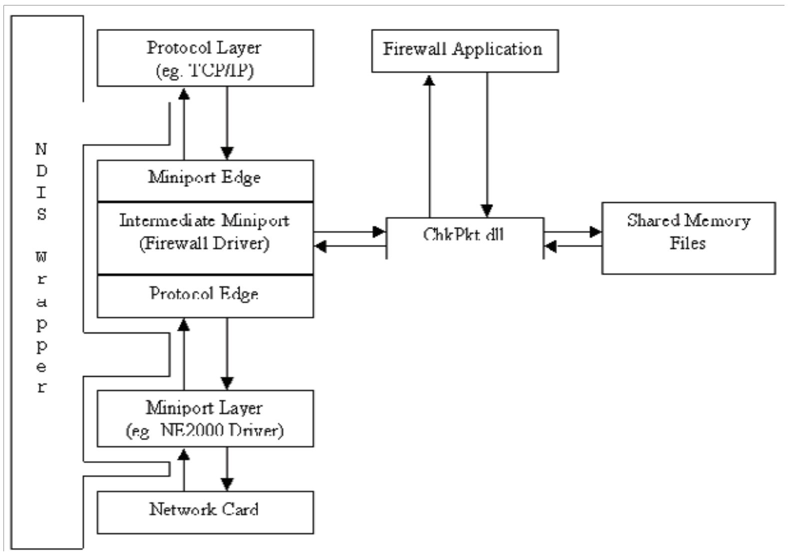
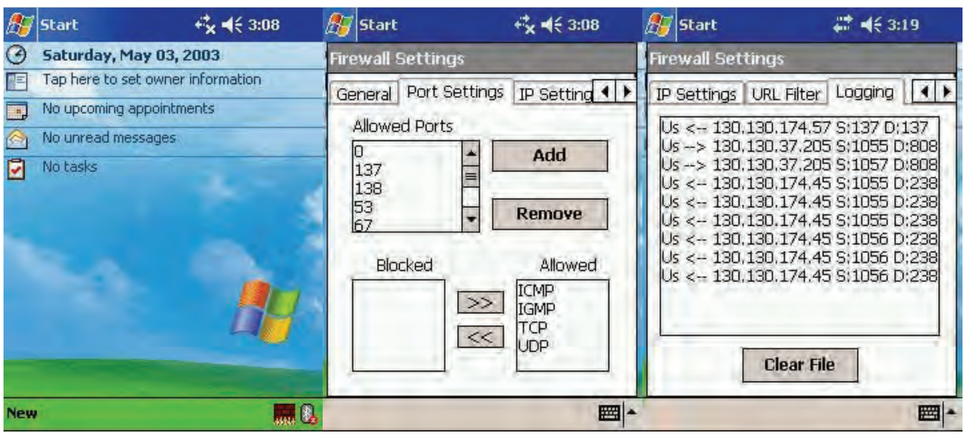


Figure 2. Screen captures of Personal Firewall for Pocket PC 2003 devices



of TCP/IP. It monitors TCP handshaking between packets to determine whether a requested session is allowed. It creates a circuit then relays data between external and internal network. Stateful firewalls keep track of sessions and connections in internal state tables. It can protect against certain types of denial of service attacks. An example of this type of firewall is Checkpoint FW-1. In addition to lower levels, application-level firewalls inspect packets at the application level. A packet reaches the firewall and is passed to an application-specific proxy, which inspects the validity of the packet.

Newer types of firewalls include distributed firewalls and personal firewalls. In distributed firewalls, security policy is centrally defined but is enforced at individual network end points, such as routers, gateways, servers or user PCs. Policy can be distributed to end points in various forms. For example, policy can be pushed to end points where such policy should be enforced. Policy may also be provided to the users as credentials that will be presented when needed.

Unlike traditional firewall sitting between the external and internal networks, a personal firewall runs on a user's computers. Personal firewalls do not have problems such as end-to-end network encryption, malicious insider attacks, single point of failure, etc.

Personal Firewall for Pocket PC 2003 Handheld devices

We have designed and implemented a personal firewall for Pocket PC 2003, which filters TCP/IP traffic passing to and from the device via an Ethernet card. The concept can be extended to allow other protocols such as 802.11b to be covered. Although personal firewalls exist in the market (e.g., Norton Personal Firewall, etc.), their design and implementation are not widely available. In this chapter, we present the first open source firewall for Pocket PC 2003 operating system devices. The personal firewall is a packet-filtering type firewall, which analyzes the contents of network packets. The firewall examines, if applicable:

1. Source and destination IP address
2. Source and destination port
3. Protocol
4. HTTP domain and URL

It establishes the validity of network packets, based upon a strict set of rules. Depending on the packet validity, the firewall either accepts or rejects the network packet. If the firewall encounters a packet that is suspicious or unknown, the user will be notified, depending on the security level, and asked to respond with the appropriate action to be taken.

The design of our firewall consists of three separate modules:

1. `IMDrv.dll` — network device driver
2. `ChkPkt.dll` — dynamic link library
3. `FWApp.exe` — application

The relation between these three modules are illustrated in Figure 1.

We have successfully developed and implemented our personal firewall on Pocket PC 2003 handheld devices. Some of the screen captures of our implementation are shown in Figure 2.

THE NEED FOR MOBILE DEVICES IN HEALTHCARE

In healthcare, mobile healthcare involves conducting healthcare related activities through using mobile devices such as a wireless phone, personal digital assistant (PDA), wireless enabled computer, iPod and so on. Portability and accessibility of these devices enhance their use of them in healthcare environment. Moreover, ease of use, affordability, and ability to read handwriting attracted use of these devices (Al-Ubaydli, 2004). The devices would be suitable for conducting patient's healthcare activities in emergencies, ward rounds, homecare, chronic disease management, conducting clinical trials, and so on.

Mobile computing is suitable for healthcare as healthcare providers are mobile. Healthcare providers would like to enter information in the patient record and process information during the ward round and at the point of care (Reuss, Menozzi, Buchi, Koller, & Krueger, 2004). Health information system located at the fixed terminal is inefficient because the necessary information will not be available at the time of query and that will interrupt the workflow (Eisenstadt et al., 1998).

USAGE OF THESE DEVICES

Mobile and wireless devices can assist in ensuring patient's safety by providing easy availability of the data at the point of care. Computerised physician order entry systems are being developed in most healthcare institutions to reduce medical errors (Ash, Stavri, & Kuperman, 2003). Mobility of mobile devices embrace usage of computerised physician order entry systems (Schuerenberg, 2005) and enhance patient safety. There are various projects using mobile devices to enhance patient's care. With the advancement of medical informatics, telemedicine, and information technology, mobile data devices play an enormous role in healthcare systems. Nursing staffs at St. Clair Hospital in Pittsburgh use PDAs equipped with peripheral scanning technology. This led to identification and prevention of 5000 errors (Anonymous, 2005). A survey and interview conducted on the usage of Personal Digital Assistants by medical residents noted that the current usage is as a single information source for personal organisation and general reference materials (Barrett, Strayer, & Schubart, 2004). However, a study conducted by Fischer, Stewart, Mehta, Wax, and Lapinsky (2003) indicated that these devices have been used for accessing medical literature, as drug information databases (pharmacopoeias), patient tracking, medical education, research data collection, business management, prescribing, speciality-specific applications such as family medicine, paediatrics, pain management, critical care, and cardiology (Fischer et al., 2003).

Two types of use of these devices can be seen in healthcare, which are download on demand (DoD) and downloading data periodically via the synchronization process (Afrin & Daniels, 2001). DoD allows the real time access of data and it is suitable for medical literature, patient tracking and various speciality specific applications. Downloading data periodically is applicable to medical educational application such as iPathology application on the iPod. A study conducted on all the American Academy of Family Physicians and American College of Osteopathic Family Practice in 2000 indicated that most commonly used operating system is the Palm on handheld computer (Criswell & Parchman, 2002). Currently, most healthcare providers use handheld devices and usage increases with emergence of smart phones and blackberry. However, only a few clinical applications are available at this stage for these devices (Al-Ubaydi, 2004).

Data integrity and confidentiality of information in these devices need to be ensured to provide safe, effective and efficient healthcare. Protecting patient confidentiality and privacy during the use of handheld device is a concern in the healthcare industry (Fischer et al. 2003). Implementing personal firewall for these devices enhances security and maintains the data integrity of health information.

CONCLUSION AND FURTHER WORK

In this chapter, we identified handheld devices as important tools for the future. We briefly reviewed the two well-known handheld devices, namely Palm OS-based and Windows CE-based handheld devices. We also illustrated our work to secure these devices. Finally, we concluded with the use of these devices in the area of healthcare. In the future, we will be developing applications that will use handheld devices for specific purpose required in the healthcare area.

REFERENCES

Afrin, L. B., & Daniels, M. S. (2001). PalMER: PalmOS-based access to the enterprise clinical data repository and clinical documentation assistant. *Proceedings of the American Medical*

Informatics Association Symposium 2001 Retrieved from http://www.amia.org/pubs/proceedings/symposia/start.html

Al-Ubaydli, M. (2004), Handheld computers. *British Medical Journal*, *7449*(328), 1181-1184.

Anonymous. (2005). *Provider turns PDAs into scanners, mobile health data and source media*. Retrieved from http://www.mobile healthdata.com/article.cfm?articleeid=1418

Ash, J. S., Stavri, P. J., & Kuperman, G. J. (2003). A consensus statement on considerations for a successful CPOE implementation. *Journal of American Medical Informatics Association, 3*(10), 229-234.

Barrett, J. R., Strayer, S. M., & Schubart, J. R. (2004). Assessing medical residents' usage and perceived needs for personal digital assistants. *International Journal of Medical Informatics, 1*(73), 25-34.

Criswell, D. F., & Parchman, M. L. (2002). Handheld computer use in U.S. family practice residency programs. *Journal of American Medical Informatics Association, 1*(9), 80-86.

Eisenstadt, S. A., Wagner, M. M., Hogan, W. R., Pankaskie, M. C., Tsui, F. C, & Wilbright, W. (1998), Mobile workers in healthcare and their information needs: Are 2-way pagers the answer? *Proceedings of the AMIA Symposium 1998* (pp. 135-139).

envi.com. (online). *Pilot install: Palm pilot installation*. Retrieved from http://envicon.de/e/pinstall/tentacle.html

Fischer, S., Stewart, T. E., Mehta, S., Wax, R., & Lapinsky, S. E. (2003). Handheld computing in medicine. *Journal of the American Medical Informatics Association, 2*(10), 139-149.

Hansmann, U., Merk, L., Nickous, M. S., & Stober, T. (2001). *Pervasive computing handbook*. Berlin: Springer-Verlag.

Harris, D. (online). *The "Liberty" crack: The first Palm OS Trojan Horse*. Retrieved from http://www.sans.org

Kingpin, M. (2001). Security analysis of the Palm OS and its weaknesses against malicious code threats. *Proceedings on the 10th USENIX Security Symposium* (pp. 135-151).

Knight, W. (online). *Palm Pilot open to a denial service attack*. Retrieved from http://www.zdnet.co.uk

Krause, A., Hartl, D., Theis, F., Stangl, M., Gerauer, K. E., & Mehlhorn, A. T. (2004). Mobile decision support for transplantation patient data. *International Journal of Medical Informatics*, *5*(73), 461-464.

Nutheesing, N. (online). *Padlock your Palm*. Retrieved from http://www.forbesbest.com

Reuss, E., Menozzi, M., Buchi, M., Koller, J., & Krueger, H. (2004). Information access at the point of care: What can we learn for designing a mobile CPR system? *International journal of medical informatics, 4*(73), 363-369.

Schuerenberg, B. K. (2005). *Mobility aims to boost CPOE use*. Mobile Health Data and Source Media, May 2005. Retrieved May 2005, from http://www.mobilehealthdata.com/

@Stake. (online). *Palm OS password retrieval and decoding.* Retrieved from http://www.securiteam.com/

Wikipedia, The free encyclopedia. (n.d.). Retrieved from http://en.wikipedia.org/wiki

KEY TERMS

Bluetooth: Bluetooth is an industrial specification for wireless personal area network (PANs). Bluetooth provides a way to connect and exchange information between devices like personal digital assistants (PDAs), mobile phones, laptops, PCs, printers and digital cameras via a secure, low-cost, globally available short range radio frequency (Wikpedia, http://en.wikipedia.org/wiki/).

General Packet Radio Service (GPRS): GPRS is a mobile data service available to users of GSM mobile phones. It is often described as '2.5 G', that is, a technology between the second (2G) and third (3G) generations of mobile telephony (Wikpedia, http://en.wikipedia.org/wiki/).

Global System for Mobile Communications (GSM): GSM is the most popular standard for mobile phones in the world. GSM phones are used by over a billion people across more than 200 countries. The ubiquity of the GSM standard makes international roaming very common with "roaming agreements" between mobile phone operators. GSM differs significantly from its predecessors in that both signalling and speech channels are digital, which means that it is seen as a second generation mobile phone system. This fact has also meant that data communication was built into the system from very early on (Wikpedia, http://en.wikipedia.org/wiki/).

IrDA: IrDA refers to Infrared data association, a standard for communication between devices (such as computers, PDAs and mobile phones) over short distances using infrared signals (Wikpedia, http://en.wikipedia.org/wiki/).

Wireless LAN or WLAN: WLAN is a wireless local area network that uses radiowaves as its carrier: the last link with the users is wireless, to give a network connection to all users in the surrounding area. Areas may range from a single room to an entire campus. The backbone network usually uses cables, with one or more wireless access points connecting the wireless users to the wired network (Wikpedia, http://en.wikipedia.org/wiki/).

This work was previously published in Handbook of Research on Mobile Multimedia, edited by I.K. IBrahim, pp. 534-543, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 7.6
E-Health Security and Privacy

Yingge Wang
Wayne State University, USA

Qiang Cheng
Wayne State University, USA

Jie Cheng
Wayne State University, USA

INTRODUCTION

The widespread and fast-developing information technologies, especially wireless communications and the Internet, have allowed for the realization of greater automation systems than ever in health-care industries: E-health has become an apparent trend, and having a clinic at home or even anywhere at anytime is no longer a dream.

E-health, including telemedicine featured by conducting health-care transactions over the Internet, has been revolutionizing the well-being of human society. Traditionally, common practices in the health-care industry place tremendous burdens on both patients and health-care providers, with heavy loads of paper-based documents and inefficient communications through mail or phone calls. The transmission of medical data is even messy for cases in which patients have to transfer between different health providers. In addition, the medical documents prepared manually are prone to errors and delays, which may lead to serious consequences. The time, energy, and resources wasted in such processes are intolerable and unimaginable in any fast-paced society. For these problems, e-health provides powerful solutions to share and exchange information over the Internet in a timely, easy, and safe manner (Balas et al., 1997).

Incorporating fast and cost-efficient Internet and wireless communication techniques has enabled the substantial development of e-health. The use of the Internet to transmit sensitive medical data, however, leaves the door open to the threats of information misuse either accidentally or maliciously. Health-care industries need be extremely cautious in handling and delivering electronic patient records using computer networks due to the high vulnerabilities of such information. To this extent, security and privacy issues become two of the biggest concerns in developing e-health infrastructures.

BACKGROUND

As early as 1987, Dr. Thomas Ferguson proposed online health care for consumers. In 1993, Dr. Ferguson, together with several other pioneers, initiated the first national conference on e-health (Nelson & Ball, 2004). The efforts laid the very foundation for the early development of e-health. However, e-health did not make a big step until the late 1990s, mainly due to the technical difficulties and high infrastructure cost. The striking development of information technology, particularly that of the computer, Internet, and wireless communications, dramatizes the reemergence of e-business and e-health (Collen, 1999). Thus, significant improvements to a new health-care infrastructure are anticipated so that health care can take place in a ubiquitous and security-assured manner.

The Internet as a fast, open, and cost-efficient way of exchanging information still faces the big challenge of protecting medical information security and privacy. The information transmitted through the Internet could be accessed, altered, deleted, or copied illegally, jeopardizing the patients' security and privacy. The security issues of e-health, in general, represented by all the precautions taken when safely accessing, collecting, and transferring the health information, must be addressed. In fact, the exchange of health information can be made more secure than in a paper-based system when carefully designed with proper security technologies. Information privacy is controlling whether and how personal data can be gathered, stored, processed, or selectively disseminated (Fischer-Hubner, 2001). Medical information may contain some of the most sensitive information about topics such as one's HIV (human immunodeficiency virus) status, emotional and psychiatric care, and abortions. Thus, the privacy of medical information needs to be especially safeguarded.

Ensuring security and privacy in e-health while preserving the fast transaction of medical data is, however, not an easy task. Security by itself is a complicated and tough task to accomplish in every sense, and there seems to be always a balance between the optimum efficiency and cost vs. maximized security (Fischer-Hubner, 2001). Security in e-business has being studied extensively, yet not a single system has been found to meet the requirements of all levels of protection. Healthcare systems need a higher level of protection because medical data are more sensitive and vulnerable to various misuses or attacks (Mac McMillan, 2002). When accessing medical data, possible errors and attacks could occur during the identification, authentication, and authorization processes. Potential threats and dangers incurred by the transmission of e-health data may come from computer viruses such as Trojan horses and droppers, and from intercepting threats such as masquerading, IP (Internet protocol) spoofing, misrouting, information modifying, and packet sniffing. General security mechanisms, which have been widely used at present, consist of the protection of individual servers and applications, firewalls, and secure data channels during transmission.

An early work conducted by the University of California, San Diego, and others in 1996, titled Patient Centered Access to Secure Systems Online (PCASSO), successfully developed a robust security architecture for Internet access (Baker & Masys, 1999). Since then, more efforts have been directed toward developing e-health security measures for virus protection, firewalls, authentication and access control, encryption, and so forth. Many businesses and research organizations have been developing and marketing their techniques and products, for example, ActiveCard Inc., MediTrust, National Health Key Corroborative, and so forth. Current technologies exploit smart cards, digital signatures, biometric devices, digital watermarking, public-key repository infrastructures, privacy-enhancing techniques, and so on (Ball, Chadwick, & Mundy, 2003; Cheng, Wang, & Tan, 2004). We believe that effective solutions to security and privacy in e-health must rely on a unified framework, with the deployment of wide-range security and privacy technologies from various vendors.

E-health security and privacy are challenging not only due to the difficulties of developing an error-free, complex framework, but also because of the complications of various moral and legal issues among all the stakeholders in the e-health industry such as the consumers, vendors, and health providers. To protect security and privacy in health care, governments around the world need to establish necessary regulatory standards. The Health Insurance Portability and Accountability Act (HIPPA) created in 1996 is a standard made by the U.S. federal government to provide the guidelines and policies that protect medical records. HIPAA presents both challenges and opportunities to improve the way in which medical data are acquired, exchanged, and distributed (Hippaadvisory, 2004). Meeting the challenges of HIPAA legislation requires a careful study of the legal infrastructure and a thorough understanding of the HIPAA evaluation process.

CONCEPTS AND REQUIREMENTS

Security in e-health is an integrated concept requiring the confidentiality, accountability, integrity, and availability of medical data. Confidentiality is ensuring that the data are inaccessible to unauthorized users. Accountability is the ability to trace back all the actions and changes made to the data, for example, through security logs used for recording log-ins, dates, accessed content, and changes. Integrity is preventing information from being modified by unauthorized users. Availability is ensuring the readiness of the information when needed. These four features are equally important and need to be satisfied simultaneously, and by working all together, they encompass the concept of security (Schneier, 2000; Stajano, 2002).

Security issues, in general, cover strategies in four different areas: (a) access security including user-authorization identification and management, (b) communication security related to the secure communication of messages, (c) content security including the protection of content such as data confidentiality, integrity, and availability, and (d) security management including security and vulnerability assessments, the implementation of policies, and guidance (Stajano, 2002; Van de Velde & Degoulet, 2003). To address the above issues, a solution to protecting the privacy and security of e-health needs to accomplish at least the following functions.

- Authentication, the process to validate the identity of the user
- Access control, the process to ensure that only authorized users see the authorized content or information
- Encryption, the process to prevent illegal access or use during data communication
- Intrusion detection and theft termination, the process to automatically detect and disable the devices if they are being accessed or attacked illegally

Privacy is the right of an individual to determine the disclosure and use of this personal data on principle at his or her discretion (Fischer-Hubner, 2001). In e-health applications over the open Internet, privacy may be seriously endangered without sufficient protection by privacy legislation and privacy-enhancing technologies (PETs; Goldberg, Wagner, & Brewer, 1997). Besides the confidentiality and integrity of personal data considered by data security techniques, PETs need to protect user identities in terms of anonymity, pseudonymity, unlinkability, and unobservability, and also user identities in terms of the anonymity and pseudonymity of data subjects (Fischer-Hubner, 2001). PETs usually exploit encryption tools and access control, and can be enforced with security policies.

TECHNOLOGIES AND CHALLENGES

Security and privacy issues arise when accessing and communicating the health data. Basically, there are three steps for security after logging into the computerized healthcare system (Van de Velde & Degoulet, 2003).

- Identification to check who the end user is
- Authentication to check whether the information provided during the identification step is correct or not
- Authorization to check whether the user is authorized to perform certain tasks

Authentication and access control are required for security when accessing medical data through an e-health infrastructure. Many techniques have been applied for such proposes, for example, smart cards, firewalls, digital certification, biometric devices, password and PIN (personal identification number) generators, and others (Ferrara, 1998). Encryption can be incorporated into different protocols such as NCSA's Secure HTTP (hypertext transfer protocol; SHTTP) and Microsoft's Private Communications Technology (PCT; Goldman & Rawles, 2004). Theft termination is an effective and active protection strategy (Goldberg et al., 1997), and more studies are needed in developing such functionalities in e-health. In the following, we shall briefly discuss several technologies for e-health security and privacy.

Firewalls

Firewalls are a common practice set on the Internet for a business' Web site to prevent unauthorized access to confidential data. A firewall can act as an isolating layer between the inside and outside networks: All information entering the firewall is filtered or examined to determine whether the users have the rights to access the network, and whether the information entered meets the requirements of the inside network for further dissemination over the network (Gollmann, 1999).

The architecture of the firewall consists of packet filters, proxies, and internal firewalls. Packet filters are programs designed at the port level that determine access by checking the source and destination addresses of the incoming data. Proxies, also known as application-level filters or application gateways, take a step further to exam the validity of the request for the entire set of data. Proxies are also entitled the right of refusing connections based on traffic directions; for example, certain files can be uploaded but not downloaded. To prevent internal attacks from insiders, internal firewalls emerged as an access control mechanism inside the network (Stajano, 2002).

Firewalls have found widespread applications in e-health (and other industries). Some popular products include the Firewall-1 from Checkpoint, Alta Vista Firewall from Compaq, and Gauntlet from Network Associates. However, there are certain limitations to this technique. First, designing firewalls often assumes the threats or attacks are from the outside network, which is not always true in reality. Second, the firewalls cannot resist outside threats at all if we are not able to control or monitor the outside connections. Third, the packet filters cannot prevent hackers from IP spoofing, and the proxies are unable to detect many computer viruses such as Trojan horses or macro viruses. Last but not the least, there is no standard established for firewall architectures, functionalities, or interoperability (Schneier, 2000, Stajano, 2002).

Smart Card

Smart cards are actually one form of token-authentication technology, and they generate the session password that can be authenticated by the server. The whole authentication process involves one or more of the following.

1. The client software, to enter the password and communicate with the server software
2. The server software, to verify the password and record the card history
3. The application software, to integrate the token-authentication technology with other technologies (Fischer-Hubner, 2001)

In general, there are two approaches in token-authentication technology: challenge-response and time-synchronous token authentication. The former works in the following way.

- The server software generates a challenge based on the ID and password that a user entered.
- The smart card generates a response number based on the challenge number and user ID.
- The server checks the response from the smart card. The log-in session is enabled if it matches the correct one.

Time-synchronous token authentication simplifies the challenge-response authentication in that there is no server-to-client step. The user can directly enter the access code from the display of the smart card, which can be matched by the server software since every 60 seconds the smart card and the server software will generate a new access code (Goldman & Rawles, 2004). ActiveCard is an example of using smart cards in e-health services. Its one-time password method is similar to the time-synchronous token-authentication approach, while its device-PIN method incorporates digital certificates or biometric information stored on the card. Graded authentication controls different levels of access rights and needs of the confidential medical data. For example, the users are entitled different levels of rights based on their roles as physicians, clinicians, nurses, technicians, administrators, insurers, and patients. The shortcomings for using smart cards include the fact that the authentication may fail when the physical device degenerates over time, and they are easy to get lost or stolen.

Biometrics

Biometric authentication is the authentication process using the unique physiological and behavioral characteristics of a user such as fingerprints, palm prints, retinal patterns, hand geometry, facial images, voice, or others. This process requires establishing a large database for the user's biometric features as the reference during validation (Jain, Bolle, & Pankanti, 1999). Accuracy and sensitivity are important issues. Usually, a good trade-off between them needs to be achieved because accuracy depends on the ability of the equipment to detect despite small variations between biometric characteristics. Though perfect biometric devices have not yet been developed, adding biometric features to smart cards provides an improved way to achieve security and key management. Such commercial products are gradually replacing the old smart cards in e-health applications. Biometric techniques for security are an active research area, and their applications to e-health industries appear rosy.

An important challenge is how to protect privacy while exploiting biometrics. It is often possible to obtain someone's biometric sample without that person's knowledge, thus, those who desire to remain anonymous in any particular situation can be denied their privacy by biometric recognition. This covert recognition of previously enrolled people may raise privacy concerns over unintended application scopes. Besides this, unintended functional scopes also pose a privacy concern. For example, collectors may obtain or infer additional personal information from biological measurements. Such derived information could lead to discrimination against segments of the population perceived as risky. An effective way to enhance privacy while permitting all the advantages of biometric-based recognition is to decentralize the biometric system. The biometric information is stored in decentralized, encrypted databases over which the individual has complete control. To enhance privacy in using biometric information, legislation by government is necessary. The European Union has legislated against sharing biometric identifiers and personal information. Another solution to the privacy problem includes building autonomous enforcement by independent regulatory organizations like a central biometric authority.

Data Encryption of Health Data

During transmission over the Web, medical data could be intercepted and manipulated maliciously with the information confidentiality and integrity compromised. A common data-protection technique uses data encryption with either symmetric or asymmetric algorithms (Stinson, 2000). The conventional encryption algorithms include Blowfish, CAST-128, IDEA (International Data Encryption Algorithm), Data Encryption Standard (DES), and so forth. Concurrent algorithms include AES (Rijndael), Serpent, RC6, Twofish, and so on (Stinson). For symmetric encryption, the same key is used during both encryption and decryption such as DES developed by IBM. The problem for symmetric encryption lies in the difficulty of key management, where the security process will fail if the third party obtains the key.

For asymmetric encryption, a pair of keys is used including the private key that is only known to the owners, and the public key that others can process. The private key has the highest security level, which cannot be inferred from the public key though they are mathematically related. When the public key decrypts the messages encrypted with the private key, the identity of the sender can be authenticated. This technique is also known as digital signature. When the private key decrypts the messages enciphered with the corresponding public key, the identity of the receiver can be authenticated. Commonly used asymmetric algorithms include RSA (Rivest, Shamir, & Adleman, 1978), elliptic curve cryptosystems (Koblitz, 1997), and so forth. Digital signature is a common technology using asymmetric cryptography to authenticate the author of electronic transmitted documents. The biggest problem with asymmetric encryption is its high computing cost and slow process (Van de Velde & Degoulet, 2003). Thus, asymmetric cryptosystems are usually used for transmitting symmetric keys as it is inefficient for large amounts of data such as medial images.

COMPLIANCE WITH HIPPA

HIPPA has been the standard in the United States for health-care industries, with the guidelines and policies to protect the privacy and security of medical records since it was created in 1996. It includes more than 68 information security conditions in three areas that must be satisfied to ensure compliance with HIPAA. These areas include the following:

1. technical security services such as user authorization and authentication, access control, and encryption;
2. administrative procedures such as formal security planning, record maintenance, and audits; and
3. physical safeguards such as the security of buildings, and the privacy of offices and workstations that handle patient information.

The standard provides a uniform level of protection for all health information, and safeguards physical storage and maintenance, transmission, and access to individual health information. It does not require specific technologies to be used, and solutions will vary from business to business depending on the needs and technologies in place. However, all stakeholders in health-care systems including health plans, healthcare providers, and clearinghouses are required to comply with HIPPA. Such compliance ensures the protection of patient privacy and data security through e-health, and will ultimately increase public confidence in using e-health as their preferred means for healthcare. How can the e-health strategies, technologies, services, and management be deployed to ensure HIPPA compliance? Such questions are faced by the health-care industries in promoting e-health. In sum, HIPAA has a far-reaching effect on the health-care industry, presenting health-care organizations with numerous planning, operational, technical, and financial challenges.

TRENDS

Security mechanisms are usually regarded as technical means to protect privacy. However, they can also contradict privacy. Oftentimes required is the collection of sensitive personal information from the users to better protect security. The user must trust the security system and provide such private information. The government needs to enforce the laws and keep the e-health system running healthily. Technically, one global trend in e-health is the development of privacy-protection mechanisms in e-prescription systems, patient record databases, and medical information systems. The development of security business management also represents an important trend. A healthy security system requires not only the measures to prevent current threats, risks, and attacks, but also reasonable protections against future generations of attacks and misuses. With the rapid evolvement of cyberattacks, excellent security business management skills and strategies are necessary in providing visionary preparation for future threats beforehand, and in designing and managing smart security systems. Because each security technique has its own strengths and weaknesses, an important development is to build a holistic framework, taking advantage of various techniques while compensating for their drawbacks. An industrial standard is lacking but urgently needs to integrate various security techniques so that maximum efficiency can be achieved.

CONCLUSION

E-health, a fast-developing area in healthcare industries, faces the challenges of privacy and security issues. Only with high confidence in security and privacy can both patients and health providers greatly benefit from the high efficiency and quality of service of e-health systems. We discuss relevant concepts, technologies, limitations, challenges, and trends in e-health security and privacy. Also discussed are strategies and standards such as HIPAA. We conclude our article by emphasizing that only through the collective efforts of technologies, businesses, management, and legislative regulation can the security and privacy of e-health be safeguarded.

REFERENCES

Baker, D. B., & Masys, D. R. (1999). PCASSO: A design for secure communication of personal health information via the Internet. *International Journal of Medical Informatics, 54,* 97-104.

Balas, E. A., et al. (1997). Electronic communication with patients: Evaluation of distance medicine technology. *JAMA, 278,* 152-159.

Ball, E., Chadwick, D. W., & Mundy, D. (2003). Patient privacy in electronic prescription transfer. *IEEE Security and Privacy*, 77-80.

Cheng, Q., Wang, Y., & Tan, J. (2004). Integrated selective encryption and data embedding (I-SEE) for medical images case. In *E-health paradigm shift: Perspectives, domains and cases.* Wiley.

Collen, M. F. (1999). The evolution of computer communications. *MD Comput*, *16*(4), 72-76.

Ferrara, F. M. (1998). The standard healthcare information systems architecture and DHE middleware. *International Journal of Medical Informatics*, *52*(1-3), 39-51.

Fischer-Hubner, S. (2001). *IT-security and privacy: Design and use of privacy-enhancing security mechanisms.* Berlin, Germany: Springer.

Goldberg, I., Wagner, D., & Brewer, E. (1997). Privacy enhancing technologies for the Internet. *Proceedings of COMPCON*, San Jose, CA.

Goldman, J. E., & Rawles, P. T. (2004). *Applied data communications: A business-oriented approach* (4th ed.). Wiley.

Gollmann, D. (1999). *Computer security.* John Wiley & Sons.

Hippaadvisory. (2004). *E-health white paper.* Retrieved from http://www.hippaadvisory.com/action/ehealth/

Jain, A. K., Bolle, R., & Pankanti, S. (1999). *Biometrics: Personal identification in networked society.* Norwell, MA: Kluwer.

Koblitz, N. (1997). Elliptic curve cryptosystems. *Mathematics of Computation*, *48*(117), 203-209.

Mac McMillan, M. H. (2002). Cognitive security management: Meeting the challenges of e-health. *HIMSS Proceedings: Educational Serious, 48*, 1-7.

Nelson, R., & Ball, M. J. (2004). *Consumer informatics: Applications and strategies in cyber health care.* New York: Springer-Verlag.

Rivest, R. L., Shamir, A., & Adleman, L. (1978). A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM, 21*(2), 120-126.

Schneier, B. (2000). *Secrets and lies: Digital security in a networked world.* New York: Wiley.

Stajano, F. (2002). *Security for ubiquitous computing.* New York: John Wiley & Sons.

Stinson, D. R. (2000). *Cryptography: Theory and practice.* Boca Raton, FL: CRC Press.

Van de Velde, R., & Degoulet, P. (2003). *Clinical information systems: A component-based approach.* New York: Springer-Verlag.

KEY TERMS

Cryptography: The conversion of data into secret codes for transmission over a public network to prevent unauthorized use.

Data Accountability: The ability to trace back all the actions and changes made to the information.

Data Availability: The ability to ensure the readiness of the information when needed.

Data Confidentiality: The ability to ensure the data is inaccessible to unauthorized users.

Data Encryption: The process of "scrambling" the data for transmission to ensure that it is not intercepted along the way.

Data Integrity: The ability to prevent the information from being modified by unauthorized users.

E-Health: Refers to the market, companies, and initiatives for conducting healthcare-related transactions electronically using the Internet and/or wireless communications.

Firewall: Special software used to prevent unauthorized access to a company's confidential data from the Internet through filtering all network packets entering the firewall at one or more levels of network protocols.

HIPAA (Health Insurance Portability and Accountability Act): Standard guidelines and policies enforced by the U.S. federal government to protect confidential medical records.

Privacy: The right of individuals to control or influence what information related to them can be disclosed, by whom it can be disclosed, and to whom it can be disclosed.

Security: Defined as the combination of processes, procedures, and systems used to attain the confidentiality, accountability, integrity, and availability of the needed information.

Smart Cards: Used in token-authentication systems. It can be either a hardware-based card or smart IDs the size of about a credit card with or without a numeric keypad.

Telemedicine: Refers to the delivery of health care at a distance using telecommunications or the Internet.

This work was previously published in Encyclopedia of E-Commerce, E-government, and Mobile Commerce, edited by M. Khosrow-Pour, pp. 385-390, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.7
Security in E-Health Applications

Snezana Sucurovic
Institute Mihailo Pupin, Serbia

ABSTRACT

This chapter presents security solutions in integrated patient-centric Web-based health-care information systems, also known as electronic healthcare record (EHCR). Security solutions in several projects have been presented and in particular a solution for EHCR integration from scratch. Implementations of Public key infrastructure, privilege management infrastructure, role based access control and rule based access control in EHCR have been presented. Regarding EHCR integration from scratch architecture and security have been proposed and discussed. This integration is particularly suitable for developing countries with wide spread Internet while at the same time the integration of heterogeneous systems is not needed. The chapter aims at contributing to initiatives for implementation of national and transnational EHCR in security aspect.

INTRODUCTION

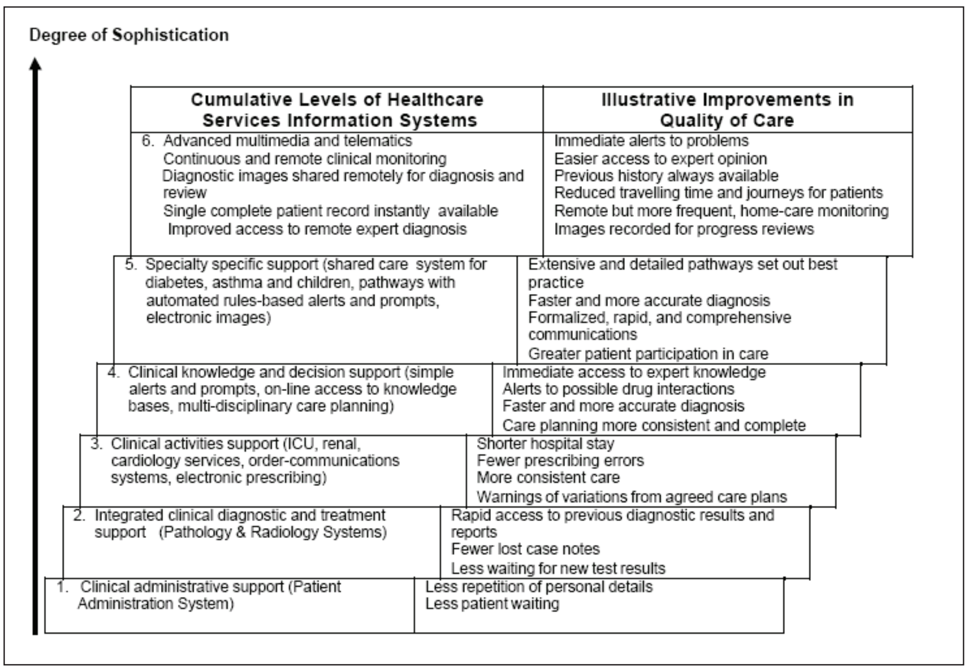
E-health has become the preferred term for healthcare services available through the Internet. While the first generation of e-health applications comprises educational and informational Web sites, at present e-health has grown into national and transnational patient centric healthcare record processing. A patient centric healthcare record, also called electronic healthcare record (EHCR) and electronic patient record (EPR), enables a physician to access a patient record from any place with Internet connection and give a new face to integration of patient data. Such integration can improve healthcare treatment and reduce the cost of services to a large extent. Benefits are based on extended possibilities for collaboration through sharing data between a physician and a patient and between physicians. In such large scale information systems, which spread over different domains, standardization is highly required. The second paragraph describes the main issues in e-health security as well as the results of EU projects EUROMED and TRUSTHEALTH, while the third paragraph presents MEDIS prototype of national healthcare electronic record suitable especially for developing countries where the Internet is widespread and healthcare information systems are not developed to large extent and therefore integration from scratch is proposed.

EXISTING SOLUTIONS

In general, the following lines of development for healthcare information system were considered as important (Reichertz, 2006): (1) the shift from paper-based to computer-based processing and storage, as well as the increase of data in health care settings; (2) the shift from institution-centered departmental and, later, hospital information systems towards regional and global HIS; (3) the inclusion of patients and health consumers as HIS users, besides health care professionals and administrators; (4) the use of HIS data not only for patient care and administrative purposes, but also for health care planning as well as clinical and epidemiological research; (5) the shift from focusing mainly on technical HIS problems to those of change management as well as of strategic information management; (6) the shift from mainly alpha-numeric data in HIS to images and now also to data on the molecular level; (7) the steady increase of new technologies to be included, now starting to include ubiquitous computing environments and sensor-based technologies for health monitoring.

As consequences for HIS in the future, the need for institutional, national, and international HIS-strategies is first seen; second, the need to explore new (transinstitutional) HIS architectural styles is needed; third, the need for education in

Figure 1. Degree of sophistication in healthcare information systems. Note. From Information Systems, Sao Paolo University Technical Report, 2006)



health informatics and/or biomedical informatics, including appropriate knowledge and skills on HIS are needed. As these new HIS are urgently needed for reorganizing health care in an aging society, as last consequence the need for research around HIS is seen. Research should include the development and investigation of appropriate transinstitutional information system architectures, of adequate methods for strategic information management, of methods for modeling and evaluating HIS, the development and investigation of comprehensive electronic patient records, providing appropriate access for health-care professionals as well as for patients (e.g., including home care and health monitoring facilities). All these requirements have implications on security issues. See Figure 1 for an example of the degree of sophistication in healthcare information systems.

Security is a very complex issue related to legal, ethical, physical, organizational, and technological dimensions defined as security policy. In that context, security addresses human, physical, system, network, data, or other aspects.

Legal Issues

Hipocrate's oath contains the obligation of keeping health data secret as a part of professional ethics «What I may see or hear in the course of the treatment or even outside of the treatment in regard to the life of men, which on no account one must spread abroad, I will keep to myself, holding such things shameful to be spoken about". As far as today's practice is concerned, several countries have adopted acts on medical data privacy protection, and especially on privacy protection of the electronic form of medical data. One among such documents is European Directive on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of such Data of September 25, 1995, intended for privacy protection in data processing systems. Medical data privacy protection is presented in Section 3 of paragraph 2 «Special Processing Categories», whose article 1 states that member countries should forbid the processing of personal data on political attitudes, religious and philosophic beliefs, racial and ethnic origin as well as medical data unless they satisfy particular, precisely specified conditions. For medical data, these conditions are as follows:

- When data processing is performed for purposes of preventive medicine, medical diagnostics, the provision of medical treatment and management of medical protection services where these data are processed by health professionals who are bound to keep professional secrecy by national laws or rules established by competent bodies or by some other persons subject to an equivalent obligation.
- Persons to which these data refer have given an explicit consent to the processing of such data.
- Processing is required for protecting the vital data of the person to which these data pertain or of some other persons, when the person in question is physically or legally incapable of giving a consent.
- Processing of data relating to persons which have committed a criminal act or to persons which may violate safety is performed under the supervision of authorized officials.

Section 4 of paragraph 2 «Information to be submitted to a person» states the conditions under which a person has to be given the information on the processing of that person's private data, and especially the informataion about forwarding these data to a third party.

Section 5 of paragraph 2 of this Document «A subject's right to access his own data» states a person's rights to access his personal data in data processing systems.

Paragraph 3 of this document says that member countries should ensure for each person, which considers that he/she has suffered a loss because of illegal processing of his/her data, to be entitled to a compensation.

Results of EUROMED Project

All European Commision funded e-health projects are in compilance with EU directive. One of first implemented was EUROMED (Katsikas,1998). This projects (started in 1997) examines use of trusted third party (TTP) services in distributed healthcare information systems. A trusted third party (TTP) is an entity which facilitates interactions between two parties who both trust the third party; they use this trust to secure their own interactions. One of TTPs is a certificate authority (CA). CAs are defined in X. 509 standard (ITU-T Standard, 1997).

X.509 standard defines a framework for the authentication service which a directory provides to all interested users. A directory is taken to mean that part of the system which possesses authentic information on system users. A dirctory is implemented as a certificate authority (CA) which issues certificates to users. X.509 defines two authentication levels:

- Simple authentication, which uses a password for identity verification
- Strong authentication, which involves credentials—additional means of identification obtained by cryptographic

Strong authentication is based on an asymmetric cryptosystem involving a pair of keys: a public and a private key. The standard does not prescribe mandatory usage of a particular crypto system (DSA, RSA, etc.) and thus supports modifications in methods to be brought about by the development of cryptography.

Each user should have a unique distinguished name. A naming authority is responsible for assigning a name.

A user is identified by proving that he/she possesses a private key. To be able to verify a private key, a user–partner in the communication process must possess a public key. The public key is available on the directory.

A user should be given a public key from a trusted source. Such a source is a CA which uses its own public key to certify a user's public key and produces a certificate in this way. A certificate has the following properties:

- Each user having the access to a CA's public key can disclose the public key on which a certificate has been created.
- No party, except for the CA, can make a modification to a certificate without such a modification being detected. Owing to this property certificates may be stored on a directory with no need for additional protection efforts.

A certificate is obtained by creating a digital signature on a set of information about a user, such as a unique name, a user's public key and additional information about a user. This set of information also contains a certificate's validity period. This period includes the interval during which the CA has to keep the information about certificate status, i.e., publish an eventual certificate revocation. A certificate is presented in the ASN.1 notation as Shown in Box 1.

Three types of strong authentication are described in the standard:

a. **One-way authentication:** Includes only one transfer from a user A to an intended user B
b. **Two-way authentication:** Includes a reply from B to A as well
c. **Three-way authentication:** Includes an additional transfer from A to B

An example of using two-way authentication is given in CEN ENV 13729 standard which prescribes the use of strong authentication in health information systems.

Box 1.

```
Certificate                    ::= SIGNED {SEQUENCE{
        version                [0]          Version DEFAULT v1,
        serialNumber                        CertificateSerial Number,
        signature                           AlgorithmIdentifier,
        issuer                              Name,
        validity                            Validity,
        subject                             name,
        subjectPublicKeyInfo                SubjectPublicKeyInfo,
        issuerUniqueIdentifier [1]          IMPLICIT UniqueIdentifier OPTIONAL,
        subjectUniqueIdentifier [2]         IMPLICIT UniqueIdentifier OPTIONAL
        extensions             [3]          Extensions OPTIONAL   }}

Validity                       ::=          SEQUENCE {
        notBefore                           Time,
        notAfter                            Time
}

SubjectPublicKeyInfo           ::=          SEQUENCE {
        algorithm                           AlgorithmIdentifier,
        subjectPublicKey                    BIT STRING
}
Extensions                     ::=          SEQUENCE OF Extension
```

CEN ENV 13729 has defined local and remote two way strong authentication using X509 standard.

EUROMED-ETS provides integrity, authentication and confidentiality services using measures such as:

- Digital signatures to ensure data integrity
- Encryption to provide confidentiality

TTP sites were established in four different locations in Europe: Institute of Computer and Communication Systems - ICCS (Athens-Greece), University Hospital Magdeburg - UHM (Magdeburg-Germany), University of the Aegean - UoA (Samos-Greece), and University of Calabria - UniCAL (Calabria-Italy).

Among the functions performed by the Certification Authority are: initialisation, electronic registration, authentication, key generation and distribution, key personalisation, certificate generation, certificate directory management,

Figure 2. Challenge-response authentication protocol using X.509 public key certificates

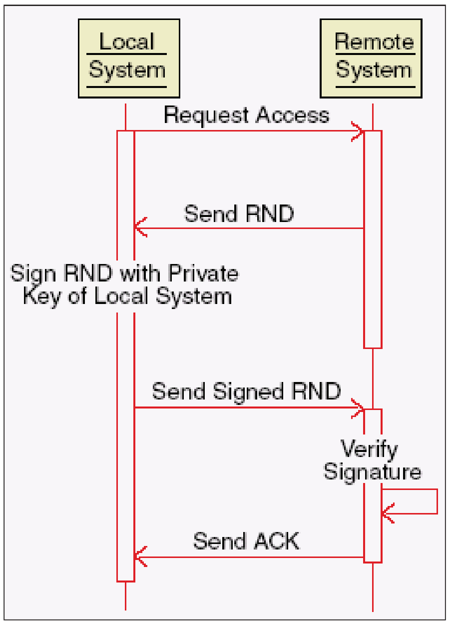
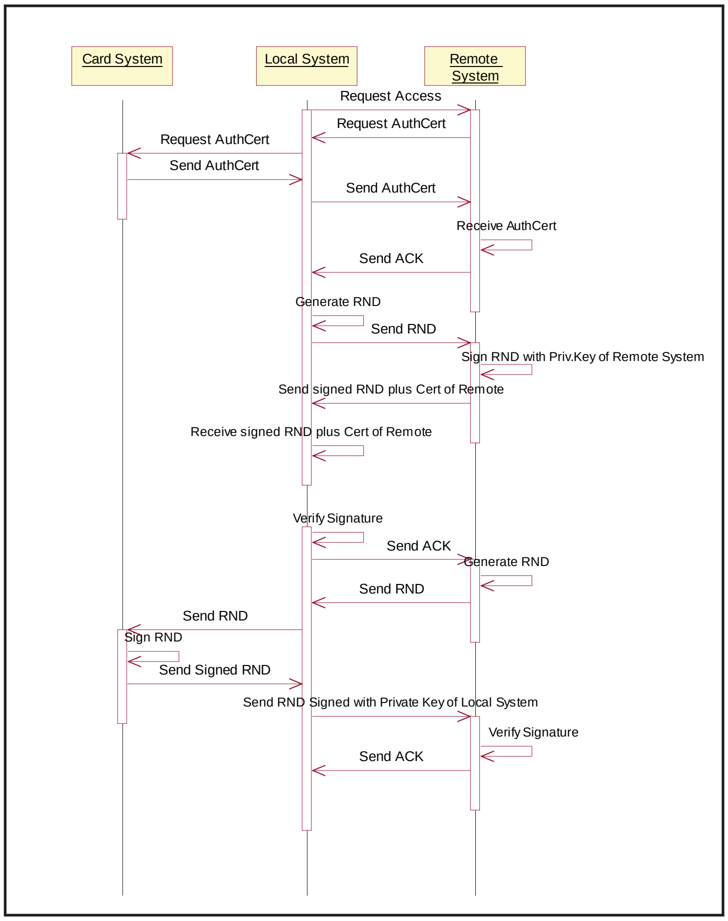


Figure 3. Remote strong authentication according to CEN ENV 13729



certificate revocation, CRL generation, maintenance, distribution storage and retrieval.

The Directories have served in that way as a repository for identification and authentication information; this information was utilised automatically by the EUROMED-ETS pilot Secure Web Servers to identify potential users and grant or deny to them rights; this identification information was also accessible anonymously through the Internet by the use of LDAP search tools.

Results of Trusthealth Project

«Trustworth Health Telematics 2» (Blobel, 2001) project was started in June 1998 under the auspices of the European Commission. The aim of the project is to create a national infrastructure that would provide the security on the communication and application level of health information systems in: Belgium, France, Germany, Norway, Sweden and Great Britain and the focus is on interstate interoperability. The implementation of Trusthealth I Project, which took part from 1996 to 1997 involved the introduction of:

- The use of Health Professional Cards (HPC), microprocessor cards used in the authentication process
- Card reader services
- Services of Trusted Third Party which manages certificates

Within TH-1 the following services have been provided:

- Sending of a physician's report with all relevant data to a central register.
- Execution of prespecified and freely formed SQL queries relating to a patient's EHCR.
- Statistical analyses by various criteria, also by making a SQL query
- Exchange of any information, including HL7 messages, images, etc.

A TTP involves several independent organizations responsible for defining TTP services. TTP components may be:

- Key generation instance
- Naming authority
- Registration authority
- Directory service authority

In the TrustHealth Project the public key registration authority (PKRA) is an entity that identifies, in a unique way, a user requiring the provision of a digital signature service. The professional registration authority (PRA) is an entity registering an individual as a physician, i.e. as a medical care professional. The naming authority (NA) is an entity assigning to users a unique name to be used in certificates. The NA may be involved in assigning unique names to classes within the medical profession (e.g., internal medicine) and subclasses (e.g. nephrology). The public key certification authority (PKCA) is an entity certifying the relation between a user's unique name and public key by issuing a public key certificate with a PKCA digital signature. The PKCA is also responsible for revoking and repeated issuance of a certificate with a public key, whereas the professional certification authority (PCA) certifies the relation between a user's unique name and professional status, after having created a digital signature on these data. The PCA is also reponsible for revoking and repeated issuance of professional certificates. The card issuing system (CIS) is an entity issuing microprocessor cards that must contain a private key and may contain certificates as well. The generation of a private-public key pair may be performed using a local or a central key generator (LKG), an entity that may be located locally (with a user or PKRA) or centrally (with the PKC or CIS). Certificates have to be stored on the certificate directory (DIR). The DIR is an entity that isues, on a request, certificates with a public key, professional certificates, revoked certificate lists as well as other information about users. In the TrustHealth Project the TTP services (NA, PRA, PCA, LKG, PKRA i DIR) are proposed to be implemented in institutions such as the chamber of physicians.

In this project the NA is implemented so as to assign a unique name to a user by using the name of a state, a unique number assigned to each physician on a state level and a physician's name. The RA is implemented so as to certify a user's identity and attributes such as profession or qualification. On a user's request, relevant information is verified, associated to a distinguish name –DN and sent online to a certification entity. The RA uses the information provided by the qualification authentication authority (QAA) or by the profession authentication authority (PAA). The former instance may be a university, for example, while the latter may be a chamber of physicians. In the part of TrustHealth Project implemented in Magdeburg, the chamber of physicians of the state of Saxony-Anhalt has implemented a majority of TTP services. The chamber of physicians has also included QAA services such as qualification, specialization, etc. All these pieces of information have been transmitted online to a certification body.

Based on data obtained from the RA, the Certification Authority creates certificates. Certificates that associate a user's distinguish name and the remaining relevant information to a user's public key are referred to as public-key certificates. Certificates that associate information about profession, qualification are attribute certificates. The first service is provided by the CA and the second by the PCA. CA Management Toolkit from the SECUDE package is used for X509v3 certificate creation and management in the TrustHealth Project.

The DIR directory service includes the publication and revocation of certificates using public directories. An X.500 compatible solution implemented in the SECUDE package is used in the TrustHealth Project. In this Project DIR maintains both public-key and attribute certificates. It is planned to use the Lightweight directory access protocol (LDAP) server later.

Results of PCASSO Project

The Patient Centered Access to Secure Systems Online (http://medicine.ucsd.edu/pcasso/index.html) was developed in 1997-2000 at the University of California San Diego School of Medicine. It is intended primarily to permit patients and health care providers to access health information, including sensitive health data. Access control is achieved by combining role-based access control (RBAC), mandatory access control (MAC) and discretionary access control (DAC). PCASSO is patient-centered and all data are stored on a single server in the current project stage.

According to DAC, when a user requests accessing an object, it is checked whether there is a rule allowing that user to access that object in a given mode. If there is, access is allowed, otherwise it is forbidden. Such an approach is viewed as very flexible and has found a wide usage, especially in commercial and industrial environments. Its shortcoming is the lack of information flow control. It is thought that it is easy to avoid access restrictions imposed by authorization (a set of rules stating which subject is allowed to access a particular object and in which mode). For example, when a user has read some data once, he/she may forward them to an unauthorized user without data owner's knowledge. In contrast to this, information flow from a higher-level object to a lower-level one is prevented in the MAC approach.

In MAC access rights are based on the classification of subjects and objects in the system. A particular protection level is assigned to each subject and each object. The protection level assigned to an object reflects data sensitivity level. The protection level assigned to a subject reflects the level of confidence in that subject that it will not forward accessed information to persons that do not have such rights. These levels are arranged in a hierarchy where each protection level dominates lower levels. A subject has the right of access to an object only if there is a particular relation between the protection level belonging to that subject and the protetcion level belonging to that object. One among such relations is the following: the protection level belonging to the subject has to dominate the protection level belonging to the object. MAC is used in defense and governmental departments.

Some researchers have expressed a view that DAC and MAC approaches cannot satisfy many practical requirements. The MAC approach is suitable for a military environment, whereas DAC is suitable for communities where cooperative work predominates, such as academic institutions. This is why a number of alternatives have been offered. Role based access control (RBAC) is the most widely used among them. RBAC controls a user's access right on the basis of user's activities performed in the system. A role may be defined as a set of activities and responsibilities relating to a particular activity. The advantages of RBAC approach include:

- Simpler authorization control. Authorization specification is divided into two stages: assigning roles to users and assigning object access rights to roles.
- Role hierarchy is easy to create, which is suitable for many systems

- Roles permit a user to work with a minimal-privilege role and use only exceptionally a role having maximal privileges. Error occurrence possibilities are reduced in this way.
- Separation of duties. It is possible to provide that not a single person can autonomously abuse the system. An example is the introduction of ... : each person performs only a portion of an operation instead of the entire operation.

In PCASSO users may have one of the following roles: patient, primary care provider (PCP), secondary care provider (SCP) or Emergency Caregiver. Information and functionalities available to a user depend on the role belonging to her/him. PCASSO employs the following security policy:

- The system controls all accesses to data for each single user.
- Primary care providers are allowed to access all parts of a patient's EHCR.
- PCP is privileged to mark some data in an EHCR as accessible to or forbidden for a patient or other care providers.
- A patient is allowed to access all parts of an EHCR except for those marked as "patient deniable"
- Care providers marked as PCP may change protection attributes in a patient's record. PCPs may authorize and give rights to consultants referred to as Secondary care providers. A patient's PCP may declare a SCP to be a PCP for a particular time period, after the expiration of which a previous role is resumed.
- A possibility is given to care providers to deny access to a part of a child's EHCR to parents.
- In emergency cases care providers may have an unlimited access (reading only) to a patient's EHCR.

PCASSO distinguishes 5 patient data protection levels: Patient-deniable, Parent/Guardian-deniable, Public-deniable, Standard and Low. A user will be allowed to access data having the same or a lower label (protection level) compared with the user's label (MAC approach) and belong, at the same time, to the group having the right of access to that piece of data (DAC approach).

EHCR ARCHITECTURE - CEN ENV 13 606 STANDARD

All EU transnational projects are in compliance with CEN ENV 13 606 standard.

The Comité Européen de Normalisation European Standard (CEN ENV 13606, 2002) "EHCR Communication" is a high level template which provides a set of design decisions which can be used by system vendors to develop specific implementations for their customers.

It contains several parts:

- **Part 1. Extended Architecture:** Defines component-based EHCR reference architecture.
- **Part 2. Domain Term List:** Defines terms which are used in extended architecture.
- **Part 3. Distribution Rules:** Defines data structures which are used in distribution and shared access to EHCR.

Communication as an act of imparting or exchanging information is the primary concern of this standard. In its Part 1 the standard defines an EHCR Communication View as the reuse of stored clinical data in a different context. There can be many such Communication Views and they provide presentation of information in a chronological order , "problem-oriented" manner or some other convention. This is provided by use of architectural components which are rich enough to be able to communicate data by a combination of components. There are a root component, which contains basic information about a patient, on one hand, and, on the other hand, a record component established by original component complexes (OCCs), selected component complexes (SCC), data items (DI) and link items (LI). An OCC comprises (according to data homogeneity) four basic components: folders, compositions, headed sections and clusters. A SCC contains a collection of data representing an aggregation of other record components that is not determined by the time or situation in which they were originally added to the EHCR. It may contain a reference to a set of search criteria, a procedure or some other query device whereby its members are generated dynamically (for example "current medication"). A Link Item is a component that provides a means of associating two other instances of architectural component and specifying the relationship between them ("caused by", for example). A data item is a record component that represents the smallest structural unit into which the content of the EHCR can be broken down without losing its meaning.

As a result of cooperation of CEN Technical Committee 251 and Australian Good Electronic Health Record (GEHR) project, CEN ENV 13606 Part 1 was revised in 2002. The revised standard adopted the GEHR concept in which object-oriented EHCR architecture is distinguished from a knowledge model. A knowledge model contains specifications of clinical structures named archetypes. There are many benefits of that two-layer model and one is that archetypes can be developed by clinicians at the same time when IT specialists develop EHCR object oriented architecture. In the revised CEN standard architectural components contain an identifier of archetype (for example "vitals", "blood pressure" etc.).

In part 2 CEN ENV 13 606 standard defines a list of terms, such as category names for Compositions ("Notes on Consultations", "Clinical Care Referrals" etc.) and Headed Sections ("Former Patient History", "Ongoing Problems & Lifestyle" etc.).

According to part 3 of CEN ENV 13606 standard, each Architectural Component has a reference to a Distribution Rule. A Distribution Rule comprises When, Where, Why, Who and How classes. Class Why is mandatory, i.e. one of its attributes has to be "not null". Instances of these classes define When, Where, Why, Who and How is allowed to access that component (see Figure 4).

IMPLEMENTING SECURE DISTRIBUTED EHCR: MEDIS EXAMPLE

The MEDIS project aims at developing a prototype secure national healthcare information system. Since clinical information systems in Serbia

Figure 4. Distribution rule (CEN ENV 13606 Part 3)

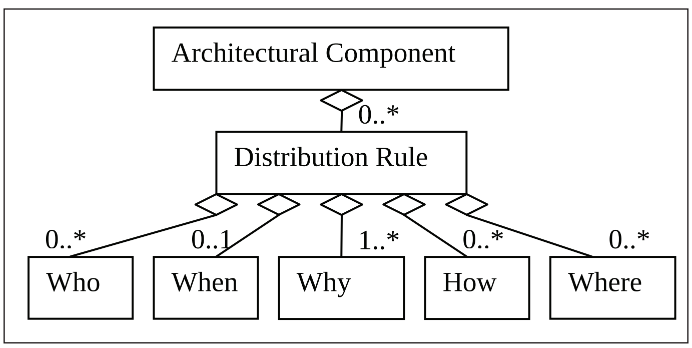


Table 1. Projects' characteristics

	EUROMED	TRUSTHEALTH	PCASSO	MEDIS
Internet based architecture	X	X	X	X
Component based architecture	X	X		X
Public key infrastructure	X	X	X	X
Privilege management infrastructure		X		X
Web service security				X
Role based access control	X	X	X	X
Rule based access control				X

and Montenegro have not been implemented to a large extent, we have focused our efforts on integration itself from the very beginning, instead of on studying how to integrate various systems. In recent years Internet has become widespread in our country and using Internet to implement a shared care paradigm is becoming a reality. MEDIS is based on CEN ENV 13606 standard and follows a component-based software paradigm in both EHCR architecture and software implementation. MEDIS has been implemented as a federated system where the central server hosts basic EHCR information and clinical servers contain their own part of patients' EHCR. CEN ENV 13 606 requirements have been strictly fulfilled in clinical servers as well as in the central server. In our opinion the user interface has to be standardised and we give our proposal for standardisation. As for the security aspect, MEDIS implements achievements from recent years, such as Public Key Infrastructure and privilege management infrastructure, SSL and Web Service security as well as pluggable, XML based access control policies. Table 1. presents characteristics of EUROMED, TRUSTHEALTH and MEDIS projects.

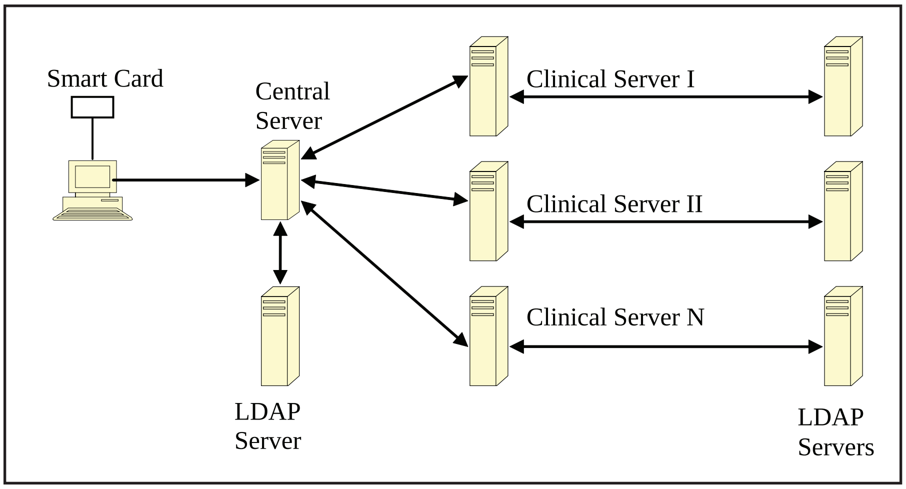
MEDIS Architecture

Since the MEDIS project refers to integration from the very beginning, EHCR reference architecture has been followed in defining the database model. Data have been stored in a hierarchical manner, where architectural components contain pointers to a supercomponent, linked components and also a selected component complex.

MEDIS has been implemented as a federated system. Architectural components are created in compliance with CEN ENV 13606 and stored there where they are created – at hospitals and clinics and are accessed via a central server which contains a root component and the addresses of the clinical and hospital servers. Architectural components that are hosted on the clinical and hospital servers have pointers to supercomponents and linked components (see Figure 5). User interface has been standardised in the following way. HTML pages are created on the central server and contain five frames: the required architectural component (AC) in the right frame, links to subcomponents and linked components in the upper left frame, links to selected component complex (actually distributed queries) in the lower left frame, the AC position in the hierarchical structure of EHCR in the upper frame and information about a user in the lower frame (see Figure 6). A physician can define the position in EHCR (and therefore HTML page) which will appear when he requests EHCR for a patient.

Currently, in the MEDIS architecture there are two types of selected component complex. Firstly, there are SCCs given as a union of queries

Figure 5. System architecture



on all clinical servers such as «current medication» or «current diagnosis». Search criteria are made according to the identifier of archetype. Secondly, there are SCCs related to architectural components on clinical servers and they contain search criteria for architectural components on that clinical server.

In MEDIS there is an authentication applet (Sucurovic, 2005) which is processed in a browser and, after successful authentication, a HTML page has been generated using JSPs on the central server. The clinical servers tier has been implemented in Java Web Services technology using Apache

Figure 6. CEN ENV 13 606 Composition Component example

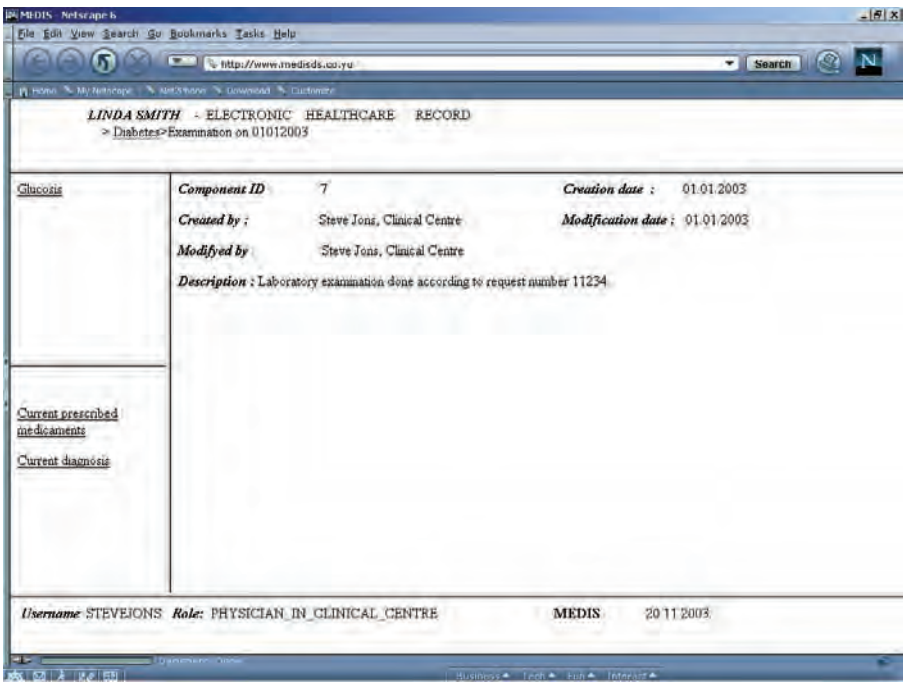
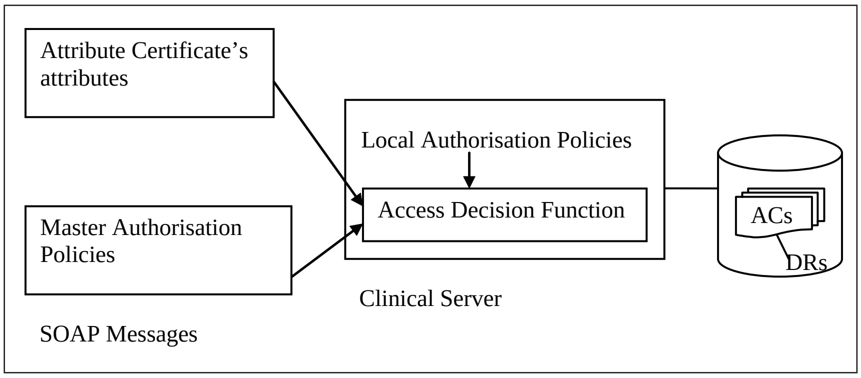


Figure 7. MEDIS access control components



Axis Web Service server and Tomcat Web Server. Business logic has been implemented in reusable components–Java Beans. Authorisation policy is decomposed into components which can be plugged in Beznosov (2004) (see Figure 7).

Implementing Security

Using a password as a means of verification of claimed identity has many disadvantages in distributed systems. Therefore, MEDIS implements CEN ENV 13 729 (CEN ENV, 2000) which defines authentication as a challenge-response procedure using X.509 public key certificates. Solutions related to attribute certificates management have also been implemented (Sucurovic, 2006). Originally, X.509 certificates were meant to provide nonforgeable evidence of a person's identity. Consequently, X.509 certificates contain information about certificate owners, such as their name and public key, signed by a certificate authority (CA). However, it quickly became evident that in many situations, information about a person's privileges or attributes can be much more important than that of their identity. Therefore, in the fourth edition of the X.509 Standard (2000), the definition of an attribute certificate was introduced to distinguish it from public-key certificates from previous versions of the X.509 Standard (ITU-T, 2000).

In the MEDIS project X.509 PKCs are supposed to be generated by the public certificate authority, while ACs are supposed to be generated by the MEDIS attribute authority. The public key certificates are transferred to users and stored in a browser (Sucurovic, 2005). The attribute certificates are stored on LDAP server (Sucurovic, 2006), because they are supposed to be under control of the MEDIS access control administrator. In the MEDIS approach, attribute certificates contain user's attributes as XML text. There are two types of public key and attribute certificates: the Clinicians' and a Patients' as distribution rules contain a flag which denotes if the architectural component is allowed to be read by a patient. Patients will be granted reading the architectural component if physicians set a corresponding flag.

Access Control

In a complex distributed system, such as MEDIS, access control is consequently very complex and has to satisfy both a fine grained access control and administrative simplicity. This can be realised using plugable, component based authorisation policies (Beznosov, 2004). An authorisation policy is the complex of legal, ethical, social, organisational, psychological, functional and technical implications for trustworthiness of health information system. One common way to express policy definition is an XML shema-data. These schemas should be standardised for interoperatibility purposes (Blobel, 2004). The MEDIS project aims at developing the authorisation policy definitions, using XML schemadata (MEDIS Tech. Report, 2005), which are based on CEN ENV 13 606 Distribution rules attributes (CEN ENV, 2002).

Distribution rules define the attributes of architectural components related to access control (CEN ENV, 2002). A distribution rule comprises Who, When Where, Why and How objects (see Table 2). The object Why is mandatory, i.e. one of its attributes has to be "not null". The attributes and entities contained within the objects Who, When, Where, Why and How shall be processed as ANDs. If, however there is more than one Who, How, Where, When or Why object present in a distribution rule, the occurrences of each of those object types shall be processed as ORs.

The MEDIS project has adopted XML as the language for developing constrained hierarchical role based access control and, at the same time, has its focus on decomposing policy engines into components (Beznosov, 2004; Blobel, 2004; Chadwick et al., 2003; Joshi et al., 2004; Zhou, 2004).

The MEDIS project authorisation policy has several components (MEDIS Technical Report, 2005). First, there is an XML schema of user attributes that corresponds to the attribute certificate attributes. Second, there are distribution rules attached to each architectural component (see Figure 4). Third, there are local authorisation policies on local LDAP servers. Fourth, there are Master Authorisation Policies on central LDAP server. There are several types of policies:

- Authorisation policy for hierarchy. It defines hierarchies of How, When, Where, Why and Who attributes (hierarchy of roles, professions, regions etc). In that way, a hierarchical RBAC can be implemented, with constraints defined by security attributes (software security, physical security rating etc.) and nonsecurity attributes (profession, specialisation etc.).

Table 2. Distribution Rule objects [1]

Classes	Attributes	Type
Who	Profession Specialization Engaged in care Healthcare agent	String String Boolean Class
Where	Country Legal requirement	String Boolean
When	Episode of care Episode reference	Boolean String
Why	Healthcare process code Healthcare process text Sensitivity class Purpose of use Healthcare party role	String String String Class Class
How	Access method (read, modify) Consent required Signed Encrypted Operating system security rating Physical security rating Software security rating	String Class Boolean Boolean String String String

- Authorisation policy for hierarchy combinations. It defines which combination of, for example, role hierarchy and profession hierarchy is valid.
- Authorisation policy for DRs. It defines which combinations of attributes in a Distribution Rule are allowed.

There is an enable/disable flag, which defines whether the policy is enabling or disabling. There are in fact, two administrators: one on the clinical server and another on the central LDAP server. In that way, this approach provides flexibility and administrative simplicity. Our future work is to explore the best allocation of these policies between the central server and local servers.

Encryption

The MEDIS project implements Web Service security between the clinical and central server and SSL between the central server and a client (Microsoft, IBM, 2004). We use Apache's implementation of the OASIS Web Services Security (WS-Security) specification–Web Service Security for Java (WSS4J) (W3C Recommendation, 2002). WSS4J can secure Web services deployed in most Java Web services environments; however, it has specific support for the Axis Web services framework. WSS4J provides the encryption and digital signing of SOAP messages. The RSA algorithm has been chosen for signing and the TripleDES for encryption. Communication between the browser and Web Server has been encrypted using SSL. Currently, Netscape 6 browser and Tomcat 5.0 Web Server are used and the agreed chiper suite between them is SSL_RSA_WITH_RC4_128_MD5.

FUTURE WORK

If an application has a large number of users and requires a large number of roles and fine-grained access control at the same time, a formal verification of security policy properties is needed. Secondly, MEDIS project is going to develop a powerful search engine based on intensive use of distributed queries with corresponding security questions solved. MEDIS is a medical record and we are planning to make it a multidisciplinary and multiprofessional record. Nowadays, integration in electronic record comprises the integration of previously introduced HIS using a communication protocol, such as HL7, and Web Services. Hospitals and clinics are connected in grids with protocols similar to Web Services SOAP protocol.

CONCLUSION

Regarding the basic requirements of secure communication and secure cooperation in distributed systems based on networks, basic security services are required. These services have to provide identification and authentication, integrity, confidentiality, availability, audit, accountability (including nonrepudiation), and access control. Additionally, infrastructural services such as registration, naming, directory services, certificate handling, or key management are needed. Especially, but not only in health care, value added services protecting human privacy rights are indisputable. This paper presents existing solutions in transnational electronic healthcare records and also gives a proposal for EHCR architecture and EHCR security solutions in developing countries with widespread Internet.

ACKNOWLEDGMENT

The author would like to thank Mrs. Vesna Zivkovic, PhD, Director of Institute Mihailo Pupin, Computer Systems, for her overall support.

REFERENCES

Beznosov, K. (2004). *On the benefits of decomposing policy engines into components*. Third Workshop on Adaptive and Reflect Middleware, Toronto, Canada.

Blobel, B. (2001). The European TrustHealth project experiences with implementing a security infrastructure. *International Journal of Medical Informatics*, *60*, 193-201.

Blobel, B. (2004). Authorisation and access control for electronic health record system. *International Journal of Medical Informatics, 73,* 251-257.

Blobel, B., Hoepner, P., Joop, R., Karnouskos, S., Kleinhuis, G., & Stassinopoulos, G. (2003). Using a privilege management infrastructure for secure Web-based e-health applications. *Computer Communication*, *26*(16), 1863-1872.

Chadwick, D., et. al. (2003, March-April). Role based access control with X.509 attribute certificates. *IEEE Internet Computing*, 62-69.

Commite Europen de Normalisation ENV 13606 Standard. (2002). *Extended architecture*.

Commite Europen de Normalisation ENV 13729 Standard. (2002). *Secure user identification*.

ITU-T Standard X. 509. (1995, October 24). *Information technology. Open systems interconnection-The directory: Public-key and attribute certificate frameworks*. Directive of the European Parliament and of the Council of 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Retrieved January 10, 2006, from http://www.dsv.su.se/jpalme/society/eu-personal-privacy-directive.html

Information Systems and Information Technology Solutions. (n.d.). *Sao Paolo University report.* Retrieved June 10, 2006, from http://www.virtual.epm.br/material/healthcare/B01.pdf

Joshi, J., et. al. (2004, November–December). Access control language for multidomain environments. *IEEE Internet Computing*, 40-50.

MEDIS technical report. (n.d.). Retrieved September 28, 2005, from http://www.imp.bg.ac.yu/dokumenti/MEDISTechnicalReport.doc

Microsoft and IBM white paper. (2005). *Security in Web service world: A proposed architecture and roadmap*. Retrieved September 28, 2005, from http://www-128.ibm.com/developerworks/webservices/library/ws-secmap

Katsikas S., Spinellis D., Iliadis J., Blobel B., 1998, Using trusted third parties for secure telemedical applications over the WWW: The EUROMED -ETS approach. *Intern. Journal of Medical Informatics*, *49*, 59-68

Reichertz, P. (in press). Hospital information systems: Past, present, future. *International Journal of Medical Informatics*.

Sucurovic, S., & Jovanovic, Z. (2005, February). *Java cryptography & X.509 authentication*. San Francisco: Dr. Dobb's Journal.

Sucurovic, S., & Jovanovic Z. (in press). *Java cryptography & attribute certificate management.* San Francisco: Dr. Dobb's Journal.

Wei, Z., & Meinl, Z. (2004). *Implement role based access control with attribute certificates, ICACT 2004*. International Conference on Advanced Communication Technology, Korea.

This work was previously published in Advances in Enterprise Informatino Technology Security, edited by D. Khadraoui and F. Herrmann, pp. 104-119, copyright 2007 by Information Science Publishing (an imprint of IGI Global).

Chapter 7.8
E-Health and Ensuring Quality

Prajesh Chhanabhai
University of Otago, New Zealand

Alec Holt
University of Otago, New Zealand

INTRODUCTION

The Internet is one of the most utilized resources for obtaining information, learning, communication, and as a source of advice. The most sought after advice and information are related with health matters. In the United States, for example, over 16 million people per year visit WebMD (http://my.webmd.com/webmd_today/home/default), an online portal dedicated to providing health information and services (Sass, 2003). Health information on the Internet has grown exponentially, with up to 88 million adults predicted to access medical information online in 2005 (Ansani et al., 2005). This merging of medical knowledge and information knowledge has given birth to e-health.

Despite the growth and application of information and communications technology (ICT) in health care over the last 15 years, e-health is a relatively new concept, with the term being introduced in the year 2000 (Pagliari et al., 2005). Its use has grown exponentially, and as Pagliari et al. (2005) reported, there are over 320,000 publications addressing e-health listed in MEDLINE alone. However, there is still no clear definition of e-health. There have been two international calls, in 2001 and 2004, for a clear and concise definition of e-health, but both failed to produce an internationally acceptable definition. In the same paper, Pagliari et al. (2005) found 24 different definitions, highlighting the fact that this is a gray area. Hence, without a clear and standardized definition, the opportunities to conduct unethical behavior are made easier.

In this article Eysenbach's (2001) definition will be used, as it provides a comprehensive overview of the term e-health. It has also been used, as Eysenbach is regarded as an expert in the area of e-health and consumer informatics. He has defined e-health as:

An emerging field in the intersection of medical informatics, public health and business, referring *to health services and information delivered or enhanced through the Internet and related technologies. In a broader sense, the term characterizes not only a technical development, but also a state-of-mind, a way of thinking, an attitude, and a commitment for networked, global thinking, to improve health care locally, regionally, and worldwide by using information and communication technology.* (Eysenbach, 2001, p. 1)

This phenomenon has led to consumers becoming more educated and aware of their health condition. This advice is not only accessed by people who may be suffering from a health condition, but also those who are healthy and want to remain so. E-health is being used as a tool for preventive and predictive treatment, as well as a means of locating others in a similar situation and sharing various experiences. In most developed countries e-health is seen to be the first step that one will take towards healing themselves, and if not avoiding the doctor, it is a useful second opinion.

However, the quality of the information on medical Web sites is highly variable and thus some users are reluctant to utilize the information given on mainstream sites. Through social engineering, some Web sites exploit the public's weakness for trusting information on the Web. There are numerous cases of online financial transactions in which unsuspecting members of the public have lost large amounts of money as they have been victims of exploitation. As with a financial setting, there is a great potential for good use and misuse for health information. With e-health experiencing rapid growth, it is becoming increasingly important to consider the various ethical issues that are involved with this form of health information. Unlike financial transactions, inaccurate and unethical information on health informing Web sites could lead to greater complications and even death (Theodosiou & Green, 2003).

Medicine and those that practice it have always had ethics as a core component of their field. There should be no difference when the more traditional aspects of medicine are modernized and utilized as e-health. All those involved with the running of online health care sites have to realize that they are running a site that could potentially mean the difference between life and death for those who access it. It is imperative that such Web sites follow codes and guidelines to prevent individuals' personal medical information, including patterns of use and interests, from involuntarily entering the hands of unauthorized people.

This article focuses on a number of privacy issues that are associated with e-health. Among these are concerns about determining the quality of technologically mediated care, ensuring and managing privacy, and allowing freedom of choice. It is well known that the Internet has the potential of exposing the public to unregulated volumes of misleading information on health and illness. This article will give a short summary of the various regulatory bodies that have been set up to try and ensure that any health information that is put up on the Internet is accurate and in no manner misleading. The *Codes of Conduct* proposed by the Health on the Net Foundation, *URAC* guidelines, and the *e-Code of Ethics* are examples of some regulatory ventures. It has to be made clear that medicine is a practice in which the interests of the patient are the priority rather than the exception, and with rules and regulations e-health can be as "safe" as going to one's local general practitioner. With the exponential growth of e-health, the need to determine the safety, security, and ethical behavior in relation to the traditional services is of paramount importance.

BACKGROUND

The Impact of E-Health

The Internet is seen as the primary medium for the expansion of e-health (Maddox, 2002). This has already resulted in a shift away from the traditional health care delivery model. Patients need not rely on their health care provider for information, and the development of telemedicine has removed the geographic limits that have always been present in health care. The advent of electronic health records with decision support systems will enable medical records to be stored in large data warehouses, thus promoting easier disease management. Also the shift in financial and clinical relationships between all parties involved in health care are all a result of the e-health revolution.

The importance of the Internet in the e-health revolution is crucial. The expansion through this medium has produced eight different types of e-health Web sites. According to Sass (2003), these are:

- Internet-based medical education
- medical expert sites for patient management of a complex nature
- general health information sites for laypeople
- specialized medical sites where physicians play an active role
- cyber doctors who give second opinions for tests carried out in the real world
- sites provided by insurers and drug companies informing about various products
- community Web sites where people suffering from similar ailments share experiences and advice
- research Web sites that focus on new treatment techniques and methods

With such a large variety of e-health Web sites, it can be seen why more than 88 million people in America alone utilize the Internet to search for health information (Ansani et al., 2005).

This rapid expansion has led to the development of new challenges that will require a readjustment in current regulatory systems. Two primary concerns that are a direct result of the information explosion are the quality of information and the possibility of misinterpretation and deceit (Fried, Weinreich, Cavalier, & Lester, 2000). Health consumers and health care providers who access the Internet to obtain knowledge and information have to be able to trust the quality of information being presented to them. Also, with the large volume of content available, the chance of manipulating data and putting up incorrect information is a major concern and a real problem (Goldman & Hudson, 2000). It is this element of uncertainty that has proved to be the main barrier for the growth of e-health at an even faster rate (Fried et al., 2000). Apart from the problems with actual content, the major fear concerns the lack of privacy and security (Chhanabhai, 2005).

The fear arises from the inherent nature of the way business is run on the Internet. Currently most online businesses, especially health Web sites, utilize cookies to collect information about people that visit their Web sites. The other method in which information is collected is through the use of online registration forms. It is the collection of this information that consumers fear. As noted by Goldman and Hudson (2000):

The trail of transactional data left behind as individuals use the Internet is a rich source of information about their habits of association speech and commerce. When aggregated, these digital fingerprints could reveal a great deal about an individual's life. (Goldman & Hudson, 2000, p. 1)

As they lack trust, health consumers are reluctant to fully and honestly disclose health information. The *Ethics Survey of Consumer Attitudes About Health Web Sites* (2000) found that 75% of people surveyed are concerned about health Web sites sharing information without their permission. A large proportion of the survey sample stated that they refused to engage in any online health-related activity as a result of their privacy and security concerns. However, 80% of respondents also said that if there was a specific policy regarding their involvement in information sharing, they would be more willing to move towards the e-health system.

Health consumers' concerns are justified by the numerous reports of health information mismanagement within electronic systems. Win (2005) highlights the following reports:

- Kaiser Permanente accidentally sending e-mail responses to the wrong recipients. The e-mail contained sensitive information that was obtained through online activities
- Global-Healthtrax, an online health store, revealing all contact information including credit card numbers of their its customers on its Web site
- The University of Michigan Medical Center mistakenly placing thousands of its patients records for public access on its Web site

These are a few examples of the reasons that the health consumer is reluctant to fully embrace e-health. Despite that, health consumers are visiting online health Web sites for information and confirmation of various diseases and treatment plans. Thus their primary fear arises when they have to share personal and private information. As a consequence, privacy may not be an issue when only visiting Web sites for information. The issues of quality and misinterpretation hence take on a bigger role in this scenario.

Health Web sites may seem to have high-quality information, but because of the nature of the health information, even this may cause unintentional harm to users of that information. Risk and Dzenowagis (2001) report that unlike other information, health information on the Internet is very susceptible to the following problems:

- language and complexity barriers
- inappropriate audience or context
- unavailability of certain services or products in different parts of the world
- difficulty in interpreting scientific data
- accuracy and currency of information
- potential for source bias, source distortion, and self-serving information

The wide-ranging sources of potential problem areas and the nature of the health Web sites makes them vulnerable to numerous ethical problems. The key ethical issues surrounding e-health are privacy and confidentiality, and quality of information (Ansani et al., 2005; Eysenbach, 2001).

Privacy and Confidentiality

Privacy is found to be the major concern with e-health (Westin, 2005). Health consumers feel that placing information on the Internet or on any electronic system makes it easily accessible to unauthorized persons. Unauthorized persons may be people within the particular company to people such as hackers (Goodman & Miller, 2000). These persons may use the obtained information in various ways. The information is valuable to insurance companies, pharmaceutical organizations, and fraudulent individuals. The fear is that health information can be used to blackmail individuals, especially those in high-profile positions (Givens, 1996). For the average person, the fear is that the information may be used by insurance companies to deny coverage and claims or to increase premiums. Potential employers could use the service of hackers to obtain health information and health Web site surfing habits on current or potential employees, and use this information to fire or not even hire a person (Goodman & Miller, 2000). As systems may be insecure and health information is of high value, there is a great potential for unethical behavior.

There are numerous types of people that have tried and will try to obtain health information. Each of these people will use different ways to try to obtain this information. These methods can be grouped into five distinct categories, identified by the Committee on Maintaining Privacy and Security in Health Care Applications of the National Information Infrastructure (National Research Council, 1997):

1. insiders who make “innocent” mistakes and cause accidental disclosures
2. insiders who abuse their record access privileges
3. insiders who knowingly access information for spite or for profit
4. the unauthorized physical intruder
5. vengeful employees and outsiders

These categories are not clear cut, as an intruder may fall into all the categories depending on his or her time period or situation. Outside attackers and vengeful employees make up one of the five categories; the other four categories indicate that the people that one should fear are those on the inside. Unlike other attackers who might be more interested in the actual breaking of the system, the insider who is a trusted individual is after the information (Spitzner, 2003).

Computer systems by their nature are prone to many other problems other than those that involve an intrusive nature. Being a system that is based on software and various pieces of hardware, an EHR system is vulnerable to both software bugs and hardware failures. These failures have the potential of corrupting medical records and thus diminishing the integrity of the system. When the postal, fax, or telephone system fails, there is a clearly evident impact on the message that is meant to be delivered (Anderson, 1996). In an EHR, this corruption may not be as easily noticeable. Examples include altering numbers in a laboratory report or deleting large amounts of important information. Unlike in other systems where an error would mean a financial disadvantage, an error in the health care system, due to badly designed software or poor hardware construction, may have a detrimental effect on both the caregiver and the patient.

Quality of Information

The quality of Internet-based medical information is extremely variable as there is no peer review or mandatory standard for the quality of health information (Edworthy, 2001). Studies by Ansani et al. (2005), Butler and Foster (2003), and Smart and Burling (2001) have found that health Web sites, irrespective of the medical condition, vary widely in their quality of information. These studies found that there were many inconsistencies in the information provided as well as in the various treatment plans proposed. This is to be expected, as many of the Web sites are sponsored by companies selling specific products or are run by individuals who are sharing their own experiences. Health information carries with it opportunities to reap large cash profits, thus many e-health Web sites are profit driven. With monetary value being associated with health information, the potential for health consumer deceit, bias, or distraction is high (Edworthy, 2001). Also, information may not come from credible sources, thus making it potentially harmful to people who may use that information as a benchmark in their health care. The problem lies with basic human nature. When individuals are suffering from an illness, especially a chronic illness, they tend to try non-conventional therapies in order to improve their condition. With the Internet teeming with alternative therapies, whether provided by individuals or a wide variety of groups, the possibility of using potentially harmful advice is very high (Center for Health Information Quality, 2005). It is this drive towards the use of non-traditional therapies that drives the need for models that can assure that good quality and correct information is published on the Internet.

Apart from possible problems that are a direct result of the information on the Internet, other negative effects of e-health affect the doctor-patient relationship (Theodosiou & Green, 2003):

- health consumers may be better informed than their health care provider about medical findings in a specialized area,
- health consumers may have found disturbing information that may not apply to their situation at all,
- health consumers may choose to use alternative therapies that hinder or oppose the mainstream treatment methods, and
- health consumers may feel they are right and refuse to listen to their health care provider.

Despite the possible negative effects on the patient-doctor relationship, the advent of e-health also has an important role in the empowerment of the patient and thus addressing patients as health consumers rather than just patients. This again highlights the importance of ensuring that the standard of information that empowers patients is of the highest quality.

As can be seen, the possibility of unethical effects is limitless. A number of regulatory bodies have come into existence to try and reduce the chances of unethical activity. The next section will give a brief description of the main quality affirming bodies.

FOCUS

Regulating E-Health

Presently there are approximately 98 different rating schemes for medical Web sites (Ansani et al., 2005). With such a large number of schemes, there is no way of ensuring that one standard that can be trusted over another, thus leading to inconsistent information being posted on the Internet. The leading organizations that have established guidelines for online e-health initiatives are:

- E-Health Code of Ethics,
- Health on the Net Code (HONCode),
- Health Internet Ethics (Hi-Ethics), and
- American Accreditation Healthcare Commission (URAC).

The four above-mentioned schemes are recognized as the leaders in developing sustainable guidelines (Wilson, 2002). The Health Insurance Portability and Accountability Act (HIPAA) is legislation that aims to establish greater privacy rights for health consumers. However, it has a strong legal background that may affect ethical considerations. It is not discussed in this article.

E-Health Code of Ethics

Founded in May 2000 by the non-profit organization Internet Healthcare Coalition, the aim of the code is to provide a guiding set of principles aimed directly at e-health stakeholders worldwide. Its goal is:

to ensure that individuals can confidently and with full understanding of known risks realize the potential of the Internet in managing their own health and the health of those in their care. (Internet Healthcare Coalition, 2006)

The code, which can be found at http://www.ihealthcoalition.org/ethics/ethics.html, is based on eight basic principles: candor, honesty, quality, informed consent, privacy, professionalism in online health care, responsible partnering, and accountability.

Its principles are all based on ethical considerations. Unlike other guidelines, it does not focus on any legal considerations. This has allowed the code to be developed in a manner that provides an ethical foundation for both the providers and receivers of information. The strength of the code results from its development process. It was developed by a method similar to that of an open source project. The draft document was created by medical ethicists and persons involved in the area of ethical behavior. This draft document was then published for public comment, and the comments were incorporated into the final version of the code (Mack, 2002).

Despite its strong ethical base, the major downfall of the e-health Code of Ethics is that it places the final burden on the user of the Web site. As there are no real enforcement measures, the final responsibility is on the provider or the reader of the health information as to whether or not they should conform to the code.

Health on the Net Code (HONCode)

Developed by the Geneva-based Health on the Net Foundation in 1996, the HONCode is the earliest quality initiative for e-health. It is available in 17 different languages. Its inherent strength, simplicity, is also its weakness. Like the eCode of Ethics, the HONCode is based on a set of eight principles. These are (Health on the Net Foundation, 2006): authority, complementarity, confidentiality, attribution, justifiability, transparency of authorship, transparency of sponsorship, and honesty in advertising and editorial policy.

This code works by allowing Web sites that conform to all eight principles to display an active HONCode logo that is considered a Web site trustmark. The presence of this logo indicates that the Web site meets the set guide of standards and ethics. However, this is a self-certification system, and even though HON does periodically review the Web sites that apply for its logo, the Foundation still depends on users of the Web site to inform it if the Web site is not conforming to the HONCode of practice (Risk & Dzenowagis, 2001).

Once again the burden of responsibility is placed on the users of the site to inform HON if there are any breaches in the code; only then can that site have its link broken. The complete code of practice can be found at http://www.hon.ch.

Health Internet Ethics (Hi-Ethics)

Launched in May 2000, Hi-Ethics is a coalition of profit-making U.S. health Web sites. It is run by the non-profit organization Hi-Ethics Inc., with the aim to ensure that individual consumers can realize the full benefits of the Internet to improve their health and that of their families. It is based on 14 principles that are intended to increase the quality of information while ensuring that sound ethical principles are maintained (Mack, 2002). Its principles look at (Hi-Ethics Inc., 2006): privacy policies; enhanced privacy protection for health-related personal information; safeguarding consumer privacy in relationships with third parties; disclosure of ownership and financial sponsorship; identifying advertising and health information content sponsored by third parties; promotional offers, rebates, and free items or services; quality of health information content; authorship and accountability; disclosure of source and validation for self-assessment tools; professionalism; qualifications; transparency of interactions; candor and trustworthiness; and disclosure of limitations and mechanism for consumer feedback.

Web sites that conform to the principles are given URAC accreditation (see next section) and are listed on the Hi-Ethics Web site as trusted sites for health information (Kemper, 2001). Each Web site undergoes a third-party accreditation process before it is approved with the Web site trustmark. The problem with this is that it targets only American companies and Web sites, thus covering only a small portion of the Web sites that are available. This closed coalition and the importance of maintaining the currency of its principles are the main drawbacks of this guideline. (Mack, 2002; Kemper 2001).

American Accreditation Healthcare Commission (URAC)

Launched in August 2001, the URAC standards are based on the 14 Hi-Ethics principles. URAC also draws upon the eCode of Ethics in order to develop ethical frameworks for health Web sites. Unlike the eCode of Ethics, which is completely voluntary with no formal compliance assurance mechanism, URAC works in cooperation with Hi-Ethics to act as a third-party verification system. The standards address the following concerns (Mack, 2002): health content editorial process, disclosure of financial relationships, linking to other Web sites, privacy and security, and consumer-complaint mechanisms.

Any Web site that applies for accreditation with URAC undergoes a two-step review process. Initially the Web site is reviewed by a member of the URAC accreditation staff. The staff member travels to physically meet the owners of the site, to ensure that the online site and the people running it conform to their standards (URAC, 2006). Having passed that stage, the Web site is then assessed by the Accreditation and Executive Committees. Any Web sites that fail to comply with the guidelines after being accredited will have their accreditation withdrawn (Mack, 2002). Each Web site that wants to be accredited has to pay a fee that includes the cost of travel for URAC certifiers.

URAC also has a number of issues that have to be addressed. It is currently aimed at American Web sites, thus like the Hi-Ethics initiative, it is very limiting in its target market. Also it requires a substantial number of fee-paying members to make it truly viable. It cannot be considered to be a successful program if it only has a select group of Web sites conforming to its requirements.

There are numerous other guidelines and systems in place, and the report by Risk and Dzenowagis (2001) provides in-depth analyses of each of the different systems. These four mechanisms alone highlight that there is a problem in trying to control the type of health information that is made available on the Internet. The ideal solution would be to regulate the actual Internet itself. This move has been supported by the World Health Organization, which has called for tighter controls on the placing of health information on the Internet. However, this was opposed by Eysenbach (1998) who correctly stated that attempting to control information on the Internet is both unrealistic and undesirable. As the Internet cannot be controlled, the next best thing is to ensure that the guidelines that are available can and will do a good job. They must promote the rapid development of e-health by making available health information that is of high quality and conforms to stringent ethical standards.

CONCLUSION

The Internet was created to produce a medium of communication that is both flexible and under no specific governing body. By its nature, the Internet has grown extremely successful and has produced an environment that is difficult to regulate, as free speech is promoted. It is in this context that e-health has been developed, and its very nature is an adaptation of the nature of the Internet. The health consumer now has the opportunity to control the way their health is treated. This being so, an area has opened up for unscrupulous and unethical behavior. The problem is further heightened as health care itself suffers from unclear privacy rules. Thus e-health enjoys the benefits of both the electronic medium and medical treatment, but also suffers from the weakness of unclear privacy rules.

This article has highlighted a number of examples where e-health has suffered negative consequences resulting from fraudulent activity and incorrect information on the Internet. Quality of Internet health information is of utmost importance, as it can have both beneficial and fatal consequences for the wide audience that it reaches out to. The unclear definition of e-health does not help the cause either. With e-health not clearly defined, it leaves another question open, “What can be defined as good quality health information?” There are numerous standards and guidelines available, but none will work if there is no way of assessing what good quality health information is.

The four models discussed in this article aim to ensure the validity and reliability of e-health information while maintaining a strong ethical framework. The problems with the models are the lack of consistency and the often complex burden that is placed on the user of the Web site to understand and follow the different criteria. Also some of these models involve high costs to ensure accreditation (e.g., URAC), thus making it not viable for smaller and individual-run Web sites to subscribe to such standards. The models are focused on mainstream Web sites as well as mainstream therapies, thus they do not take into account alternative therapies or treatments plans; yet, it is in these alternative methods that the chances for unethical behavior arise.

E-health is unquestionably the future of health care. With e-health, health consumers will be empowered with deep knowledge and self-care strategies, and will be linked to large communities of other health consumers and health care providers. For the health care providers, e-health brings the opportunity to immediately access patients’ records while promoting evidence-based medicine and keeping updated with the most current treatment techniques. However, the full potential of e-health may not be realized, as there is great fear concerning the privacy, security, and unethical behavior connected with the Internet. Currently there are many alarming areas regarding e-health. What is needed is a single regulatory body that will ensure that ethics and privacy are essential elements in the design of e-health initiatives.

Promoters of e-health must remember that e-health is a branch of medicine and thus they should subscribe to the ethical practices that have been established through the Hippocratic Oath. They must remember that "the practice of medicine involves a commitment to service to the patient, not one to economic interests" (Dyer & Thompson, 2001, p. 1).

REFERENCES

Ansani, N.T., Vogt, M., Fedutes-Henderson, B.A., McKaveney, T.P., Webber, R.J., Smith, R.B., Burda, M., Kwoh, C.K., Osial, T.A., & Starz, T. (2005). Quality of arthritis information on the Internet. *American Journal of Health System Pharmacy, 62,* 1184-119l.

Butler, L., & Foster, N.E. (2003). Back pain online: A cross sectional survey of the quality of Web-based information. *Spine, 28*, 395-401.

Center for Health Information Quality. (2005). Retrieved November 31, 2005, from http://www.hfht.org/chiq

Chhanabhai, P. (2005). *EHRs: Fear of breach? The New Zealand public's opinion*. Unpublished Masters Thesis, University of Otago, New Zealand.

Dyer, K.A. (2001). Ethical challenges of medicine and health on the Internet: A review. *Journal of Medical Internet Research, 3*(2). Retrieved December 25, 2005, from http://www.jmir.org/2001/2/e23/

Edworthy, S.M. (2001). Crawling through the Web: What do our patients find? *Journal of Rheumatology, 28*(1), 1-2.

Eysenbach, G. (2000). Recent advances: Consumer health informatics. *British Medical Journal, 320,* 713-16.

Eysenbach, G. (2001). What is e-health? *Journal of Medical Internet Research, 3*(2). Retrieved July 23, 2005, from http://www.jmir.org/2001/2/e20/

Fried, B.M., Weinreich, G., Cavalier, G.M., & Lester K.J. (2000). E-health: Technologic revolution meets regulatory constraint. *Health Affairs, 19*(6). Retrieved December 11, 2005, from http://content.healthaffairs.org/cgi/reprint/19/6/124

Goldman, J., & Hudson, Z. (2000). Virtually exposed: Privacy and e-health. *Health Affairs*, *19*(6). Retrieved December 11, 2005, from http://content.healthaffairs.org/cgi/reprint/19/6/140

Goodman, K.W., & Miller, R.A. (2000). Ethics and health informatics: Users, standards, and outcomes. In E.H. Shortliffe & L.E. Perreault (Eds.), *Medical informatics: Computer applications in health care and biomedicine*. New York: Springer-Verlag.

Grimes-Gruczka, T., & Gratzer, C. (2000). *Ethics survey of consumer attitudes about health Web sites*. Retrieved November 11, 2005, from http://www.chcf.org/documents/consumer/Ethics2ndEdition.pdf

Health on the Net Foundation. (2006). Retrieved January 3, 2006, from http://www.hon.ch/

Hi-Ethics Inc. (2006). Retrieved January 3, 2006, from http://www.hiethics.com/

Internet Healthcare Coalition. (2006). Retrieved January 3, 2006, from http://www.ihealthcoalition.org/

Kemper, D.W. (2001). Hi-Ethics: Tough principles for earning consumer trust. In J. Mack (Ed.), *The new frontier: Exploring e-health ethics* (pp. 145-150). URAC/Internet Healthcare Coalition.

Mack, J. (2002). *Global e-health ethics□access + quality = equity*. Retrieved November 6, 2005, from http://www.wma.net/e/publications/pdf/2001/mack.pdf

Maddox, P.J. (2002). Ethics and the brave new world of e-health. *Online Journal of Issues in Nursing*. Retrieved December 20, 2005, from http://www.nursingworld.org/ojin/ethicol/ethics_10.htm

Pagliari, C., Sloan, D., Gregor, P., Sullivan, F., Detmer, D., Kahan, J.P., Oortwijn, W., & MacGillivray, S. (2005). What is e-health (4): A scoping exercise to map the field. *Journal of Medical Internet Research*, *7*(1). Retrieved December 23, 2005, from http://www.jmir.org/2005/1/e9/

Risk, A., & Dzenowagis, J. (2001). Review of Internet health information quality initiatives. *Journal of Medical Internet Research*, *3*(4). Retrieved August 5, 2005, from http://www.jimir.org/2001/4/e28/

Sass, H.M. (2003). *Ethical issues in e-health and the promotion of health literacy.* Retrieved October 5, 2005, from http://www.ruhr-uni-bochum.de/zme/healthliteracy/publofpro.htm

Sass, H.M., & Zhai, X. (2004). E-health ethics: A yet to be recognised issue in medicine and medical ethics. *Eubios Journal of Asian and International Bioethics*, *14*, 147-148.

Smart, J.M., & Burling, D. (2001). Radiology and the Internet: A systemic review of patient information resources. *Clinical Radiology*, *56*, 867-870.

Theodosiou, L., & Green, J. (2003). Emerging challenges in using health information from the Internet. *Advances in Psychiatric Treatment*, *9*, 387-396.

URAC. (2006). Retrieved January 4, 2006, from http://www.urac.org/

Wilson, P. (2002). How to find the good and avoid the bad or ugly: A guide to tools for rating quality of health information on the Internet. *British Medical Journal*, *324*, 598-600.

Win, K.T. (2005). A review of security of electronic health records. *Health Information Management*, *34*(1), 13-18.

Westin, A.F. (2005). U.S. public sharply divided on privacy risks of electronic medical records. *Proceedings of the Hearing on Privacy and Health Information Technology*, Washington, DC. Retrieved March 17, 2005, from http://www.pandab.org/

KEY TERMS

Consumer Health Informatics: "The branch of medical informatics that analyses consumers needs for information; studies and implements methods of making information accessible to consumers; and models and integrates consumers' preferences into medical information systems" (Eysenbach, 2000).

E-Health: "An emerging field in the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the Internet and related technologies. In a broader sense, the term characterizes not only a technical development, but also a state-of-mind, a way of thinking, an attitude, and a commitment for networked, global thinking, to improve healthcare locally, regionally, and worldwide by using information and communication technology" (Eysenbach, 2001).

Electronic Health Records: A longitudinal collection of health information about individual health consumers and populations which is stored in an electronic format. The stored information will provide a picture of "cradle to the grave" consumer health information.

Health Informatics (Medical Informatics): The emerging discipline in which information technology is applied to the medical field. It involves the development of structures and algorithms to enhance the treatment, management, and prevention of medical conditions.

Informed Consent: The process by which a health consumer is firstly given all the information about various procedures, which will possibly include information on how consumer information will be treated. The health consumer can then make a legally binding decision to carry on with the process or decline taking part.

Medical Ethics: The field that considers the activities, risks, and social effects that affect the field of medicine. It is strongly based on the Hippocratic Oath and uses this as a guideline to determine the ethical conduct of professionals in the field of medicine.

Medical Internet Ethics: "An emerging interdisciplinary field that considers the implications of medical knowledge utilised via the Internet, and attempts to determine the ethical guidelines under which ethical participants will practise online medicine or therapy, conduct online research, engage in medical e-commerce, and contribute to medical Web sites" (Dyer, 2004).

MEDLINE: A comprehensive electronic literature database of life sciences and biomedical information produced by the U.S. National Library of Medicine. It covers the fields of medicine, dentistry, nursing, veterinary medicine, health care administration, and the pre-clinical sciences dating back to 1966.

Telemedicine: The use of telecommunications technology to aid in medical diagnosis and consumer care when the health care provider and health consumer are separated geographically.

This work was previously published in Encyclopedia of Information Ethics and Security, edited by M. Quigley, pp. 170-178, copyright 2007 by Information Science Publishing (an imprint of IGI Global).

Chapter 7.9
Developing Trust Practices for E-Health

Elizabeth Sillence
Northumbria University, UK

Pamela Briggs
Northumbria University, UK

Peter Harris
Sheffield University, UK

Lesley Fishwick
Northumbria University, UK

ABSTRACT

The number of people turning to the Internet to meet their various health needs is rising. As the prevalence of this form of e-health increases, so the issue of trust becomes ever more important. This chapter presents a brief overview of e-health and describes how and why people are using the Internet for health advice and information. In order to understand the trust processes behind this engagement, a staged model of trust is proposed. This model is explored through a series of in-depth qualitative studies and forms the basis for a set of design guidelines for developing trust practices in e-health.

INTRODUCTION

E-health is a term widely used by many academic institutions, professional bodies, and funding organizations. Rarely used before 1999, it has rapidly become a buzzword used to characterize almost everything related to computers and medicine. As its scope has increased, so have the trust issues associated with the term. A systematic review of published definitions identified a wide range of themes but no clear consensus about the meaning of the term e-health, other than the presence of two universal themes (health and technology) (Oh, Rizo, Enkin, & Jadad, 2005). Eysenbach's (2001) commonly cited definition allows a conceptualization that goes beyond simply "Internet medicine."

> *E-health is an emerging field in the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced through the Internet and related technologies.*

Despite an absence of any single definition of the concept, the key themes of health and technology allow a wider exploration of the domain from multiple standpoints. E-health can encompass a wide range of technologies (Internet, interactive television, personal, digital assistants, CD-ROMS), a range of health services and information types (family practitioner's surgeries, public settings, consultations, decision making), and a range of different stakeholders (medical professionals, patients and careers, business).

Sixty two percent of Internet users have gone online in search of health information, and it has been estimated that over 21 million people have been influenced by the information provided therein (Pew, 2000). Given the prevalence and use of Web sites concerned with health, the focus of this chapter is specifically on the use of the Internet for health advice and information. There has been a rapid increase in the use of technology, specifically the Internet, in health information and advice. Worldwide, about 4.5% of all Internet searches are for health related information (Morahan-Martin, 2004). Sieving (1999) lists a number of different push and pull factors that have led to this increase, at least from an American perspective. These include an increasingly elderly population and a change in emphasis from healthcare providers treating illness to patients having primary responsibility for maintaining and improving their own health. Meanwhile, a range of content providers have recognized the educational and financial benefits of providing online health information.

Given the range of health advice and information available and the possible delivery modes available, patients are faced with decisions about which information, providers, and technologies to trust. How do people make decisions about trust in this context and how can guidelines for trust practices help in this respect? This chapter aims to explore the issue of trust within the context of e-health. The rest of this chapter is organized as follows. In the next section, we discuss the role of technology in the context of health information and advice. In section "Trust and Mistrust in E-Health," we examine the context of trust in relation to e-health, and in particular online or Web-based health advice and information, and present a staged model of trust that helps reconcile differences in the literature. "Validating the Model" presents a validation of the staged model through in-depth qualitative work. "Guidelines for Developing Trust in E-Health" draws together the literature and the results of the qualitative work to present a set of guidelines addressing how trust practices can be promoted in e-health and especially Web-based e-health.

THE ROLE OF TECHNOOGY IN THE CONTEXT OF HEALTH INFORMATION AND ADVICE

Technology has traditionally played an important role within health care. Through physician focused systems such as telemedicine and more recently the development of an electronic patient record system within the UK to an increase in patient-focused information and advice channels, accessing information about health can now be achieved via the telephone, the Internet, or even via the television (BBC Online, 2004). The UK's National Health Service is one service provider exploring new ways of disseminating health information to the general public. The NHS Direct service allows patients to access information online, speak to a nurse on the telephone, or retrieve information via a public access kiosk (Dobson, 2003).

Increasingly, the Internet has been seen as a major source of advice and information on health. More and more patients prefer not to let physicians dictate medical care and as such, have turned to alternative sources of advice. An estimated 80% of adult Internet users in the US (Pew Research, 2003) and over two thirds of Internet users in Europe have turned to the Internet for health information and advice (Taylor & Leitman, 2002). People search online for health information and advice for a number of reasons. Some people want to be better informed, better prepared when meeting the doctor, or are searching for support, alternative answers, or reassurance (Rozmovitis & Ziebland, 2004). Over 40% of Europeans said they would like to discuss information found on the Internet with their doctor, and one in five European consumers has asked their doctor about a symptom or diagnosis after having read information online (WHA, 2002). Many people searching online for health advice believe that it will enable them to better deal with their health and will convey health benefits (Mead, Varnam, Rogers, & Roland, 2003).

Time constraints in the consulting room have also led to an increase in online searching. The average length of a doctor's appointment in the UK is currently about 8 minutes. In this short period of time, both the doctor and the patient often find it difficult to explain and discuss all their issues. Patients often find it difficult to recall the specifics of their discussions with the doctor after the consultation (Kalet, Roberts, & Fletcher, 1994). Young people, in particular, are turning to the Internet rather than to a family doctor or a parent to get health information and advice (Kanuga & Rosenfeld, 2004), and the appeal of the Internet is particularly strong for those people who wish advice on important but sensitive matters (Klein & Wilson, 2003). The Internet allows people to communicate and interact with a far greater variety of people across all walks of life. It provides up-to-date information as well as increased social support (Eysenbach, Powell, Englesakis, Rizo, & Stern, 2004). It allows information to be shared in the form of text and images and can put people in touch with the most up-to-date information from some of most eminent sources in the medical profession. The Internet can offer people second and even third opinions and, in short, can provide people with information and advice that they simply cannot find anywhere else.

The advice that health consumers are seeking online ranges from information and advice about treatment decisions, interpreting diagnoses, and understanding any long-term medical implications through to finding support groups, seeking alternative therapies, and understanding drug requirements and drug interactions (Bernhardt & Felter, 2004; Gray, Klein, Noyce, Sesselberg, & Cantrill, 2004; Peterson, Aslani, & Williams, 2003; Reeves, 2001; Ziebland, Chapple, Dumelow, Evans, Prinjha, & Rozmovits, 2004).

But does the Internet leave patients better informed, and does reading online information and following advice lead to better outcomes? Coulson (personal communication), for example, found that people using Internet resources for health topics such as HIV and Aids often took on medical advice that was incorrect. Another key issue is the extent to which users recognize quality advice and information. Despite there being at least 70,000 health related sites available on the Internet (Pagliari & Gregor, 2004) less than half of the medical information available online has been reviewed by the medical profession (Pew Research, 2000), and few sites provide sufficient information to support patient decision making, with many also being heavily jargon laden and difficult to read (Smart & Burling, 2001).

One of the benefits (and the disadvantages) of the Internet is the range of material available regarding health and lifestyle advice. Sites can vary in the type of advice and information they present (see Table 1) and in terms of their scope, health topic, ownership, country of origin, and levels of interactivity. To indicate the range of different sites available, Table 2 provides an overview of three different UK-based sites highlighting their main features and the range of services available on each site.

Given the huge variety of sites available, how are consumers searching for and evaluating the trustworthiness and quality of health Web sites? There is a large body of research assessing the quality of information available on the Internet (Jadad & Gagliardi, 1998; Smart & Burling, 2001; Wyatt, 1997), embracing diverse topics such as Viagra, rheumatoid arthritis, and diabetes. A systematic meta-analysis of health Web site evaluations noted that the most frequently used quality criteria included accuracy, completeness, and readability and design (Eysenbach, Powell, Kuss, & Sa, 2002). Seventy percent of the studies they reviewed concluded that quality is a problem on the Internet.

Almost all of these quality assessments have been conducted from a medical perspective. Relatively few studies have tried to understand the patient experience. Despite forming a cornerstone of NHS IT policy (NHS, 1998) and numerous studies reporting the widespread consumer use of the Internet for health information, relatively few have attempted to understand how different stakeholders access and evaluate the information and the way in which Internet sources affect the patients' experience or health outcomes. Notable exceptions are studies of the way in which patients with HIV or AIDS (Reeves, 2001), cancer (Ziebland et al., 2004), paediatric health information needs (Bernhardt & Felter, 2004), or medicine

Table 1. The range of health Web sites available to consumers

Type of site	Description of site
Web providers and portal sites	Information and advice supplied by Web provider rather than a physical organization. Portals act as catalogues of information providing a gateway to many other sites providing information and advice.
Support groups	Often run by individuals or on behalf of support groups. May be local, national, or global in scale. Often contains forums where consumers can read comments and contribute to discussions.
Charity sites	Registered charity sites provide information and advice on specific health issues and provide a focal point for fund raising activities.
Government Web sites	Provide patient information in the form of news, features, and fact sheets.
Pharmaceutical sites	Sponsored by pharmaceutical companies these sites are often biased in favour of their own drug remedies and regimes.
Sales sites	Sales sites promote and sell certain drugs, medical devices, or health plans, often in addition to some information.
Personal sites	Contains personal experiences of illness and health issues.
Medical databases	Provide access to research papers on health and illness issues.
Media sites	Extensions of print or television media sites that provide the latest news and commentary on health features.
Clinician sites	Information on specific health issues or specialist clinics run by medical professionals.

Table 2. Health Web sites and range of features/services

	Web site		
	Netdoctor http://www.netdoctor.co.uk	**Dipex** http://www.dipex.org	**NHS direct** http://www.nhsdirect.nhs.uk
Domain	General health	High blood pressure	General health
Owner	Web provider	Charity	Government
Type of advice	Patient fact sheets Ask the doctor service News Discussion boards	Personal accounts Medical information Frequently Asked Questions	Health encyclopaedia Self-help guide Clinical evidence-based research reports
Commercial aspects	Health related advertising	Donations requested	None
Notable features	SMS alert service (e.g., contraceptive pill reminder Discussion boards)	Video clips Links to support groups	Locate nearest NHS services facility

requirements (Peterson et al., 2003) use online information to help them make treatment decisions, interpret information, find support groups, and seek alternative therapies. Other studies have suggested that whilst the Internet can increase patients' knowledge about their health conditions, they can be left feeling too overwhelmed by the information available online to be able to make an informed decision about their own health care (Hart, Henwood, &Wyatt, 2004). Reviews of quantitative Web-based intervention studies have generally indicated beneficial effects of the Internet on health outcomes, although issues concerning the methodological quality of such studies remain (Bessell, McDonald, Silagy, Anderson, Hiller, & Sansom, 2002). A more recent meta-analysis by Wantland et al. (Wantland, Portillo, Holzemer, Slaughter, & McGhee, 2004) concluded that in the majority of studies, knowledge and or behavioural outcomes improved for participants using Web-based health interventions.

The paucity of research from the patient perspective is a problem, because we know that ordinary health consumers are likely to adopt different quality criteria to experts, being more readily influenced by the attractiveness of the design, for example (Stanford, Tauber, Fogg, & Marable, 2002). In an empirical study of over 2,500 people who had sought advice online, for example, most began their search via a general information portal—gaining access to information of variable quality indiscriminately (Briggs, Burford, DeAngeli, & Lynch, 2002). Eysenbach and Köhler (2002) noted that consumers (as opposed to experts) failed to check the authorship or owners of the Web site or read disclosure statements, despite suggesting these as important quality markers beforehand. Their study made use of an artificial search task, however, and the authors themselves suggested that people in a "real setting" with a greater stake in the outcome may well pay more attention to the content of the Web sites, in terms of markers of quality. This failure to engage with real Internet consumers in a realistic setting is a common problem in studies of online trust. Thus, there is a real need for systematic explorations of the ways in which people evaluate the trustworthiness of health information and advice online.

TRUST AND MISTRUST IN E-HEALTH

The literature suggests that trust is multifaceted, multidimensional, and not easy to tie down in a single space (Marsh & Dibben, 2003). A key point to note is that trust is intimately associated with risk; indeed, it is possible to argue that in the absence of risk, trust is meaningless (Brien, 1998). Even within seemingly simple acts of trust, complex sets of judgements are invoked and a risk assessment is involved. Indeed, some of the trust models that have been developed in recent years have explicitly included risk (Corritore, Kracher, & Wiedenbeck, 2003). In online advice, there is no doubt that people are more willing to trust a site if perceived risk is low. This came out very clearly in a study of over two and a half thousand people who said they had sought advice online (Briggs et al., 2002). Those that sought advice in relatively high risk domains (e.g., finance) were less likely to trust, and subsequently act on, the advice than those who sought advice in low risk domains (e.g., entertainment).

The literature regarding trust in an e-commerce setting provides a useful starting point for exploring the ways in which people evaluate the trustworthiness of health information and advice online (see Grabner-Krauter & Kaluscha, 2003 for a recent review.) Based on this literature, we can assume that various factors are likely to govern the extent to which individuals feel they can trust health advice online. Firstly, they may be influenced by the look and feel of the site; trusting, for example, those sites rated high in visual appeal and mistrusting those sites with poor visual design or with unprofessional errors. Secondly, they may be influenced by the branding of the site or by presence of familiar images or trusted logos. Thirdly, they may be influenced by the quality of information available on the site, trusting those sites with greater perceived expertise, and fourthly, they may be influenced by the extent to which the advice is personalized to the individual, that is, the extent to which the advice appears to come from and be directed to similar individuals (i.e., those with a shared social identity).

Whilst these different factors appear to be important, researchers disagree over their relative importance in fostering trust. For example, some researchers argue that consumer trust (or a related construct, credibility**)** is primarily driven by an attractive and professional design (Fogg, Kameda, Boyd, Marchall, Sethi, Sockol, & Trowbridge, 2002; Kim & Moon, 1998; Stanford et al., 2002) or is influenced by the presence or absence of visual anchors or prominent features such as a photograph or trust seal (Riegelsberger, Sasse, & McCarthy, 2003). Others argue that trust reflects the perceived competence, integrity predictability, and/or benevolence of the site (Bhattacherjee, 2002; McKnight & Chervany, 2001). A few authors also highlight the importance of personalisation in the formation of trust judgments (Briggs, De Angeli, & Simpson, 2004) or the notion of good relationship management (Egger, 2000).

One way of reconciling these different findings is to consider a developmental model of trust or the way in which trust develops over time. For example, it is worth distinguishing between the kinds of trust that support transient interactions and those that support longer-term relationships (Meyerson, Weick, & Kramer, 1996). Riegelsberger, Sasse and McCarthy (2005), in their model of trust, describe different stages of trust that develop over time: early, medium, and mature forms of trust. A number of authors (Egger 2000, 2001; Sillence, Briggs, Fishwick, & Harris, 2004a) have suggested that three phases are important: a phase of initial trust followed by a more involved exchange that then may or may not lead to a longer-term trusting relationship. If one considers trust in this developmental context, then some of the findings in the literature make more sense. In particular, consideration of a developmental context helps to reconcile the tension between those models of trust, which suggests that it is a concept grounded in careful judgment of institution and process factors such as vendor expertise and experience, process predictability, degree of personalization and communication integrity, and those models that suggest trust decisions depend much more heavily on the attractiveness and professional feel of a site.

A staged model of trust (see Figure 1) makes it possible to distinguish between relatively "hasty" and more "considered" processing strategies for the evaluation of trust in high- and low-risk environments. Chaiken (1980) identified two processing strategies by which an evaluation of trustworthiness may be made: firstly, a heuristic strategy that follows a "cognitive miser" principle, where people base decisions on only the most obvious or apparent information; and secondly, a systematic strategy that involves the detailed processing of message content. A number of other studies in the persuasion literature support the two-process model, namely that people use cognitively intense analytical processing when the task is an important or particularly engaging one, whereas they use affect or other simple heuristics to guide their decisions when they lack the motivation or capacity to think properly about the issues involved. Such different processing strategies also reflect the distinction between the preliminary stage of (1) *intention to trust* and the later stage of (2) *trusting activity* (McKnight & Chervany, 2001).

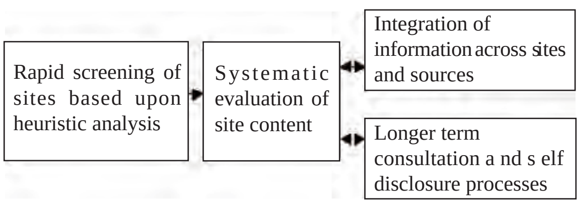
The process does not stop there, however, and a more realistic assessment of the development of trust should include a third stage in which a trusting relationship develops between the consumer and the Web site. This final stage has been rather overlooked in the trust literature, although it was originally proposed in the Cheskin/Sapient report (1999), and also appears in MoTEC (a Model of Trust for E-Commerce) (Egger, 2000, 2001), where the authors described a stage of trust maintenance. There is, however, little empirical evidence for a staged model of trust in e-commerce. Only two of the studies reviewed by Grabner-Krauter and Kaluscha (2003) investigated real transactions and these were limited to short-term interactions. Most published studies of trust do not investigate the act of trusting, but rather investigate the intention to trust (Briggs et al., 2004).

Stage 1: First Impressions and Heuristic Analysis

Some studies have asked participants to briefly visit a site or sites and then offer some evaluation of trustworthiness. According to Chaiken (1980), such participants would have very low involvement in this process, and we would expect that their trust judgements would be based on first impressions and as such, highly influenced by the attractiveness and ease of use of the site. A number of studies support this.

In their studies on credibility (a concept closely related to trust), Fogg et al. (2002) asked consumers to compare two sites drawn from 1 of 10 different domains, and to make a judgement about which of the sites was more credible. They were asked to

Figure 1. Staged model of trust



supplement this judgement with comments, and notably 46.1% of those comments reflected design qualities, as indicated by the following sample of comments drawn from four participants.

More pleasing graphics, higher quality look and feel.

Not very professional looking. Don't like the cheesy graphics. Looks childish and like it was put together in 5 minutes.

Such comments are remarkably similar to those elicited by Briggs et al. (2002) in a qualitative investigation of trust in sites offering advice and information to potential house buyers (Study 1). In that study, participants had to search for relevant information and then discuss which sites they would return to and which they would reject. Once again, a positive first impression was linked to good design and an absence of amateur mistakes as well as to indications of expertise, while a negative first impression was more explicitly tied to poor design.

Similar results have been noted within the health domain. Peterson et al. (2003) explored consumer experiences in searching for and evaluating information on medicines online. They found that participants reported quickly, rejecting sites that were slow to load, that contained too many graphics, and that had pop-up advertisements. Social cues are also important in the design of trustworthy Web sites. Appropriate graphics and photographs can add to a sense of social presence and inclusion, whilst inappropriate mission statements or alienating language can have the opposite effect (Wang & Emurian, 2005).

Faced with a vast array of information, it is likely that people, and in particular unmotivated students or people under time pressures, will fall back on heuristic processes. Sillence et al. (Sillence, Briggs, Fishwick, & Harris, 2004c) have shown that this rapid screening is very effective in the initial stages of contact with a health Web site. Within a 30-second window, participants were able to efficiently sift information, recognizing and rejecting general portals and sales sites quickly. This may be because such sites have distinctive design features associated with them, but some content processing is also underway.

Good design and credible information do not always go "hand in hand." Sillence et al. (2004c) noted that credible health sites such as those run by BUPA and the NHS were often rejected because they failed to convey through their design the right cues to their content. Well-designed sites with large financial backing, such as those owned by pharmaceutical companies or large sales companies, may well contain credible yet very biased information and advice. Fraudulent Web sites designed to be used in phishing scams are often very well-designed sites containing false information. Experts in the health domain warn against consumers' over reliance on the design of health Web sites as an indicator of credibility (Stanford et al., 2002).

Stage 2: FurtherInvolvement with the Site and Careful Evaluation of Site Content

Those investigations that involved real consumers or that required some protracted engagement with a site or that have asked customers about the general principles underpinning e-commerce transactions have generated a family of trust models with reasonable agreement. In general, the models suggest that trust, which supports online engagement, is influenced by perceived integrity and expertise, predictability, or familiarity of content and reputation (e.g., Bhattacherjee, 2002; Briggs et al., 2002, study 2; McKnight & Chervany, 2001). A number of studies also highlighted the importance of interface factors (ease of use and functionality) that help to reduce the transaction costs of an exchange (e.g. Egger, 2000, 2001; Lee, Kim, & Moon, 2000).

For example, Bhattacherjee (2002) developed a psychometric scale for trust in online transactions that was tested in two field trials and modified accordingly. The resultant seven item scale tapped into three trust elements:

1. **Ability:** both in terms of expertise and information access.
   Example: "Amazon has the skills and expertise to perform transactions in an expected manner."
2. **Integrity:** encompassing issues of fairness of conduct in transactions, customer service, and data usage.
   Example: Amazon is fair in its use of private user data collected during a transaction."
3. **Benevolence:** in terms of keeping the customers interests in mind in terms of showing empathy and responsiveness to customer concerns.
   Example: "Amazon is open and receptive to customer needs."

In a large scale study of credibility conducted at the Stanford persuasion laboratory, over 1,400 participants completed a questionnaire concerned with those factors they felt made Web sites more or less credible (Fogg, Marshall, Laraki, Osipovich, Varma, Fang, et al., 2001). The authors found that "commercial implications" of a Web site negatively affected its perceived credibility. Users penalized sites that had an aggressive commercial flavour. This included sites that required a paid subscription to gain access or sites that had a commercial purpose as opposed to an academic purpose. Bernhardt and Felter (2004) also noted that parents mistrusted commercial sales sites when searching online for health advice for their children. Conversely, sites that convey expertise through the inclusion of authors' credentials and references were viewed in a positive light. External links to outside materials and sources were also seen favourably as evidence of the site's honest and unbiased intention. (Fogg et al., 2001). Within a health context, however, research suggests that perceptions of credibility in relation to commercialism may be more flexible and are highly dependent on topic and content. Walther, Wang, and Loh (2004) found that the presence of advertising on health Web sites only had a deleterious effect on the credibility of sites with .org domains. It actually had a positive effect on .com and educational domains (.edu).

In the health domain, Peterson et al. (2003) found that consumers' opinions on credible sources of information on medicines varied, with some participants viewing pharmaceutical companies as the "official" information on a medicine and others preferring what they considered to be more impartial sources, such as Government organizations and educational establishments. Perceptions of impartiality vary according to the health domain (see our work on the Measles, Mumps, Rubella (MMR) vaccination (Sillence, Brigs, Fishwick, & Harris, 2004b), but remain a key factor in predicting trust. Bernhardt and Felter (2004) found that parents thought that ulterior motives of a site undercut the reliability of the health information they provided.

Stage 3: Subsequent Relationship Development and Integration

As previously mentioned, few studies within the e-commerce trust literature have looked at longer-term relationships between consumer and Web site. This is surprising given that early psychological models of trust between individuals focused explicitly on the build up of a relationship over time.

Trust is a consequence as well as an expectation of action, which means that initial trust judgements will be modified by experience. This interpretation of trust was evident in Rotter's (1967) original view of interpersonal trust, where trust in a generalized other could develop from successful and consistent exchanges with parents and siblings. Gambetta (1988) emphasized the developmental nature of trust by arguing that the point where we shift from saying, "I don't trust X" to "I trust X," is a threshold on this continuum that will vary with individual tendencies (e.g., a disposition to trust) and experience: "trust is not a resource that is depleted through use; on the contrary, the more there is the more there is likely to be" (Gambetta, 1988, p. 234).

Those online studies taking a longer-term perspective on trust emphasize the importance of shared values between customer and vendor. In these studies, trust in a longer term e-commerce relationship is a function not only of competence and predictability, but is highly influenced by the extent to which e-vendors are good communicators and show sensitivity to the personal values and circumstances of the consumer. In this way, good personalization practices are shown to be important for the development of a trusting relationship

In Florian Egger's (2000, 2001) studies of online customers, good personalized communication between customer and vendor was shown to be vital to the development of trust. Egger developed a model of trust for electronic commerce (MoTEC), where trust is initially determined by three factors: (a) the users knowledge of the domain and reputation of the vendor, (b) the impression made by the interface, and (c) the quality of the informational context as assessed by the consumer, but where a fourth factor, relationship management, becomes influential over time:

Relationship management reflects the facilitating effect of timely, relevant and personalized vendor-buyer interactions on trust development (pre-purchase) and maintenance (post-purchase). (Egger, 2001)

Fogg et al. (2001) also noted the influential role of good personalized transactions in their large-scale study of credibility. One of the scales in the questionnaire measured "tailoring" of content and included the following four items:

- The site sends e-mails confirming the transactions you make.
- The site selects news stories according to your preferences.
- The site recognizes that you have been there before.
- The site requires you to register or log in.

Tailoring was found to increase credibility, although the effect was more profound for the older users. In other words, older respondents reported higher credibility evaluations for sites that used some type of tailoring. The role of personalization in online trust was also apparent in a study in which people were asked about advice on the Internet (Briggs et al., 2002). This was an online questionnaire-based study in which a total of 2,893 respondents said they had actually sought advice online. These individuals were asked to give information about the site they had used in terms of issues related to trust as identified in an extensive review of trust literature. In a regression analysis, a clear three-factor model of trust emerged. Personalization was one of these factors and included items such as:

- Did the respondent feel involved in the process?
- Was the site interactive?
- Was the information tailored to the participant?
- Were different courses of action suggested?
- Was a peer commentary available?

Personalization then seems to be an important enabling factor for trust in online advice and indeed in other studies of trust in e-commerce that have explored full engagement with Web sites.

The online community literature also suggests that long-term relationships online are based upon a sense of trust that develops through the exchange of often personal information. The ability to personalize content and to build personalized content is important. The ability to share empathy is also key, especially in health-based communities (Preece & Ghozahti, 2001), but the community context allows individuals some control over disclosure (Gray et al., 2005). Studies within the health domain have examined people's motivations for continued use of the Internet. Rozmovits and Ziebland (2004) noted that in cancer patients, long-term support was the key factor. The Internet was often used to share experience and advice, and to contact support groups and chat rooms. In other health domains, for example, hypertension, personal experiences are also important (Sillence, Briggs, & Herxheimer, 2004).

There is some debate as to the impact of the Internet on health decision making and outcomes, for example, in terms of choosing one form of medication over another or choosing to make a lifestyle choice such as giving up smoking. Coulson (personal communication), for example, found that people using Internet resources for health topics such as HIV and Aids often took on medical advice that was incorrect. However, recent studies indicate that there is some interplay between an individual's use of the Internet and other personal health information sources. Gray et al. (2005) noted that adolescents using the Internet for health advice would check the information that they had received from personal sources on the Internet for consistency. Bernhardt and Felter (2004) also noted that information convergence is a strategy used by parents looking online for health advice about their children, to assess whether or not the information is trustworthy. Parents thought that information that they had read across a number of sites and across a range of different sources (off-line and online) was more trustworthy.

VALIDATING THE MODEL

The staged model of trust proposed by Briggs et al. (2004), and illustrated in Figure 1, reconciles the different findings with regard to the literature concerning online trust. It also allows for the development of guidelines that relate to the stage of engagement with a Web site. As discussed previously, the majority of online trust studies have suffered from using unrealistic settings, unmotivated participants, and from focusing purely upon initial impressions of trust. In keeping with the staged model of trust, we report on a longitudinal in-depth qualitative study that investigates first impressions of a Web site and early analysis of content, as well as longer-term engagement with health Web sites.

The qualitative study is a 4-week investigation of real consumers' attitudes towards online health advice. Four groups of consumers, all with an interest in a specific health issue, searched the Internet for advice over a 4-week period. A variety of data collection methods were used including data logging, log books, verbal protocols, and group discussion. The aim was to elicit the key themes surrounding initial trust and mistrust of health Web sites.

Qualitative Investigations of Trust in Online Health Advice

A novel methodology was developed in order to examine how genuine health consumers search for and appraise online health information and advice. The methodology allows for the examination of people's decision-making processes, in particular with respect to trust, to be observed through the first few seconds of interaction with a Web site to 1 year of engagement.

Four longitudinal studies examining groups of people faced with risky health decisions were undertaken. The health topics were menopause and HRT, hypertension, and healthy living. In addition, a small-scale study examining MMR was also carried out. The methodology is described in more detail in Sillence et al. (2004a). In total, 40 participants (30 females and 10 males) took part in four separate studies focusing on four different health topics. Despite including one solely female health topic, it was anticipated that the final sample would have been more balanced. We can only assume that the gender bias in our sample reflects the higher female interest in health advice within an online setting that has been noted in other studies (Briggs et al, 2002; Pew, 2000) and thus were more willing and committed to take part in the longitudinal study.

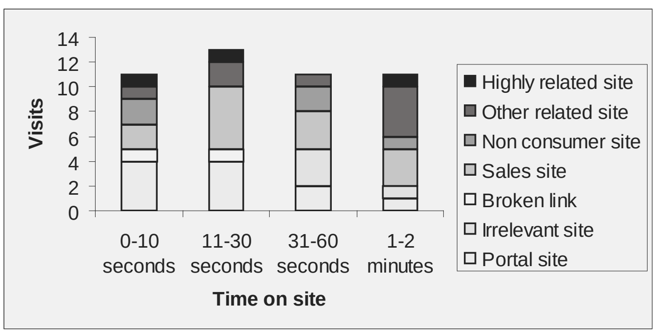
The methodology used was the same in each study. Participants were invited to attend a total of four 2-hour sessions at Northumbria University, UK. During all four sessions, participants used the Internet to search for information and advice on the health topic in question, followed by a group discussion with a facilitator. Participants were told to freely surf the Web during Sessions 1 and 4, and were directed to specific Web sites during Sessions 2 and 3. These specific sites were chosen for their trust design elements, and to reflect the range of sites actually used by consumers. All the Web sites visited by the participants were logged and the amount of time spent on each site was recorded. In addition, the participants were asked to record their perceptions of each site visited in a logbook and use this information during the discussion sessions. All the discussions were transcribed and subject to content analysis. At the end of the fourth week, the participants were given diaries in order to record their ongoing information and advice searches over a 6-month follow-up period.

Rejecting or Selecting Health Sites

The first study indicated that during the first 10 seconds of scanning Web sites, participants are keen to find something that is immediately relevant to them and to find content that they can "latch onto." Many of the rejected sites at this stage were portal sites. Portal sites often require participants to carry out a deeper search or to input additional search terms. At 30 seconds, the largest category of sites rejected at this point were sales based. Participants were clearly able to detect and reject signs of commercial activity and a sales orientation very rapidly. At 2 minutes, many sites rejected at this stage were either sales based or were classified as "other related." Some sales sites do not reveal their commercial element until the user has explored the site in some detail. Participants may search through the site and then make a decision to reject the site after discovering its commercial motivation. Figure 2 illustrates the time to reject sites and the type of content the site provided in the case of the HRT and menopause study.

The transcripts from the group discussions and the verbal protocols were examined in terms of the selection (trust) or rejection (mistrust) of Web sites. A number of themes relating to the first impressions of the Web site and characteristics of trustworthy sites emerged. In terms of rejection, the overwhelming majority of comments related to the design of the Web site. The look and feel of the Web site was clearly important to the participants. Visual appeal, plus design issues relevant to site navigation, appeared to exert a

Figure 2. Rejection time and content of sites



strong influence on people's first impressions of the site. Poor interface design was particularly associated with rapid rejection and mistrust of a Web site. In cases where the participants did not like some aspect of the design, the site was often not explored further than the homepage, and was not considered suitable for revisiting at a later date. Negative comments included an inappropriate name for the Web site; complex, busy layout; lack of navigational aids; and pop-up adverts (see P10 in Box 1).

The participants mentioned a number of factors in terms of the sites they had selected to explore in more depth. The participants liked sites that contained a great deal of information that was presented in such a manner that an individual could quickly pinpoint their own specific areas of interest. Participants trusted the selected sites because they demonstrated an in-depth knowledge of a wide variety of relevant topics, and put forward unbiased clear information. Participants were more likely to trust the information if they could verify it and cross-check it with other Web sites. Most individuals preferred sites that were run by reputable organizations or had a medical or expert feel about them. They trusted the information on such Web sites, especially when the credentials of the site and its authors were made explicit. Most participants showed some distrust of the advice and information on Web sites sponsored by pharmaceutical companies or those explicitly selling products. This is interesting given that medical expert reviews of HRT and menopause Web sites recommend pharmaceutical sites as providing a comprehensive and credible source of information (Reed & Anderson, 2002). Participants were looking for sites that were written by people similar to themselves, who shared similar interests. In this way, advice feels personalized for them.

Over the Longer Term

We asked the participants in our longitudinal studies to keep diaries over a 6-month period in order to assess how they followed up the information and advice they had read online. The participants were asked to record any sites that they visited during that time and to make a note of what prompted

Box 1. Examples of features associated with early rejection

Inappropriate sounding names

Well I didn't like it, I think it was possibly the name but I didn't hold out any confidence in something called Netdoctor at all, it sounds more like an IT company to me. (HRT study, P11, 59 years old)

Commercial

It's a very visual thing but then I am a very visual person it has to appeal and the flip side of that is that anything that's covered in ads and pop ups and stuff like that, it's the principle I'm just not interested. (HRT study, P10, 48 years old)

It just looked a bit cheaply made and the donations sign just put me off straight away. (MMR study, P2, Female, 35)

Nothing to latch onto

There was one in particular I was just I went into it and then after a couple of minutes I just thought its too complicated its just too much I can't be bothered with it. (Hypertension study, P13, Female, 52) them to go online. We also asked them to note any "off-line" interactions concerning this particular health topic, for example, interactions with health professionals and friends and family. They were asked to make a note of any other resources they had used, for example, television, radio, or newspapers. We were interested to find out when and where the Internet becomes involved in the decision-making process and to assess whether or not information integration had occurred.

Integration

We noted that the Internet was affecting decision making concerning treatments and lifestyle issues differentially across health topics and across individuals. For some participants, online information and advice was affecting their thinking on a week–by-week basis. For others, the affect on their decision-making processes was ongoing over the course of the following year. Here is an example that illustrates how one participant used the information online, how it was integrated with other sources, and overall, how it affected outcomes in terms of decision making.

DT took part in the hypertension study. He was taking medication for his high blood pressure and was unhappy with what he saw as the debilitating side effects. Discussions prior to the study had

Box 2. Examples of positive features associated with trusted sites

Clear and informative content

I found an absolutely marvellous site I was really, really taken with it, it went into so such clear explanations and with a breakdown of the different, oestrogen, progesterone, testosterone and what they actually do and how they link together all along the way it kind of encouraged you at the beginning to work through the site progressively if you wanted to get like a whole raft of background knowledge and then it would help you make decisions, it was great. (HRT, P1, 52 years old)

Impartial information

I suppose it was interesting in the sense that it was so biased against the MMR its useful to read it I mean but basically it was just a self help group website and so again it was like totally propaganda asking for donations and this kind of thing that I found it a bit disturbing actually I didn't like it, it was just so one sided I didn't like it. (MMR study, P1, Female, 33)

A Web site for someone like me

You can go on and find medical information, but what I want to find out is, you know, people giving reports of, right I've got high blood pressure and I've done this and I've done that and it brought my blood pressure down. (Hypertension study, P6, Female, 45)

Familiar language

I like sites that use familiar language, language I am comfortable with not lots of American terms that don't mean anything to me or probably mean different things like drug names for example. (Healthy living study, P2, Male, 35).

It was a British site I guess because it called hot flushes hot flushes not hot flashes like American sites do—that really annoys me that. (HRT study P4, 45 years old) left him frustrated since his doctor had advised him that side effects were very unusual for this medication. DT was convinced his symptoms were not normal. During the sessions, he came across one site in particular thta detailed other people's experiences of side effects whilst taking this particular medication. During the group discussions, he told everyone of his findings and his intention to show these results to his doctor. The diaries indicated that he returned to the Web site several times before he visited his doctor. In consultation with his doctor, he had agreed on a change in his medication.

During the follow-up interviews, all the participants expressed the opinion that they were now more careful evaluators of online and off-line content. They also reported a new found confidence with respect to doctors and medical information.

Registration and Self-Disclosure

KK took part in the healthy living study. She was interested in planning a new exercise regime. She looked at ivillage (http://www.ivillage.co.uk) during one of the free search sessions and liked the interactive features of the Web site, such as the calculators, the alert functions, and the large message boards facility. Her diary indicates that she engaged with the site over the longer term. She registered her details with the site in order to be able to post messages and receive e-mail alerts and reminders concerning her personalized exercise plan.

Overall, the participants expressed their desire for personalized advice, but thought that currently it was difficult to obtain online. Often the information on the Web site was felt to be limited and lacking in personal detail. The participants did suggest they would be prepared to disclose more personal information in return for more personalized advice. However, the nature of the personal information was important in this respect. Whilst sites such as BBC's Big Challenge <http://www.bbc.co.uk/bigchallenge> and ivillage <http://www.ivillage.co.uk> offer something more akin to human rather than automated personalized advice, very few of the participants wanted to subscribe to sites such as these or to register their contact details in return for more tailored advice (see Sillence, Briggs, Fishwick, & Harris, 2005c for more detail).

SUMMARY

Initial design appeal is important to users. The speed with which consumers will reject information should provide a salutary lesson regarding the importance of providing the right cues to site content in a highly visible manner. Social identity emerged as a major factor for building trust in online advice. General sites with few identity cues (NHS, Bupa) were seldom selected as trusted sources. Sites with home pages containing relevant information and that offered the opportunity to browse stories from like-minded individuals were viewed very positively. These results add to and extend previous work on trust by Briggs et al. (2002). Once again, personalization and credibility appear to be important factors in predicting whether or not a respondent will act upon the online health advice offered. Additionally, ease of access to useful information also has an important role to play in this respect. This extends the role of usability within online advice and online trust. Good designs combined with the stories that resonate with individual experience are very important factors in building initial trust. There is support for the staged model of trust in which heuristic evaluation and rapid screening is followed up by careful content evaluation and then longer-term engagement with a site. The important trust factors emerging from the studies provide a useful starting point for the development of guidelines for developing trust in e-health.

GUIDELINES FOR DEVELOPING TRUST IN E-HEALTH

A composite picture of trust guidelines (taken from the literature and from the studies reported previously) is offered next. The guidelines are split into three sections to reflect the staged model of trust and to suggest the point at which they are most influential in terms of interaction.

Stage 1: Heuristic evaluation

1. Use a professional and attractive design
2. Ensure good ease of use
3. Provide good cues to content on the homepage
4. Maximize the familiarity or predictability of the layout and interaction
5. Do not mix advertising and content—avoid sales pitches and banner adverts

Stage 2: Content evaluation

1. Include background information on the knowledge and expertise of the authors
2. Make clear the motivations of the authors—provide good links to independent Web sites in the same domain to provide reassurances on bias
3. Ensure that the Web site reflects the social identity of the user
4. Offer a personalized service that takes accounts of each user's health needs and preferences
5. Include peer contributions and the ability to contribute to the site

Stage 3: Longer term engagement

1. Ensure content is regularly updated
2. Provide interactive features on the site as well as alternative ways of engaging users, for example, e-mail or text message alerts
3. Enable users to register in order to obtain personalized advice— make clear the purpose of registration and the privacy and security implications
4. Facilitate integration with other Web sites—include good links to other Web sites within the same domain. Provide local off-line addresses and contact details
5. Provide clearly stated privacy policies

FUTURE TRENDS

We began this chapter with a discussion of some of the reasons why considerations of trust are important in the expanding field of e-health. Let us end the chapter with some explicit considerations of the trust issues raised by future visions of e-health. The changing nature of e-health and its supporting technologies means that the issue of trust becomes ever more important. Health information and advice will no longer solely be accessed from a PC in the home but from public access kiosks in a variety of locations. Health alert systems on mobile phones will mean people receive personal information in public settings. The development of networked health information and the electronic patient record system in the UK means that increasing numbers of people will have access to sensitive health information shared across large networks.

The demand for increasingly personalized e-health services also raises issues of trust and self-disclosure. The qualitative studies reported in this chapter indicate that people desire personalized health services but are reluctant to register their details with Web sites. Work by Sillence and Briggs (in press) also indicated that users found it more difficult to get personalized health advice online compared, for example, to financial advice. The increase in personalization coupled with increasing "public technologies" raises a number of important points regarding privacy and security. Burgoon (1982) described two kinds of privacy, informational privacy and physical privacy, both of which will have to be addressed in future work on e-health.

As increasingly sophisticated Web-based services are being developed, health consumers' trust in the information, and increasingly in the technology that facilitates the transfer of the information, will become of particular importance. Users must be prepared to place their trust not only in the people, but also in the technology that underpins an interaction. Understanding the context for trust, therefore, involves understanding issues of encryption and data security as well as understanding the development of a psychological bond. Bollier (1986), for example, argued that it is vital to distinguish between issues of "hard trust,"—involving authenticity, encryption, and security in transactions, and issues of "soft trust,"—involving human psychology, brand loyalty, and user-friendliness.

We are involved in several related research projects that address these issues, and some things have become clear. Firstly, we need to develop research methods that invite proper participation and involvement. The trust literature is peppered with investigations of users' intention to trust rather than explorations of their trusting behaviour. It is not simply enough to ask people about trust and related issues such as self-disclosure and privacy in the abstract – because, quite simply, what people say and what they do are two different things. Paradigms that allow researchers to investigate the whole cycle of trust from initial trusting dispositions through intention to trust and trusting behaviour have been developed over the past couple of years. These methodologies are currently being refined to include further ways of determining important trust factors in e-health, for example, incorporating measures of individual differences and personalization. The use of eye-tracking, for example, provides another tool in our kit to study the attitudes and behaviours of genuine health consumers engaged in different levels of risky decision making. We also need to know a great deal more about what happens following loss of trust and how the integration process with other sources, for example, healthcare providers, affects decision making. Such questions will be crucial for the development of e-health systems that people can genuinely use.

CONCLUSION

This chapter has proposed a staged model of trust to help explain the trust processes that are involved with consumers' use of online health advice and information. The studies presented in this chapter have documented some of the factors that can influence people's perceptions of the trustworthiness of online health advice and directly infleunce behaviour. The studies provide support for the staged model of trust and have assisted with the development of guidelines for trust practices in e-health. These guidelines relate to both the design and content features of health Web sites, and will assist in prediciting the uptake of health advice. Future work will focus upon the importance of individual differences and personalization in e-health, and will explore the consequences of a loss of trust.

REFERENCES

BBC Online. (2004). Health advice through your TV. Retrieved November 1, 2005, from http://news.bbc.co.uk/go/pr/fr/-/1/hi/health/4100717.htm

Bernhardt, J. M., & Felter, E. M. (2004). Online pediatric information seeking among mothers of young children: Results from a qualitative study using focus groups. *Journal Medical Internet Research*, *6*(1), e7.

Bessell, T., McDonald, S., Silagy, S., Anderson, J., Hiller, J., & Sansom, L. (2002). Do Internet interventions for consumers cause more harm than good? A systematic review, *Health Expectations*, *5*(1), 28-37.

Bhattacherjee, A. (2002). Individual trust in online firms: Scale development and initial test. *Journal of management Information Systems, 19*(1), 211-241.

Bollier, D. (1996). *The future of electronic commerce*. A report of the Fourth Annual Aspen Institute Roundtable on Information Technology. Aspen, CO: The Aspen Institute.

Brien, A. (1998). Professional ethics and the culture of trust. *Journal of Business Ethics, 17*(4), 391-409.

Briggs, P., Burford, B., De Angeli, A., & Lynch, P. (2002). Trust in online advice. *Social Science Computer review, 20*(3), 321-332.

Briggs, P., de Angeli, A., & Simpson, B. (2004). Personalisation and trust: A reciprocal relationship? In M.C. Karat, J. Blom, and J. Karat (Eds), *Designing personalized user experiences for eCommerce*. Kluwer.

Burgoon, J. K. (1982). Privacy and communication. *Communication yearbook, 6,* 206-249.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39,* 752-766.

Cheskin Research and Studio Archetype/Sapient. (1999). *eCommerce trust study.* Retrieved July 16, 2003, from http://www.cheskin.com/p/ar.asp?mlid=7&arid=40&art=0&isu=1

Corritore, C. C. L L., Kracher, B., & Wiedenbeck. S. (2003). On-line trust: Concepts, evolving themes, a model. *International Journal of Human-Computer Studies, 58*, 737-758.

Dobson, R. (2003). News extra: Study reports on use of ”touch screen” health kiosks. *British Medical Journal, 326,* 184.

Egger, F. N. (2000). Trust me, I'm an online vendor: Towards a model of trust for e-commerce system design. In *Proceedings of CHI 2000*. ACM Press.

Egger, F. N. (2001). Affective design of e-commerce user interfaces: How to maximise perceived trustworthiness. In *Proceedings of CAHD: Conference on Affective Human Factors Design* (pp. 317-324).

Eysenbach, G. (2001) What is e-health? *Journal Medical Internet Research, 3*(2), e20.

Eysenbach, G., & Köhler, C. (2002). How do consumers search for and appraise health information on the World Wide Web? Qualitative study using focus groups, usability tests and in-depth interviews. *BMJ, 324,* 573-577.

Eysenbach, G., Powell, J., Englesakis, M., Rizo, C., & Stern, A. (2004, May 15). Health related virtual communities and electronic support groups: Systematic review of the effects of online peer to peer interactions. *British Medical Journal,* 328,1166, doi:10.1136/bmj.328.7449.1166.

Eysenbach, G., Powell, J., Kuss, O., & Sa, E-R. (2002). Empirical studies assessing the quality of health information for consumers on the World Wide Web, a systematic review. *Journal of the American Medical Association, 287*(20), 2691-2700.

Fogg, B. J., Kameda, T., Boyd, J., Marchall, J., Sethi,R., Sockol, M., & Trowbridge, T. (2002). *Stanford-Makovsky Web credibiltiy study 2002: Investigating what makes Web sites credible today.* A research report by the Stanford Persuasive Technology Lab and Makovsky & Company, Stanford University. Retrieved from http://www.webcredibility.org

Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., & Treinen, M. (2001). What makes Web sites credible? A report on a large quantitative study. In *Proceedings of CHI 2001* (pp. 61-66). Seattle, WA: ACM Press.

Gambetta, D. (1988). Can we trust trust? In I. D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations* (pp. 213-237). Oxford: Basil Blackwell.

Grabner-Krauter, S., & Kaluscha, E.A. (2003). Empirical research in online-trust: A review and critical assessment. *International Journal of Human-Computer Studies, 58,* 783-812.

Gray, N. J., Klein, J. D., Noyce, P. R., Sesselberg, T. S., & Cantrill, J. A. (2005). Health information-seeking behaviour in adolescence: The place of the Internet. *Social Science and Medicine, 60,* 1467-1478.

Hart, A., Henwood , F., & Wyatt, S. (2004). The role of the Internet in patient practitioner relationships: Findings from a qualitative research study. *Journal Medical Internet Research, 6*(3), e36.

Jadad, A. R., & Gagliardi A. (1998). Rating health information on the Internet: Navigating to knowledge or to Babel? *Journal of the American Medical Association, 25* 279(8), 611-4.

Kalet, A., Roberts, J. C., & Fletcher, R. (1994). How do physicians talk with their patients about risks? *Journal of General Internal Medicine, 9*, 402-404.

Kanuga, M., & Rosenfeld, W. D. (2004). Adolescent sexuality and the Internet: The good the bad and the URL. *Journal of pediatric and adolescent gynecology, 17*(2), 117-124.

Kim, J., & Moon, J. (1998). Designing towards emotional usability in customer interfaces: Trustworthiness of cyber-banking system interfaces. *Interacting with Computers, 10,* 1-29.

Klein, J. D., & Wilson, K. M. (2003). Delivering quality care: Adolescents' discussion of health risks with their providers. *Journal of Adolescent Health, 30*(3), 190-195.

Lee, J., Kim, J., & Moon, J. Y. (2000). What makes Internet users visit cyber stores again? Key design factors for customer loyalty. In *Proceedings of CHI 2000,* The Hague, Amsterdam (pp. 305-312).

Marsh, S., & Dibben, M. (2003). The role of trust in information science and technology. In B. Cronin (Ed.), *Annual Review of Information Science and Technology, 37* (pp.465-498). Medford, NJ: Learned Information.

McKnight, D. H., & Chervany, N. L. (2001). Trust and distrust definitions: One bite at a time. In R. Falcone, M. Singh, & Y.-H. Tan (Eds.), *Trust in cybersocieties.* Berlin: Springer-Verlag.

Mead, N., Varnam, R., Rogers, A., & Roland, M. (2003). What predicts patients' interest in the Internet as a health resource in primary care in England? *Journal of Health Services Research and Policy, 8*(1), 33-39.

Meyerson, D., Weick, K. E., & Kramer, R. M. (1996). Swift trust and temporary groups. In R. M. Kramer. & T. R. Tyler. (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 166-195). Thousand Oaks, CA: Sage Publications.

Morahan-Martin, J. M. (2004). How Internet users find, evaluate and use online health information: A cross-cultural review. *CyberPsychology and Behaviour, 7*(5), 497-510.

NHS Exec. (1998, September). *Information for health: An information strategy for the modern* (NHS 199802005).

Oh, H., Rizo, C., Enkin, M., & Jadad, A. (2005). What is eHealth (3): A systematic review of published definitions. *Journal Medical Internet Research, 7*(1), e1.

Pagliari, C., & Gregor, P. (2004). *Literature review of traditional research databases.* Retrieved from http://www.sdo.lshtm.ac.uk/ehealth.html

Peterson, G., Aslani, P., & Williams, K. A. (2003). How do consumers search for and appraise information on medicines on the Internet? A qualitative study using focus groups. *Journal of Medical Internet Research, 5*(4), e33.

Pew Research Center. (2000). *The online health care revolution: How the Web helps Americans take better care of themselves.*Retrieved from http://www.pewinternet.org

Pew Research Center. (2003). *Health searches and email have become more common place, but there is room for improvement in searches and overall Internet access.* Retrieved from http://www.pewinternet.org/

Preece, J., & Ghozati, K. (2001). Observations and explorations of empathy online. In R.R. Rice & J. E Katz (Eds.). *The Internet and health communications: Experience and expectations* (pp. 237-260). Thousand Oaks, CA: Sage publications Inc.

Reed, M., & Anderson, C. (2002). Evaluation of patient information Internet Web sites about menopause and hormone replacement therapy. *Maturitas, 4,* 135-154.

Reeves, P. M. (2001). How individuals coping with HIV/AIDS use the Internet. *Health Education Research,16*(6), 709-719.

Riegelsberger, J., Sasse, M. A., & McCarthy, J. (2003). Shiny happy people building trust? Photos on e-commerce Web sites and consumer trust. In *Proceedings of CHI 2003.* ACM Press.

Riegelsberger, J., Sasse, M. A., & McCarthy, J. D. (2005). The mechanics of trust: A framework for research and design. *International Journal of Human-Computer Studies*, *62*(3), 381-422.

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality, 35,* 651-665.

Rozmovits, L., & Ziebland, S. (2004). What do patients with prostate or breast cancer want from an Internet site? A qualitative study of information needs. *Patient Education and Counselling, 53,* 57-64.

Sieving, P. (1999). Factors driving the increase in medical information on the Web—one American perspective. *Journal Medical Internet Research, 1*(1), e3.

Sillence, E., & Briggs, P. (in press). Please advise: Using the Internet for health and financial advice. *Computers and Human Behavior.*

Sillence, E., Briggs, P., Fishwick, L., & Harris, P. (2004a, April 24-29). Trust and mistrust of online health sites. In *Proceedings of CHI'2004*, Vienna, Austria (pp. 663-670). ACM Press.

Sillence, E., Briggs, P., Fishwick, L., & Harris, P. (2004b, June 5-6). What parents make of MMR and the Internet. *He@lth Information on the Internet, 39.*

Sillence, E., Briggs, P., Fishwick, L., & Harris, P. (2004c, September 6-10). Timeline analysis: A tool for understanding the selection and rejection of health Web sites. In A. Dearden & L. Watts (Eds.), *Proceedings of the Conference HCI 2004: Design for Life*, Leeds, UK (Vol. 2, pp.113-116).

Sillence, E., Briggs, P., Fishwick, L., & Harris, P. (2005c, December). Do health Web sites offer patients personalized information and advice? *He@lth Information on the Internet, 48,* 9-10.

Sillence, E., Briggs, P., & Herxheimer, A. (2004, December). Personal experiences matter: What patients think about hypertension information online. *He@lth Information on the Internet*, *42*, 3-5.

Smart, J. M., & Burling, D. (2001). Radiology and the Internet: A systematic review of patient information resources. *Clinical Radiology, 56*(11), 867-870.

Stanford, J., Tauber, E., Fogg, B. J., & Marable, L. (2002). Experts vs. online consumers: A comparative credibility study of health and finance Web sites. *Consumer Web Watch Research Report.* Retrieved August 2003, from http://www.

consumerwebwatch.org/news/report3_credibility research/slicebread_abstract.htm

Taylor, H., & Leitman, R. (2002). The future use of the Internet in four countries in relation to prescriptions, physician communication and health information. *Health Care News 2*(13). Retrieved from http://www.harrisinteractive.com/news/newsletters_healthcare.asp

Walther, J. B.,Wang, Z., & Loh, T. (2004). The effect of top-level domains and advertisments on health Web site credibility. *Journal Medical Internet Research*, *6*(3), e24.

Wang, Y. D., & Emurain, H. H (2005). An overview of online trust: Concepts, elements and implications. *Computers in Human Behavior, 21*, 105-125.

Wantland, D., Portillo J., Holzemer W., Slaughter, R., & McGhee E. (2004). The effectiveness of Web-based vs. non-Web based interventions: A meta-analysis of behavioral change outcomes. *Journal Medical Internet Research, 11,* 10, 6(4), e40.

WHA. (2002). Internet transforming the doctor patient relationship. Retrieved from http://www.w-h-a.org/wha2/index.asp

Wyatt, J. C. (1997). Measuring quality and impact of the World Wide Web. *British Medical Journal, 314,* 1879-81.

Ziebland, S., Chapple, A., Dumelow, C., Evans, J., Prinjha, S., & Rozmovits, L. (2004). How the Internet affects patients' experience of cancer: A qualitative study. *British Medical Journal, 328*, 564.

This work was previously published in Trust in E-Services: Technologies, Practices, and Challenges, edited by R. Song, L. Korba, and G. Yee, pp. 235-258, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.10
Privacy and Access to Electronic Health Records

Dick Whiddett
Massey University, New Zealand

Inga Hunter
Massey University, New Zealand

Judith Engelbrecht
Massey University, New Zealand

Jocelyn Handy
Massey University, New Zealand

INTRODUCTION

The special relationship of trust that needs to exist between a patient and his or her physician has been recognized since the origins of the profession, and the need for doctors to keep confidential any information disclosed to them is codified in the Hippocratic Oath. A distinctive feature of the health records which arises from this relationship is the intimate nature of the information that they may contain; consequently, it is vitally important to maintain the confidentiality of the records and to protect the privacy of the patients. Privacy has long been recognized as a fundamental right in most western societies (Westin, 2003), and unless a patient can be sure that personal information will not be distributed against his or her wishes, the patient may be reluctant to disclose information that may in fact be crucial to his or her correct treatment (Ford, Bearman, & Moody, 1999; NZHIS, 1995), or he or she may refrain from seeking treatment (Sankar, Moran, Merz, &

Jones, 2003). This is particularly true when health records contain sensitive information concerning issues like drug and alcohol problems, sexual behavior, mental health, or a genetic predisposition towards certain diseases. In such circumstances, the consequences of the inappropriate release of information could be extensive and might impact on many aspects of a person's life, such as the ability to gain employment, to maintain a marriage, or to obtain loans or life insurance (Chadwick, 1999; Woodward, 1995).

Within the healthcare sector there is a constant pressure to balance patients' requirements for personal privacy against the potential benefits that may accrue to society as a whole from the more widespread use of their personal information. This issue is particularly relevant in developed countries that have been seeking to use computer-based patient records (CPRs) (Dick & Streen, 1991), electronic medical records (EMRs), and electronic health records (EHRs) to improve both organizational efficiency and the quality of care provided for patients (AHRQ, 2006).[1]

The potential benefits of EHRs are widely accepted, but there are also serious problems concerning the potential threats to patient privacy (Carter, 2000). The move from paper-based records to electronic records has greatly increased the potential threats to patients' privacy in two ways. Firstly, it has increased the risk of unauthorized access to patients' information by people both within and outside of an organization, since it is now no longer necessary to manually search through individual patient's records and it is possible to systematically search through collections of records from a distance (Goldschmidt, 2005). Secondly, the development of communications networks has greatly increased to the extent to which patient information is now routinely exchanged between different healthcare organizations so more people have access to it (Kissinger & Borchardt, 1996).

This article will explore some of the privacy issues associated with the development and use of EHRs. The first part describes the background and development of EHRs and the various ways that patient health information can be used and distributed within modern healthcare systems. It discusses the benefits that may accrue to the individual patient and also to healthcare organizations due to improved access to information. The second part then reviews some issues that arise from the use of EHRs, and it reviews research into patient attitudes towards the distribution of their health information. The final part of the article discusses some technologies that address the security requirements of patients such as role-based security systems (Sandhu, Coyne, Feinstein, & Youman, 1996), smartcard systems (Rienhoff, 2003), and finally, e-consent systems (Coiera & Clarke, 2004; Galpottage & Norris, 2005; Scott, Jennett, & Yeo, 2004), which aim to provide patients with much greater control over the access to their information.

BACKGROUND

Electronic information systems are often justified on the grounds that having access to more complete, accurate, and timely information facilitates better decisions. In the case of health records, these benefits may accrue directly to the individual patient in terms of better treatment or to the population in general through improvements to healthcare practice or administration (Mount, Kelman, Smith, & Douglas, 2000). The application of data-mining techniques to large numbers of EHRs could facilitate epidemiological and evaluative studies (Bath, 2004; Payton, 2003), and the information may also benefit healthcare administrators and managers by providing them with more comprehensive information about service usage and costs (Hannan, 1999).

Despite the wide range of their potential benefits, the introduction of comprehensive EHRs has been relatively slow because of the complexity of the health sector from technological, organizational, and ethical perspectives (Goldschmidt, 2005). The use of computer-based information systems to store patients' records has been evolving since the 1970s. The early systems tended to focus on the administrative details of a patient and to deal with a single episode of care. Because of the high development and implementation costs, early systems were mainly used in larger hospitals (Goldschmidt, 2005; Reichertz, 2006), but as computing costs have fallen, sophisticated systems have become widespread, and systems are now found in most hospitals and in many primary care or GP practices (Didham & Martin, 2004).

In the late 1980s and throughout the 1990s, the potential benefits of an integrated lifelong electronic health record began to be recognized and explored (Haux, 2006). For the individual patient, major benefits arise from the improved continuity of care which is possible if all healthcare practitioners have a complete and detailed history of the patient's conditions and treatments on which to base their diagnoses and decisions. Easy access to a comprehensive patient history would be particularly useful when a patient is referred to a practitioner for the first time, if a patient needs treatment when he or she is traveling and is away from his or her usual practitioner, or in accident or emergency situations when the patient is unconscious or unable to answer questions (Mount et al., 2000; Hunt, Haynes, Hannah, & Smith, 1998). For example, comprehensive records will improve medication management by allowing the practitioner to quickly check whether a patient is known to be allergic to a particular medicine, or whether a particular drug might have adverse interactions with other medicines the patient may be taking (Ministry of Health, 2001).

Unfortunately, the widespread introduction of EHRs has been hampered by organizational problems caused by the dispersed nature of the healthcare sector in most countries. Since many organizations are involved in providing and funding care, a patient's information tends to get fragmented and dispersed (Goldschmidt, 2005; Kissinger & Borchardt, 1996). Treatment of a typical condition may involve a patient visiting a community-based practice, such as a general practitioner (GP), consultations and treatments by hospital-based specialists, and tests and analyses undertaken by various laboratories. Typically, information from each of these encounters will be stored by each organization within its own separate computer system, which will have been designed to support the specific requirements of each organization and its users. Furthermore, the payment for all these services may come in part or full from several sources, such as the patient, the government, or some private insurance scheme, so further inter-organizational information exchange is required to sort out the finances (Kissinger & Borchardt, 1996). Current developments of EHRs are therefore focusing on distributed structures with a centralized summary of a patient's information, with links to detailed information located in other computer systems (AHRQ, 2006). Developments of this kind are currently being supported by the governments of the UK (NHS, 2001, 2002a), Canada (CHII, 2006), Australia (NHIMAC, 1999), and New Zealand (Ministry of Health, 2000, 2001), which are all pursuing the development of nationwide EHRs to help manage their public health systems. In the United States where there is no public health system, the government is still involved in similar developments that aim to promote the efficient interchange of health information by funding the National Health Information Infrastructure (NHII) project (USDHHS, 2001) and through legislation such as the Health Insurance Portability and Accountability Act (HIPAA) of 1996 (USDHHS, 2006), which defines standards for health information structure, security, and privacy. These developments should benefit individual patients and also improve the overall efficiency of the health sector.

PRIVACY AND PATIENTS' ATTITUDES

The potential threats to the confidentiality of the information that EHRs contain and their implications for patient privacy are controversial (Carter, 2000), and large-scale health information systems have often been fraught with problems (Hannan, 1998; Sicotte et al., 1998). A classic example of this occurred within the British National Health

Service when doctors boycotted an inter-organizational network designed to improve the exchange of information on the grounds that it threatened patients' privacy (Davis, 1996; Willcox, 1995). Research in New Zealand into clinicians' attitudes towards the use of information contained within EHRs indicated that some of the potential uses were unacceptable and would also lead them to withhold information (Handy, Hunter, & Whiddett, 2001).

Despite the fact that many western countries have adopted privacy legislation that is based on the OECD principles (OECD, 1980) which should protect the confidentiality of health information, concerns about the uses made of the information are not unfounded. In the UK, the Caldicott Committee Report (1997) into patient privacy identified several breaches of confidentiality, and the report made wide-ranging recommendations to improve the security of patients' information. In a survey of Australians, Mulligan (2001) found that 1.9% of respondents reported harm arising from unauthorized disclosure of their information by health services. Anderson (2000) cites a number of cases where patient privacy has been breached in the United States. The U.S. pressure group *Patientprivacyrights* (2006) is also concerned that amendments to HIPAA will increase the abuse of patient information. The group claims that medical information is used to influence employment and promotion decisions, to determine eligibility for bank loans and insurance, and to market pharmaceuticals.

Despite the importance of personal privacy, there is a lack of research which addresses the issue from the patients' perspective. A recent review of the literature (Sankar et al., 2003) identified nearly 6,000 articles related to issues of patient privacy, but the vast majority of them were written from the practitioners' perspective or addressed legal or regulatory issues. Only 6% of the identified articles were written from the patients' perspective, and of these only 110 (2%) were based on research. Furthermore, most of this research focused on specific groups or particularly sensitive issues, such as adolescents who are concerned about their information reaching their parents or people who are having HIV tests. Very little work has addressed the perspective of ordinary patients.

Some research into attitudes of patients or the public towards the distribution of their information has been undertaken in the UK in association with the National Health Service's Electronic Record Development and Implementation Program (ERDIP) (NHS, 2002a), the PERIC (Patient Electronic Record: Information and Consent) project (Shickle, Carlisle, & Wallace, 2002), and the 'Share with Care' project (NHS, 2002b). Patients' attitudes have also been studied in New Zealand (Whiddett, Hunter, Engelbrecht, & Handy, 2006), and in Australia the issues of confidentiality of patient information and consent for access have been addressed as part of the Health*Connect* project (Health*Connect* Program Office 2002, 2003). These projects have recognized the increasing levels of concern about personal privacy among the general public and have begun to try to identify their expectations. However, to date the exploration of the attitudes of patients and the development of systems to meet their requirements has not received the attention that these issues would seem to warrant.

Although the various studies of patients' attitudes have addressed different specific issues and have used different research methodologies, such as questionnaires, interviews, telephone interviews, and focus groups, it is possible to draw some broad overall conclusions. A key finding of all the studies is that on many detailed issues there is little overall public consensus, and a wide range of views are often found, with some people being happy for their information to be used in certain a way, while others would be most upset by it. The studies usually find that there is little or no correlation between peoples' attitudes and demographic variables such as age, gender, ethnicity, and so forth. This means that it is virtually impossible to develop a standard system that will satisfy all patients, or even to predict the specific privacy requirement of any particular individual. However, the studies do reveal some general trends.

Firstly, the studies have all confirmed that people do want restrictions to be applied to access to their information. While most people are happy for their information to be accessed by practitioners who are treating them, there is much less consensus about information being accessed by other professionals for secondary purposes, such as clinical research or organizational management and planning.

Secondly, many people also wish for different levels of security to be provided for different parts of the record so that some information could be made available on a 'need to know' basis only, and for information about particularly sensitive issues to be hidden from general access. For example, the Share with Care project (NHS, 2002b) found that 32% of respondents had at least some sensitive information over which they would like particular control through the use of a 'virtual sealed envelope'.

Thirdly, people are more willing to allow their information to be used for purposes other than their care if the information is first made anonymous by removing any identifying features such as their name and address. However, many people expressed some reservations about sharing even anonymous information with people other than health professionals, and many people feel that they should still be consulted about the use of their information, even when it is anonymous. The desire for this level of control contrasts with some current practices which are permitted under the privacy legislations of the UK, Australia, and New Zealand (Galpottage & Norris, 2005) which do not control the use of unidentifiable data.

MANAGING ACCESS

Overall, the studies revealed that people would like to have more control over the uses that are made of their information. However, the mechanisms that are provided in most current computer systems operate at too coarse a level of granularity and are not sophisticated enough to meet patients' requirements. Typically, access control mechanisms provide entire classes of users with access to the set of applications that they need to fulfill their role within their organization (Sandhu et al., 1996). Protection of an individual's privacy tends to be based upon the integrity of the staff who are relied upon not to abuse their positions of trust rather than being based upon technological safeguards within the systems. However, technological solutions have been investigated by the health sector, particularly in the areas of smartcards and e-consent models.

The health sector has been experimenting with the use of smartcard technologies (credit card-sized devices with on-board memories and possibly microprocessors) to store patient information and to authenticate practitioners since the early 1990s, and several systems in the United States and Europe (particularly France and Germany) have used smartcards to store patient information with varying degrees of success (see Reinhoff, 2003, for a comprehensive review). Having a patient's information stored electronically on a card can be very convenient for organizations, and it can also give the patients a high level of control over who has access to their information (Marschollek & Demirbilek, 2006). However, there are often problems with compatibility and inter-operability of systems that limit the uptake and use of the smartcard-based EHRs (Reinhoff, 2003). The relatively limited memory that is available on the card means that systems tend to only contain a very restricted amount of information (Chan, 2000), and an alternative use of smartcards is to provide a method of authenticating users and controlling their access to other systems over a network (Blobel, Pharow, Spiegel, Engel, & Engelbrecht, 2001; Marschollek & Demirbilek, 2006). Overall it seems that smartcards on their own will not resolve all of the issues of patient privacy, but they may provide some components of the solution.

Recently, some researchers have been focusing on the design of systems which are capable of capturing and recording a patient's privacy requirements and their consent to share information at a fine level of granularity. This information would then be used by the 'e-consent systems' to mediate information requests and to enforce a patient's requirements depending on the intended recipient, the nature of the information, and the proposed use of the information (Coiera & Clarke, 2004; Galpottage & Norris, 2005; Scott et al., 2004). However, as Coiera and Clarke (2004) note, the process of consulting with each patient about future access to information needs to be managed carefully so that it does not place an undue burden on health professionals. One solution that has been proposed to overcoming this problem during the clinical consultation is to enable patients to log onto the EHR system at some other time and to set or update the access permissions for their own personal information (Galpottage & Norris, 2005), thus giving patients the ultimate control.

CONCLUSION

Healthcare systems are exceedingly complex environments that often deal with highly personal and intimate issues; consequently, the security and privacy requirements for EHRs are more demanding than most other information systems applications. To date, the magnitude of the technical difficulties of implementing EHRs has tended to dominate the attention of researchers, and the privacy issues which are associated with the wholesale exchange of patients' information have received comparatively little attention. However, privacy issues are important since they will also influence the acceptance of EHRs; a failure to address them may ultimately lead practitioners or patients to boycott or reject such systems.

Patients' privacy needs to be protected in several ways: by technological systems that enforce security and can restrict access to information, by organization and cultural systems that emphasize and respect patients' privacy, and finally by a legislative framework that can be used to enforce patients' rights to privacy.

REFERENCES

AHRQ (Agency for Healthcare Research and Quality). (2006). *Electronic medical/health records.* Retrieved from http://www.ahrq.gov

Anderson, J. (2000). Security of the distributed electronic patient record: A case-based approach to identifying policy issues. *International Journal of Medical Informatics, 60,* 111-118.

Bath, P.A. (2004). Data mining in health and medical information. *Annual Review of Information Science and Technology, 38,* 331-369.

Blobel, B., Pharow, P., Spiegel, V., Engel, K., & Engelbrecht, R. (2001) Securing interoperability between chip card based medical information systems and health networks. *International Journal of Medical Informatics, 64*(3), 401-415.

Chan, A.T. (2000). WWW+smart card: Towards a mobile health care management system. *International Journal of Medical Informatics, 57*(2), 127-137.

Caldicott Committee. (1997). *Report on the review of patient-identifiable information.* London: NHS Executive. Retrieved from http://www.doh.gov.uk/confiden/crep.html

Carter, M. (2000). Integrated electronic health records and patient privacy: Possible benefits but real dangers. *Medical Journal of Australia, 172,* 28-30.

Chadwick, R. (1999). The Icelandic database—do modern times need modern sagas? *British Medical Journal, 319,* 441-444.

CHII. (2006). *Canada Health Infoway Incorporated.* Retrieved from http://www.canadahealthinfoway.ca

Coiera, E., & Clarke, R. (2004) E-consent: The design and implementation of consumer consent mechanisms in an electronic environment. *Journal of the American Medical Informatics Association, 11*(4), 129-140.

Davies, S. (1996). Dystopia on the health superhighway. *Information Society, 12*(1), 89-93.

Dick, R., & Streen, E. (1991). *The computer based patient record: An essential technology for healthcare*. Washington, DC: National Academy Press.

Didham, R., & Martin, I. (2004). *A review of computerised information technology in general practice medicine*. Retrieved from http://hcro.enigma.co.nz/website/index.cfm?fuseaction=articledisplay&FeatureID=040302

Ford, C.A., Bearman, P.S., & Moody, J. (1999). Foregone health care among adolescents. *Journal of the American Medical Association, 282*(23), 2227-2234.

Galpottage, P.A.B., & Norris, A.C. (2005). Patient consent principles and guidelines for e-consent: A New Zealand perspective. *Health Informatics Journal, 11*(1), 5-18.

Goldschmidt, P.G. (2005). HIT and MIS: Implications of health information technology and medical information systems. *Communications of the ACM, 48*(10), 69-74.

Hannan, T.J. (1998). Transporting an electronic patient record system across international boundaries—the lessons learnt. *Medinfo, 9*(1), 18-20.

Hannan, T.J. (1999). Variation in health care—the roles of the electronic medical record. *International Journal of Medical Informatics, 54*(2), 127-136.

Handy, J.A., Hunter, I., & Whiddett, R. (2001). User acceptance of inter-organizational electronic medical records. *Health Informatics Journal, 7*(2), 102-106.

Haux, R. (2006). Health information systems—past, present and future. *International Journal of Medical Informatics, 75*, 268-281.

HealthConnect Program Office. (2002). *Consent and electronic health records: A discussion paper*. Canberra: Commonwealth of Australia. Retrieved from http://www.healthconnect.gov.au/publications/

HealthConnect Program Office. (2003). *HealthConnect: Interim research report*. Canberra: Commonwealth of Australia. Retrieved from http://www.healthconnect.gov.au/publications/

Hunt, D.L., Haynes, R.B., Hannah, S.E., & Smith, K. (1998). Effects of computer-based clinical decision support systems on physician performance and patient outcomes: A systematic review. *Journal of the American Medical Association, 280*, 1339-1346.

Kissinger, K., & Borchardt, S. (1996). *Information technology for integrated health systems*. New York: John Wiley & Sons.

Marschollek, M., & Demirbilek, E. (2006). Providing longitudinal health care information with the new German health card—a pilot system to track patient pathways. *Computer Methods and Programs in Biomedicine, 81*(3), 266-271.

Marietti, C. (1998, May). *Will the real CPR/EMR/EHR please stand up*. Retrieved from http://www.healthcare-informatics.com/issues/1998/05_98/cover.htm

Ministry of Health. (2000). *The New Zealand health strategy*. Wellington: Author.

Ministry of Health. (2001). *From strategy to reality: The WAVE project*. Wellington: Author.

Mount, C., Kelman, C., Smith, L., & Douglas, R. (2000). An integrated electronic health record and information system for Australia? *Medical Journal of Australia, 172*, 25-27.

Mulligan, E. (2001). Confidentiality in health records: Evidence of current performance from a population survey in South Australia. *Medical Journal of Australia, 174*(12), 637-640.

NHIMAC (National Health Information Management Advisory Council). (1999). *Health online: A health information action plan for Australia*. Canberra: Commonwealth of Australia.

NHS (National Health Service). (2001). *Building the information core*. London: NHS Executive.

NHS (National Health Service). (2002a). *Electronic Record Development and Implementation Programme—EPDI*. London: NHS. Retrieved from http://www.nhsia.nhs.uk/erdip/pages/default.asp

NHS (National Health Service). (2002b). *Share with care! People's views on consent and confidentiality of patient information*. London: NHS. Retrieved from http://www.nhsia.nhs.uk

NZHIS. (1995). *Health information privacy and confidentiality.* Wellington: New Zealand Health Information Service.

OECD. (1980). *Guidelines on the protection of privacy and transborder flows of personal data.* Retrieved from http://www.oecd.org/dsti/sti/it/secur/prod/

Patientprivacyrights. (2006). Retrieved from http://www.patientprivacyrights.org

Payton, F.C. (2003). Data mining in health care. In J. Wang (Ed.), *Data mining: Opportunities and challenges* (pp. 350-365). Hershey, PA: Idea Group.

Reichertz, P.L. (2006). Hospital information systems—past, present and future. *International Journal of Medical Informatics, 75,* 282-299.

Reinhoff, O. (2003). *Integrated health data cards (smart cards): A primer for health professionals.* Washington, DC: Pan American Health Organization. Retrieved from http://www.ehealthstrategies.com

Sandhu, R.S., Coyne, E.J., Feinstein, H.L., & Youman, C.E. (1996). Role-based access control models. *Computer, 29*(2), 38-47.

Sankar, P., Moran, S., Merz, J.F., & Jones, N.L. (2003). Patient perspectives on medical confidentiality: A review of the literature. *Journal of General Internal Medicine, 18,* 659-669.

Scott, R.E., Jennett, P., & Yeo, M. (2004). Access and authorization in a global e-health policy context. *International Journal of Medical Informatics*, *73*, 259-266.

Shickle, D., Carlisle, J., & Wallace, S. (2002). *Patient Electronic Record; Information and Consent (PERIC)—public attitudes to protection and use of personal health information*. ScHARR Report Series No. 7, School of Health and Related Research, University of Sheffield, UK.

Sicotte, C., Denis, J.L., Lehoux, P., & Champagne, F. (1998). The computer based patient record challenges: Towards timeless and spaceless medical practice. *Journal of Medical Systems, 22*, 237-256.

USDHHS. (2001). *Information for health: A strategy for building the national health information infrastructure.* Washington DC: U.S. Department of Health and Human Services. Retrieved from http://www.hhs.gov

USDHHS. (2006). *HIPPA medical privacy—national standards to protect the privacy of personal health information*. Washington DC: U.S. Department of Health and Human Services. Retrieved from http://www.hhs.gov

Westin, A.F. (2003). Social and political dimensions of privacy. *Journal of Social Issues, 59*(2), 431-453.

Whiddett, R., Hunter, I., Engelbrecht, J., & Handy, J. (2006). Patients' attitudes towards sharing their health information. *International Journal of Medical Informatics, 7*(7), 530-541.

Willcox, D. (1995). Health scare. *Computing,* (October 19), 28-29.

Woodward, B. (1995). The computer-based patient record and confidentiality. *New England Journal of Medicine, 333,* 1419-1422.

KEY TERMS

Computer-Based Patient Record (CPR): Automated patient record designed to enhance and support patient care by providing access to complete and accurate data and to bodies of knowledge.

Confidentiality: Tacit understanding that information disclosed to care providers will remain protected.

Consent: Process by which an individual authorizes medical interventions or the processing of his or her information based on an informed understanding of the expected outcomes.

Data Mining: A set of techniques that explore large collections of data to discover new or unexpected relationships.

E-Consent System: Advanced computer-based system that can record a patient's consent and use it to mediate information requests from other systems.

E-Health: A term that encompasses a wide range of healthcare-related activities that involve the exchange of information using electronic communication systems.

Electronic Health Record (EHR): Integrated collection of health information relating to an individual's lifetime, stored in electronic format.

Electronic Medical Record (EMR): The health information for a patient contained within a given institution or organization stored in electronic format.

Master Patient Index (MPI): Module of a health information system used to uniquely identify a patient.

Smartcard: Credit card-sized device with on-board memory and possibly a microprocessor. Sometimes used to store a patient's personal information.

ENDNOTE

[1] There are no universally accepted definitions of the terms CPR, EMR, and EHR. The meanings of the terms have been constantly evolving and definitions often overlap; see Marietti (1998) for further discussion. This article will adopt the term EHR.

This work was previously published inEncyclopedia of Information Ethics and Security, edited by M. Quigley, pp. 534-541, copyright 2007 by Information Science Publishing (an imprint of IGI Global).

Chapter 7.11
A Trusted System for Sharing Patient Electronic Medical Records in Autonomous Distributed Health Care Systems

Surya Nepal
CSIRO ICT Centre, Australia

John Zic
CSIRO ICT Centre, Australia

Frederic Jaccard
CSIRO ICT Centre, Australia

Gregoire Kraehenbuehl
CSIRO ICT Centre, Australia

ABSTRACT

The problem of assuring secure and confidential access and transfer of medical records in health-care facilities can be partitioned into (a) secure storage and access of electronic records within a facility and (b) secure transfer of electronic records between facilities. To address the first issue, we propose a new tag-based data model for representation of electronic medical records along with patients' policy statements. This model helps to categorize the patient information as well as express patients' consent for a variety of domains, such as individual practitioner and facility. To address the second issue, this paper proposes a way of establishing a trust relationship between two interacting parties based on emerging trusted computing technologies, and describes its application and implementation in an electronic healthcare system. Our proposed

solutions have been demonstrated by developing a prototype system utilizing trusted computing components.

INTRODUCTION

There is a growing need for sharing patient records among health care providers and their associated facilities, in order to provide the best care and clinical outcomes for patients. Due to the increasing use of information technologies in health care industries, patient records are often produced and shared in an electronic form (Choudhri, Kagal, Joshi, Finin, & Yesha, 2003). The coordination of an individual's health care relies on the sharing of personal health information among health care providers and their facilities, such as clinics, test laboratories, and hospitals. It is well known that there are potential benefits and risks associated with the ease with which patients' electronic medical records may be shared (Task Force on Medical Informatics (TFMI), 1996). The ease of sharing information and its rapid dissemination by information technologies underscores the need to assure that the privacy and security requirements of the patient are met at all times and for all possible situations. For example, patients may not want to share or transfer any of their personal health information, without their knowledge and consent, and may want to retain the rights to both access and transfer of their information.

In order to address the issue of control of private medical information, the CSIRO ICT Centre was involved in the development of an electronic consent (e-consent) model (O'Keefe, Greenfield, & Goodchild, 2005) within the context of the Australian Government Department of Health and Aging (DoHA) Electronic Consent Project (Australian Government Department of Health and Aging (DoHA) Project, 2002). The term e-consent was coined to refer to a mechanism through which patients can express their consent policies for electronic records that are accessed and shared between health care facilities. The e-consent model was developed and demonstrated to stakeholders from the Australian health care sector for work commissioned by the DoHA. The most important outcome of the project is the concept of "placeholders," which were used as a means for ensuring privacy-preserving, anonymous electronic record transfer protocols. We make the following two observations on this e-consent model (referred to as the DoHA e-consent model).

First, any underlying data model must support adequate default consent statements and policies to ensure that the e-consent model can secure the privacy and confidentiality of medical data within a facility. The hierarchical data model used by the DoHA e-consent model has limitations on expressing a default policy set, as well as categorizing electronic patient records. This limitation led to the development of our new tag-based data model. In our model, each electronic medical record has a number of policy tags, which we call *e-tags* (electronic tags). Each e-tag has two fields: *category* and *policy*. The category field categorizes records into different groups, such as heart, head, and AIDS. The policy field, which we call an *e-co* (e-consent), contains a number of policy expressions. Unlike role-based access control (RBAC) (Crook, Ince, & Nuseibeh, 2003; Reid, 2003a) and the DoHA e-consent model, our approach allows the specification of access and transfer permissions in terms of (a) roles, (b) health care facilities, (c) health care providers, and (d) categories of information.

Second, in the DoHA e-consent model, the sender facility has to trust the receiver facility (and vice versa). That is, the sender as well as authenticating the identity of the receiver has to rely on the receiver having the right software and hardware system components and configuration to enforce the sender's policies on privacy and confidentiality. Similarly, the receiver also must ensure that it is always in the correct state and configuration for accepting any incoming record. Both facilities need to ensure and enforce the patients' consent policies for their electronic records, when these records are accessed and transferred. The DoHA e-consent model relies upon a person at the receiver facility entering a patient's consent requirements in the receiver facility's system. These limitations led to the development of a mutual attestation protocol to alleviate the mutual trust issue between senders and receivers.

The remainder of the paper is organized as follows. The second section presents a review of related work. We then present a motivating scenario in the third section. Our tag-based data model is presented in the fourth section. The fifth section presents two protocols: one for mutual attestation and one for transfer. We then illustrate our data model and mutual attestation protocol by describing the developed architecture and prototype implementation in the sixth section. The final section presents the concluding remarks.

RELATED WORK

In this section, we first discuss work related to protecting patient privacy in health informatics in general and then discuss the work closely related to our proposed model. This is followed by an overview of related work on remote attestation.

Security in Health Care Systems

Security and privacy of patient electronic medical records needs greater attention because of the wide adoption of electronic health care systems. Huston (2001) discusses the general security concerns of implementing electronic medical records and the technological and administrative tools that are available for safeguarding these records. Stein (1997) presents several scenarios in the use of electronic medical records and highlights the threats and opportunities in their adoption. One of the major privacy concerns is the secure transfer of electronic medical records from one service provider to another. TFMI (1996) discusses some issues related to the transfer of medical records. Chadwick and Mundy (2004) look at the security requirements for electronic transfer of prescriptions from the perspectives of confidentiality, integrity, and availability.

One of the most common privacy and security preserving techniques for sensitive medical data is the use of simple access control lists (ACLs). Evered and Bogeholz (2004) present a case study of the access control requirements for a health information system in a small aged-care facility. The study focused on the use of a static per-method ACL and found that the method was inadequate because the policy constraints became complex even for the small system considered in the case study. Effective use of an ACL system requires a carefully developed data model.

Access Control Data Model

Alternatives to simple ACLs also have been investigated and developed. Reid et al. (2003) examined the use of RBAC and found that the range of access policy expressions supported by RBAC was inadequate. Further, the paper proposed a model that supported access control through a consumer-centric role called the "care team." The advantage of this model was that a subgroup of entities within a role could be granted explicitly or denied access to health information. Motta and Furuie (2003) extended the RBAC reference model by introducing additional contextual authorization. The authorization module not only used positive and negative authorization, but also user affiliation, time, and location of access, user and patient relationship, patient status, and so forth.

Most of these hierarchical RBAC models cannot adequately represent the matrix nature of medical information. Khayat and Abdallah (2003) present a formal model for flat role-based access control, which we see as being closer to our model. Some of the access control problems are simplified because of the flat, nonhierarchical model. However, this model does not consider the problem-oriented approach, where the patient's medical condition may be divided into a list of discrete problems, such as diabetes, coronary artery disease, and lower back pain (Stein, 1997). The problem with this model is that only roles are considered. Our approach has overcome this problem by flattening, not only roles, but also the categories of medical records and medical information.

Mutual Trust

One way of establishing trust between interacting parties is to use an attestation protocol. Most of the earlier work on attestation protocols relied on software-based security, where the entire security feature is implemented in software systems. These solutions are often sufficient, but not always adequate (Trusted Computing Group (TCG), 2003). Classically, problems with key storage and management, and the location where secret keys may be stored in an accessible persistent store, remain open issues for most trusted system implementations. In recent times, integrating hardware-based solutions into software-based security systems has gained momentum because of a variety of political and economic requirements for enabling electronic commerce (e-commerce), such as those required to implement digital rights management systems. A general, hardware-based solution has been proposed by the TCG (2003). TCG first proposed a remote attestation protocol, using a trusted third party, called a privacy certifying authority (Privacy CA). One of the problems associated with a Privacy CA model is the lack of anonymity. In order to address this problem, direct anonymous attestation (DAA) was proposed as an alternative to the Privacy CA model (Brickell, Camenisch, & Chen, 2004). Yung (2005) presented a thorough analysis of implications of TCG-style remote attestation including Privacy CA, DAA, and combinations of both.

We next present a review of subsequent developments in remote attestation. There have been recent works on architectures that support enforcement of an owner's policy in client applications by attesting to the platform and integrity of the requesting applications (Sailer, Jaeger, Zhang, & Doorn, 2004a; Sandhu & Zhang, 2005). These researchers have developed and implemented such systems in a Linux environment, using the TCG remote attestation protocol and integrity measurement technologies. Several problems have been identified related to remote attestation (Halder & Franz, 2004; Reid et al., 2003). These include, for example, that remote attestation (a) makes no claims, statements, or assurances about program behaviors; (b) is static, inflexible, and inexpressive; (c) makes upgrades and patches to executable programs or operating systems difficult; and (d) is fundamentally incompatible with a widely varying, heterogeneous computing environments. Some techniques have been proposed to overcome these limitations, including semantic attestation (Halder &/ Franz, 2004), property-based attestation (Sadeghi & Stüble, 2004), virtual machines (Barham et al., 2003; Garfinkel, Pfaff, Chow, Rosenblum, & Bonch, 2003), and a fine grain attestation mechanism that measures only the required code (Shi, Perrig, & Van, 2005).

The most relevant work to our paper is the integrity measurement and attestation mechanism proposed by IBM (Sailer et al., 2004b). To perform an integrity measurement, IBM has modified the Linux boot loader and kernel to maintain the *chain of trust* from trusted platform module (TPM) to operating system to applications. The kernel maintains a list that represents the history of loaded processes and libraries. During attestation, the processes list is sent to the challenger, together with the reported platform configuration register (PCR) values. The challenger can then verify that whether the list is valid or not.

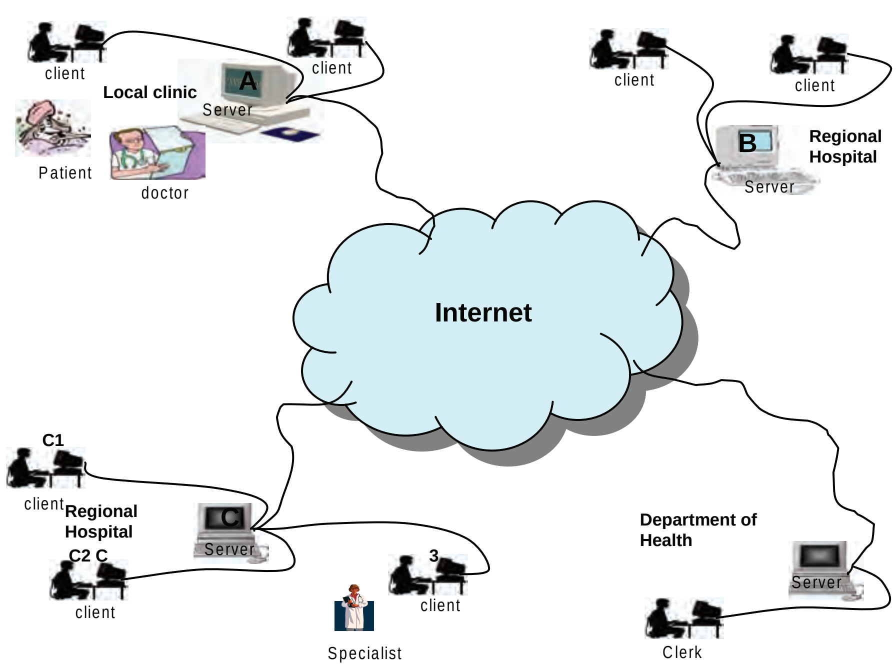
Our implementation and protocols differ from the IBM proposal as follows. First, our implementation is based on a Windows environment, rather than Linux, to explore what is achievable within the current, commercially installed infrastructure. Second, our approach suppresses the need for the facility to load the processes in a particular order, as the order is impossible to maintain in a dynamically loaded system, where software components are loaded and unloaded under application control. Third, our approach provides flexibility since each process can be specified as being in one of three states: *may*, *must*, or *must not* be loaded for the attestation to succeed. Finally, our approach can be used to force a facility to load the correct versions of the software, as we permit multiple valid digests for a process in order to support multiple versions of the same software.

A MOTIVATING SCENARIO

We consider an example from electronic health care environment that includes two hospitals, one medical clinic, and a governmental department of health as shown in Figure 1. The department of health issues credentials and permissions to operate the health care facilities, such as hospitals and medical clinics.

Though the basic set of policies are established by government rules and legislation, each facility may implement these policies differently. That is, each may have a different set of policies for different category of information. For example, Hospital A may have a different set of policies for AIDS-related records to that of Hospital B. Any data model must support a variety of default access policies for hospitals so that all medical records within that hospital are subject to these

Figure 1. Distributed health care environment



policies. However, it is a fundamental assumption of our paper, and the prior e-consent work, that patients are the ultimate owners of their own electronic medical records. Consequently, patients' privacy and consent policies may differ from those of a particular hospital. For example, patients can define a personal policy that all of the AIDS-related records are accessible to a personal doctor, but no other doctor in the hospital can access it, except in case of emergency. Similarly, patients' families can define their own policies for family members. For example, only family doctors can access immunization records of all family members. It is clear that the data model must support the definition of default family and patient policies. Similarly, patients' policies on AIDS-related records may say that only family doctors can access them, but they may want a particular blood test to be examined by an AIDS specialist. Furthermore, patients may not want to disclose their AIDS-related records to their family doctors though the families have policies that all records of family members are accessible to the family doctors.

Our access policies are defined at different granularity, based on entities, such as hospitals, patients, families, and medical records. However, some of the policies are defined for a group of medical records, such as AIDS-related records. This means medical records need to be *categorized* so that it is possible to define policies for a set of records within a category. For example, hospital default policies are normally expressed on *roles* (within the facility), rather than on *individuals.* As an illustration, Hospital A may have a policy that any individual doctor or nurse in an *emergency doctor role* can access all medical records and information. In contrast, patients' policies are less likely to depend on roles and may express their policies in terms of *individuals,* such as "grant access to AIDS-related records to Dr. Smith."

The hospitals may share patient medical information with each other to provide better services. Similar to access policies, we define *transfer policies* at the hospital, family, patient, and record levels. The hospital may have a certain policy for transferring information to another hospital. However, sharing of information is only possible if patients have given their consent. Therefore, the patients also must be able to define transfer policies. In addition to the specification of access and transfer policies, we also need to have policy enforcement mechanisms in place. This includes establishing a trust relationship between the clinic and hospital such that nothing undesirable (in terms of policy violations) ever occurs.

In summary, the data model must support (a) categorization of information, such as AIDS and heart, (b) policy expression for access and transfer, (c) default policy expression for hospitals, patients, and families, (d) default policy expression for different categories of information, and (e) policy expressions for both individuals and roles. Additionally, we need to define transfer and attestation protocols such that when the records are transferred from one facility to another, the defined policies are always enforced.

TAG-BASED DATA MODEL

This section discusses the tag-based data model and how it provides the features that are needed for modeling the health information scenario as discussed in the third section.

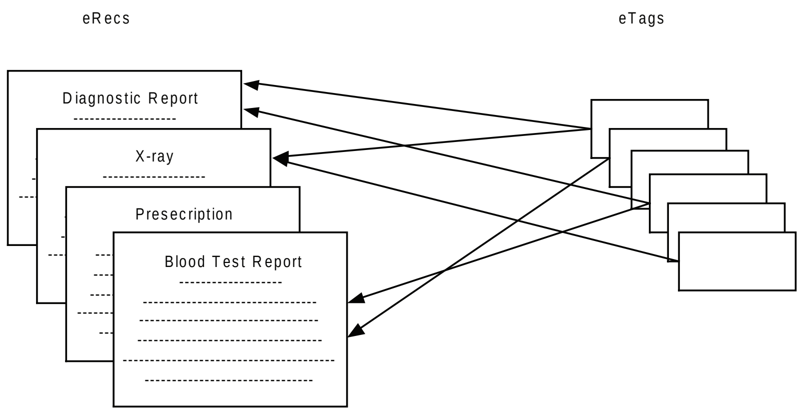
Data Model

An electronic record (e-rec) is a fundamental unit of information. It could be a diagnostic report, X-ray image, or prescription. An e-rec may have a number of electronic tags (e-tags) attached to it. Figure 2 shows the structure of an e-tag. Each e-tag has two fields: a *category* and a *policy expression* that we refer to as an e-consent. The category field consists of category *type* and *value.* The type determines whether the category is related to *patient, family,* or *facility.* The value

Figure 2. eTag structure

<Category Type, Value>	<eCo>

Figure 3. Relation between eRecs and eTags in the data model



provides the categorization within the category type. The e-consent field contains a set of access and transfer policies for the category.

Each e-tag can be grouped into one of three broad classes:

a. A special NULL category with a nonempty set of policies used to express record-specific policies. Such e-tags are defined for a specific medical record and are attached to it.
b. A category with an empty set of policies is used for categorization purposes.
c. A category with a set of policies used to group a set of records and specify group-related policies.

Figure 3 shows the many-to-many relation between e-tags and e-recs. Because this relation is many-to-many, each e-rec can be categorized into multiple groups by attaching multiple e-tags into it. Access and transfer of an e-rec is determined by the policies defined in its attached e-tags.

The category information is also sensitive, and unauthorized access to the e-tags attached to e-recs may reveal private and sensitive materials. For example, an e-tag with an AIDS category may reveal information that the patient has been visiting clinics for AIDS-related consultations. Our e-tags have an e-consent that applies both to the e-rec to which it is attached as well as to the e-tag itself. The implication of this is that the e-tag (and its associated e-rec) is inaccessible without suitable authorization, and so, for example, a medical practitioner will not have access to view the e-tags if the e-consent defined in the e-tag denies access to the practitioner.

We next provide some semiformal definitions in order to clarify the structures and relationships in the tag-based data model.

The following definitions assume the existence of some basic types:

STRING: the set of all finite-length strings composed of letters from an alphabet;
NAT: the set of natural numbers;
IMAGE: the set of all diagnostic and other images (charts, plots, etc.);
FTYPE: the set of types of facilities, such as hospitals, laboratories, and clinics;
CTYPE: the set of domain specific category types (such as facility, family, patient, and NULL);
ROLE: the set of occupational roles within a health care system, such as doctor, patient, nurse, specialist, and intern;
OPERATION: the set of permitted operations on an e-rec; and
POWERSET(S): the power set of a set S; that is, the set of all subsets of S.

We also define some identifying keywords associated with each of the defined entities in our model.

Definition 1: Subject (keyword: *Sub*)

A *subject* refers to a particular named human being or user. The identification is based on a pair: a key word *user* and unique name, expressed as a string taken from *STRING.*

$subject \in SUBJECT$
$SUBJECT = \{id \in STRING \mid (Sub, id)\}$

Definition 2: Facility (keyword: *Fac*)

A *facility* in our model is a named facility. *FACILITY* represents the set of all such possible facilities.

$facility \in FACILITY$
$FACILITY = \{f \in FTYPE, id \in STRING \mid (Fac, f, id)\}$

Definition 3: Electronic record (e-rec) (keyword: *Erc*)

An *e-rec* is a piece of information in the system for which access policies are defined, such as X-ray image, blood sample, and prescription. *EREC* represents the set of all such pieces of information.

$eRec \in EREC$
$EREC = \{e \in STRING \times IMAGE \mid (Eec, e)\}$

Definition 4: Roles (keyword: *Rol*)

Role is an occupational position in our system taken from a predefined set *ROLE.*

$role \in ROLE$
$ROLE = \{r \in ROLE \mid (Rol, r)\}$

Note that there is a many-to-many relationship between roles and subjects. A *subject* can be assigned to multiple *role*s, and, similarly, a *role* can be assigned to multiple *subjects.*

$ROLE \leftrightarrow SUBJECT$

Definition 5: Principal (keyword: *Prn*)

A *principal* is an entity for which an access can be granted.

$principal \in PRINCIPAL$
$PRINCIPAL = \{\bigcup \in SUBJECT \cup ROLE \cup FACILITY \mid (Prm, \bigcup)\}$

Definition 6: Operations (keyword: *Opn*)

Operations are the tasks that can be performed on an *e-rec*.

$operation \in OPS$
$OPS = \{o \in OPERATION, e \in EREC \mid (Opn, o, e)\}$

Definition 7: Category (keyword: *Cat*)

A *category* is either a tuple consisting of predefined types and associated values (effectively set of freely defined strings) or a special *NULL* category. Then we define the set of all categories as follows.

$category \in CATEGORY$
$VALUE \subset STRING$
$CATEGORY = \{k \in CTYPE \times VALUE \ \cup \{\mathrm{NULL}\} \ | (Cat,k)\}$

Definition 8: Policy (keyword: *Pol*)

A *policy* is a statement about operations that can be performed on an *e-rec* based on *principals*. Each policy has a time stamp associated with it (modeled as a natural number, *NAT*) that is used for conflict resolution and auditing purposes. *POLICY* is the set of all policies and we define it as follows.

$policy \in POLICY$
$POLICY = \{time \in NAT,\ \ OPERATION \times PRINCIPAL \, | (Pol,\ time,\ d)\}$

Definition 9: E-consent (keyword: *Ecn*)

An e-consent is a subset of all possible policy statements including the *NULL* policy, and *ECO* is the set of all policy statements.

$eCo \subseteq ECO$
$ECO = \{sp \in POWERSET(POLICY \ \cup \ \{\mathrm{NULL}\}) \ | (Ecn,sp)\}$

Definition 10: E-tag (keyword: *Etg*)

An e-tag is an e-consent defined for a category. *ETAG* represents all possible e-tags, except the one where all elements are empty.

$eTag \in ETAG$
$ETAG = \{t \in CATEGORY \times ECO \, | (Etg,t)\}$

When an e-tag is attached to an e-rec, it categorizes the e-rec and protects its access.

Definition 11: Object (keyword: *Obj*)

An *object* is a complete e-rec, along with its attached e-tags. *OBJECT* represents all e-recs with attached categorization and policy information.

$object \in OBJECT$
$OBJECT = \{ee \in ETAG \times EREC \, | (Obj,ee)\}$

Policy Expression Language

Our e-consent is used to express the transfer and access policies of a particular e-rec. One could define an e-consent using standard policy expression languages such as XACML (OASIS, 2005) or EPAL (Paul, Satoshi, Karjoth, Powers, & Schunter, 2003). However, most of the current policy expression languages primarily are used for expressing *access* policies, and languages that also are able to express *transfer* policies are a recent development. Examples of such languages are found in OMA 2.0 (Open Mobile Alliance (OMA), 2004) that came out of the family domain work within Motorola Research Laboratories (Messerges & Dabbish, 2003). These are rich, expressive languages allow generalized application to any digital rights management (DRM) application. However, as we did not require their full capabilities for this application, we decided to define and use our own simple application-specific policy expression language that supports both access and transfer policies.

Figure 4 shows our policy expression language. Each policy has a time stamp representing the time of creation of the policy.

The transfer policy grants or denies transfer to a certain facility (or hospital). The access policy grants or denies access to health practitioners (or subject), their roles, or facilities. The access permission can be granted with appended rights for health care practitioners or their roles as defined in the facility. A doctor can access an e-rec if, and only if, the access policies (a) allow access to the subject or to the role and (b) allow access to the facility. That is, both the facility and practitioners need to have access permission in order to access the medical records. Similarly, a doctor can transfer an e-rec if, and only if, (a) the

Figure 4. Policy expression language

```
policy :=
   policy_TRANSFER | policy_ACCESS
policy_TRANSFER :=
   (‘grant’ | ‘deny’) ‘transfer to’ (FACILITY)
policy_ACCESS:=
   policy_ACCESS_GRANT | policy_ACCESS_DENY
policy_ACCESS_GRANT:=
   ‘grant’ ‘access to’
   (((SUBJECT | ROLE)[‘with append right’]) | FACILITY)
policy_ACCESS_DENY :=
   ‘deny’ ‘access to’ (SUBJECT | ROLE | FACILITY)
```

access policies allow access to the subject and the facility and (b) transfer policies grant transfer permission to the destination facility.

Policy Enforcement and Conflict Resolution Mechanism

In order to determine whether a principal can access an e-rec or not, all the policies listed in the e-consent of all attached e-tags of an e-rec must first be evaluated. Not unexpectedly, there is a reasonable chance that there will be conflicting policies defined in different e-tags. For example, a patient's policy may state that the information is only accessible to a specific doctor at a particular, nominated hospital. However, the nominated hospital may have a conflicting policy to that of the patient, and allow any doctor or nurse, while acting in an emergency role at that hospital, to access all of the patient's information.

Our conflict resolution mechanism is based upon the following ordering on e-tag category types.

$$Facility \subseteq Family \subseteq Patient \subseteq NULL \quad (1)$$

This order ensures that any specific policies attached to a particular e-rec have a higher priority than the default policies that are attached to all the e-recs of the patient at a particular facility. Similarly, a patient default policy has priority over the facility default policy. This combination of priorities ensures that the patient always has a final say about access to their personal health information.

We now define our prioritized multistep conflict resolution mechanism as follows.

1. *Higher priority wins:* A check is first performed against the order in Equation 1. Unfortunately, this first step cannot resolve all conflicts, since a single e-consent may contain multiple policies. Similarly, an e-rec can have multiple e-tags at the same priority level. If resolution is not possible at this step, then the resolution moves onto step 2.
2. *More specific wins:* The conflict that results from multiple e-tags, possibly with the same priority, is resolved using a principle of "more specific wins." For example, a policy concerning a doctor wins over a policy concerning a role because a doctor is more specific than the role. Once again, this does not cover the situation where the policies are of the same priority and have identical specificities. If this step cannot resolve the conflict, we move onto step 3.

3. *Most recent wins:* In the case where step II cannot resolve the conflicts, we use the principle of "most recent wins," based on the time stamp attached to each policy. Finally, if the conflict still cannot be resolved using the time stamp, the resolution must be finalized at step 4.
4. *"Deny" wins over "grant":* If two policies have the same time stamp, the policy with "deny" wins over the policy with "grant."

Despite the fact that we expect that this conflict resolution mechanism will be able to resolve all of the conflicting policies, ultimately, for those that cannot be resolved by using this scheme, any implementation should provide for human intervention.

SECURE TRANSFER PROTOCOL

Secure transfer of an electronic medical record consists of two parts. First, mutual trust between interacting parties must be established so that they can be enforced to implement based on agreed access policies. Second, secure communication of electronic medical records between two parties using untrusted and unsecure public network infrastructure needs to be in place. We next describe the mutual attestation and transfer protocols we have developed to address these issues.

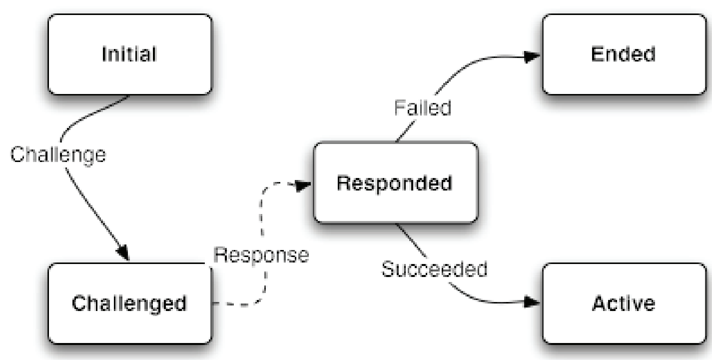
Mutual Attestation Protocol

TCG is an industrial standards body that produces a variety of specifications aiming to provide the best-practice software and hardware-based solutions to "help users protect their information assets (data, passwords, keys, etc.) from compromise due to external software attack and physical theft (2003)." One of the TCG specifications defines the cryptographic microcontroller hardware called the *trusted platform module* (TPM) that serves as a root of trust for authenticating a host's software configuration and protecting data by never releasing a root key outside itself.

The TPM chipset contains a set of *PCRs* that are used in measuring the characteristics and properties of the platform environment, such as integrity, system configuration, and executing states. For example, platform integrity (measured from boot time to the time at which applications are loaded) is kept in one of the PCRs. Any changes in the environment are reflected in a change to the PCR value (the update applies the cryptographic encryption algorithm SHA-1 to the current values concatenated with the new measured content).

The measured environment is then exchanged between the two parties wishing to establish a

Figure 5. TCG Remote attestation protocol



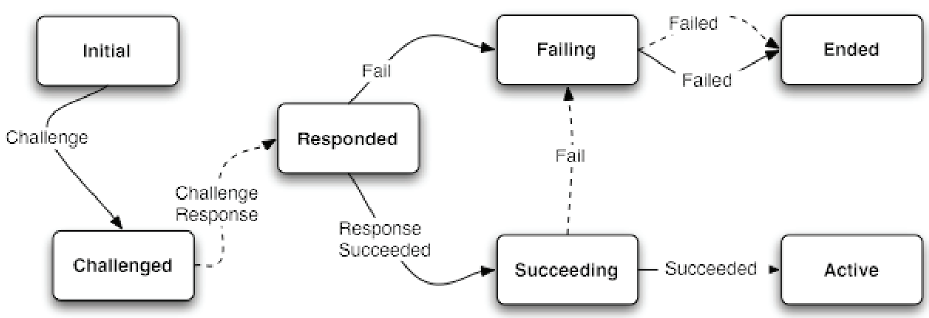
trust relationship, using a challenge-response attestation protocol. This attestation protocol is used to verify authenticity of the receiving party as shown in Figure 5. During attestation, the challenger platform (sender) sends an attestation challenge to the receiving (challenged) platform. The challenged platform then assigns one or more PCR values, using an *attestation identity key* (AIK) and sends this back to the challenger as a response. The AIK is protected by the TPM hardware[1] and the assigning is usually done using a special "*quote*" operation. The challenger then verifies that the received assigned values are the expected values, and so determines the authenticity of the challenged platform. After the successful attestation, the protocol reaches an active state, that is, the challenger trusts the platform configuration of the challenged facility and that the facility has a valid TPM. The challenger can then send encrypted messages to challenged facility, knowing that only the challenged facility can decrypt them (through the use of its TPM).

The TCG attestation protocol can support the remote authentication of a specific, challenged platform. However, in a general distributed health care environment, it is necessary to ensure that both the health care facilities engaged in a transaction have a valid TPM and that the facilities are in the correct software configuration. That is, the sender can trust the receiver, and the receiver can trust the sender for each possible transaction.

We extend the current remote attestation protocol and propose a mutual attestation protocol that permits two facilities to prove their validity to each other as shown in Figure 6. The full and dotted lines indicate the messages sent by sender and receiver facilities, respectively. The mutual attestation protocol is as follows. First, Facility A sends an attestation challenge to Facility B. Facility B then assigns one or more PCR values, representing its platform state using an AIK, and sends these values back to Facility A as a response, along with its attestation challenge to Facility A. Facility A then verifies the received assigned values with the expected values to determine whether the challenge is successful or not.

If the challenge is successful for Facility A, it then assigns one or more PCR values, representing its platform state using an AIK, and sends it/them to Facility B as a response, along with the successful result for its response to the earlier challenge. The receiver facility verifies the received assigned values with the expected values to determine whether the challenge is successful or not. The successful mutual attestation ends at the *active* state, where they can exchange electronic patient records, whereas unsuccessful attestation ends at the *ended* state.

Figure 6. Mutual attestation protocol



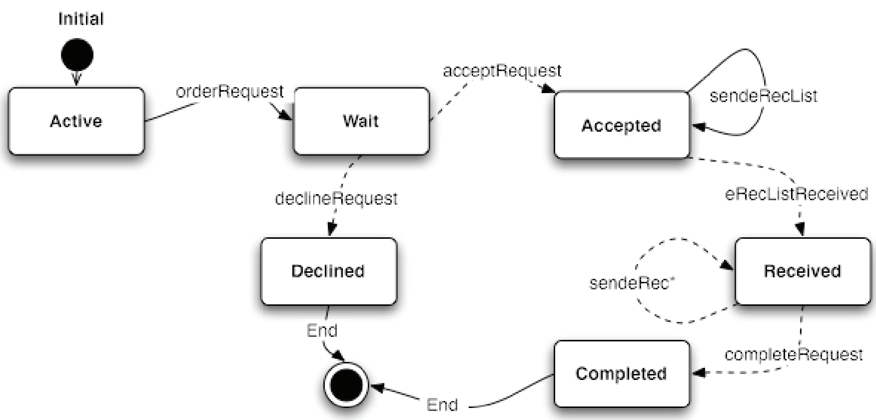
Transfer Protocols

Health care facilities exchange electronic medical records only when both facilities reach the active state. We next describe the transfer protocol we have defined for this purpose. Our transfer protocol is adapted and extended from the protocol proposed by O'Keefe et al. (2005). After successful mutual attestation, a doctor at a facility initiates a transfer by first creating and then sending an *order request* to another doctor at another facility as shown in Figure 7. The facility that initiates the transfer is called the originating facility, and the facility that receives the order request is called the destination facility. The order request is structured and contains the following information:

- Request identifier: This is a unique identifier for each request. The context of the interaction between facilities is always maintained by accompanying the protocol and data messages with the request identity.
- Short description: This is a string, for example, used by a doctor for description or other comments about a particular case. Note that this description is a permanently recorded and cannot be revoked or otherwise removed.
- Originating doctor and destination doctor identities: This is necessary to deliver messages to the right doctors.
- Policies: The originating facility also sends a set of policies that are applicable to medical records that will be created in the destination facility as part of the response to the request. These policies will be attached to any piece of information created as a response.
- Patient and family e-tags: It is necessary to register the patient in the destination facility in case they are not already in its system. This registration associates a repeatable, but anonymous, identity associated with the transferred patient record. Consequently, the e-tags are not transferred as part of the request in anonymous transfer but are replaced with anonymous tags (containing the patient and family policies and de-identifying the transferred information).

If the request is accepted by the destination facility, an *accept request* message is sent back to the originating facility. The originating facility can then fetch all the e-recs corresponding to the request in batch mode (resulting from the originating facility issuing an *e-rec list* message). All e-recs are preprocessed before being sent (as

Figure 7. State diagram illustrating the transfer protocol in originating facility



described later in our prototype system). In case of anonymous transfer, the name of the patient and the family e-tags are replaced with unique identifiers. As soon as the e-recs are received by the destination facility, the records are postprocessed (as described later in our prototype system) and stored in the facility. The destination facility then sends an acknowledgement message *e-rec list received* to the originating facility.

In case a response is required to a transfer, the destination facility can send back any number of e-recs by issuing a *send e-rec* message multiple times. Theses e-recs are similarly preprocessed and postprocessed by the destination and originating facilities, respectively. The send e-rec message contains, along with the e-rec, the request identifier. In postprocessing anonymous transfers, the anonymous identifier tags are replaced by the identifying family and patient e-tags. At the end of every transfer protocol, the destination facility sends an *end* message.

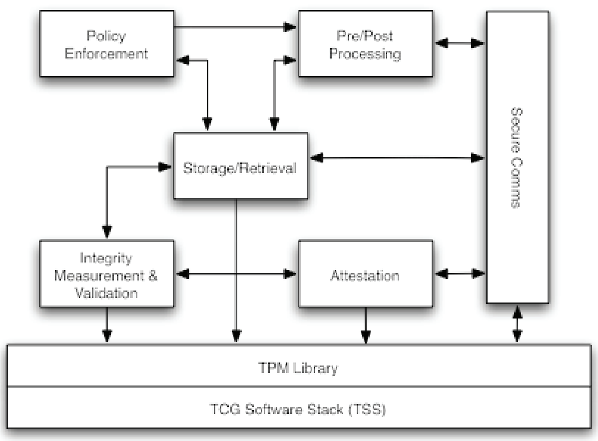
Figure 7 illustrates the transfer protocol for the originating facility for both cases—with and without a response. There is a complementary state diagram for the receiving facility, which is omitted here due to space limitations. The solid arrows represent sent messages, and the dashed arrows the received messages. The superscript star takes on the usual regular expression meaning and represents zero or more transitions (messages). In the case of a transfer without a response, such as transfer of a patient from one facility to another, the starred message is not received at all. In the case of a transfer with a response such as a referral of a patient to a specialist, it is possible to receive multiple e-recs in response to the referral.

PROTOTYPE IMPLEMENTATION

Figure 8 depicts the overall architecture of the "MedicClient" demonstration system, which is based on the concepts described in the fourth and fifth sections. The system has six major components.

- Policy enforcement: The system is responsible for enforcing policies and resolving conflicts, while accessing and transferring medical records. It is also responsible for generation and maintenance of transaction logs (required for auditing).
- Integrity measurement/validation: The system is responsible for measuring the current environment of the host computer

Figure 8. MedicClient architecture



and verifying that the measurements sent by other facility are as expected, and so can be trusted.

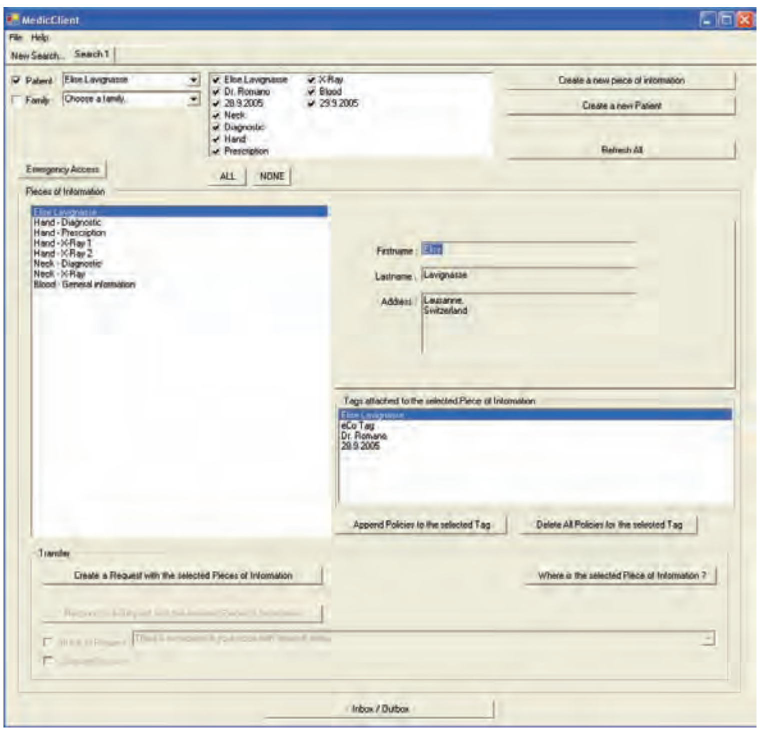
- Secure communication: The system is responsible for encrypting the outgoing information and decrypting the incoming information.
- Attestation: The system is responsible for authenticating and determining the identity of the remote facility.
- Pre/post processing: The system is responsible for processing e-tags for transfer.
- Storage and retrieval: The system is responsible for storing and retrieving e-recs from SQL databases.

Doctors access, create, and share medical information by use of the MedicClient software. Figure 9 shows a screenshot of the main interface for the system. Once a doctor is authorized to use the system and selects a patient, MedicClient displays the health-information view for the patient, as shown in Figure 9. We omit the details of operation of the system here due to the limitation of the space and focus on the technical details. Refer to Nepal, Zic, Jaccard, & Krahenbuehl (2006) and Nepal, Zic, Krahenbuehl, & Jaccard (2006) for a more comprehensive explanation of the system. We next describe the implementation details of major components.

Storage and Retrieval

Electronic medical records and all related data needed for the functioning of components are stored in a persistent, securely encrypted store and implemented using a SQL server. We use ADO.NET classes to connect data sources as well

Figure 9. MedicClient user interface for Dr. Romano



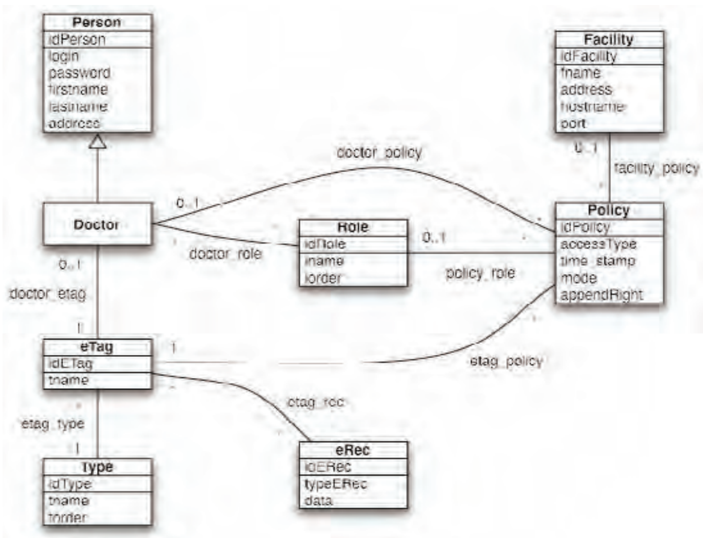
as to retrieve and update stored data. Figure 10 shows the relationships between the various entities involved in the tag-based data model. It can be seen from this figure that a policy is defined on the roles and facilities as well as individual doctors. It is worth noting that the doctor's e-tag is a *NULL* policy, as it does not have any relationship with the policy entities. The entities and their relationships were used to create the SQL tables that were held on the server. We utilized the trusted computing technology provided by the TPM microcontroller and its associated software stack to ensure the privacy and security of the SQL data. Database encryption and security is an area of research by itself, and this paper does not claim any contribution to this area. Our work was guided by the requirement that a patient's medical record privacy must be protected, which in turn suggests that at least the information held in a patient's e-rec table entry must be encrypted. This decision allowed a simple database implementation, as only the e-rec table entry data needs to be secured. Decryption of the information in an e-rec could be done without using another access to the database.

The e-rec data was encrypted using Triple DES. This requires the use of a symmetric key created at the time of installation of the MedicClient application. An asymmetric key is created and registered at the same time and is used for sealing the symmetric key to prevent unauthorized access. The seal is an encryption function that only allows the TPM and its specific software environment that created the object to decrypt it.

Policy Enforcement and Monitoring

We describe only one of the many implemented functions of this component, namely access policy enforcement. The access policy enforcement mechanism consists of two steps. The first step checks whether the doctor is allowed to access the information. The second step checks whether the facility, from which the doctor is trying to access the information, is allowed to access it. For both steps, all related policies for an e-rec are collected from e-tags in a list and checked against the conflict resolution mechanism specified in the fouth section to see whether the winner mode is grant or deny. Access is granted to the doctor in

Figure 10. E-R diagram for the tag-based model



the facility for the e-rec only if the policy for *both* doctor and the facility allows it. Otherwise, the access is denied.

Integrity Measurement

As explained earlier, in our distributed health care system scenario, a government health department provides the operational credentials and licenses for health care facilities. Our attestation protocol requires both parties to validate the issued credentials. In order to support this, we introduce a facility credential, digitally signed by the department of health or a similar authority, within the TPM credential architecture as shown in Figure 11. This credential is given to all facilities by the external authority after registration, and is linked to the endorsement credential of the registered facility.

The general principle of integrity measurement using a TPM-based system is that any software is measured using the SHA-1 digest of the executable code before being loaded and stored in PCR registers. The first (loaded) software that does the measurement is called c*ore root of trust for measurement* (CRTM), which in our case, is the machine BIOS. Our development platform was based on Microsoft's current version of the Windows XP operating system (OS), with the application development being done under the .Net environment. Despite the fact that the TPM chipset is found in most new PCs, the version of Windows XP (the most current released version at time of writing) did not support integrity *measurement* from CRTM. Further and because of this fact, there is no chain of trust that ties together the hardware BIOS to the application software.

Due to the lack of support at the OS level, we simulated the core root of trust by developing specialized loader software, called the *MedicLoader*, as shown in Figure 12. This loader intercepts calls to the loaded and unloaded applications and performs measurements on the executables before their execution. A fundamental assumption to our implementation is that the MedicLoader can be trusted to securely measure itself and the processes it loads before executing them. This means all components that are under this root of trust also must be trusted. Although this is useful for a concept demonstration, it is not a realistic solution for commercial applications. Realistic solutions must rely on OS support of both integrity measurement and being able to respond to a chain of trust.

The integrity measurements are performed as follows. For each executable file, we uniquely

Figure 11. The TPM credential architecture

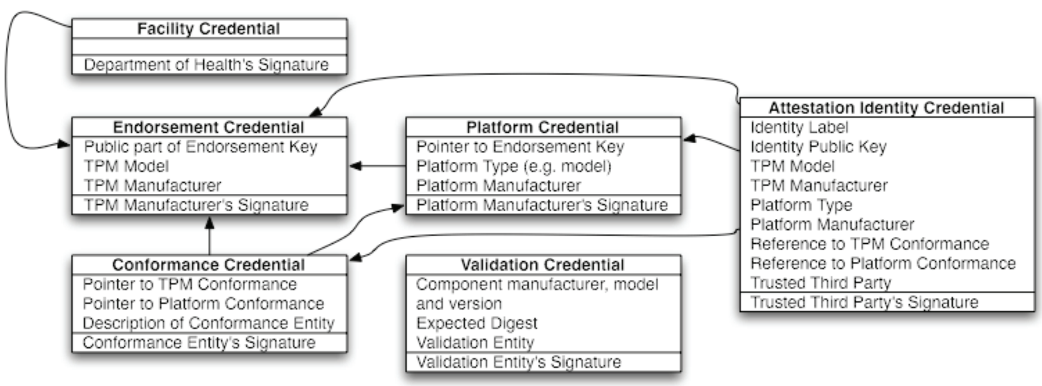
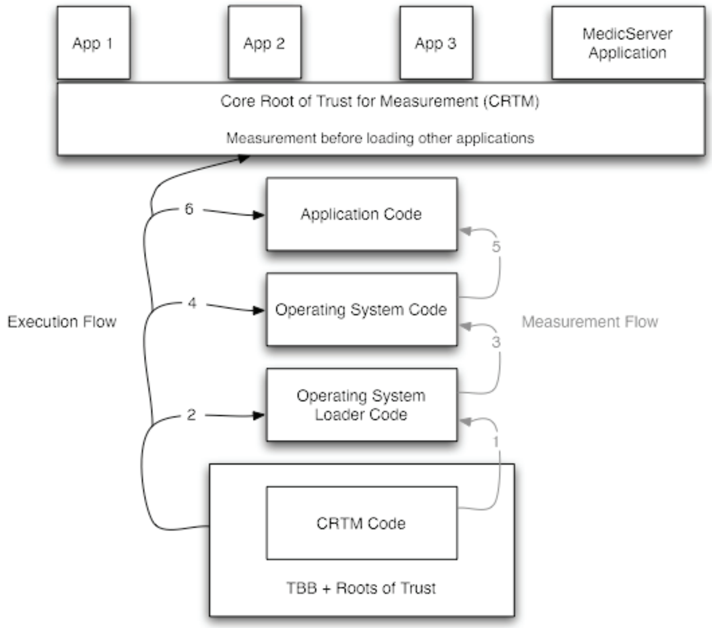


Figure 12. Raising the level of trust (for measurement)



identify the executable and capture whether it is loaded or unloaded. We do this by extending the PCR values as a triple: the state of the code ("load" or "unload"), the file name of the executable, and SHA-1 digest value of the executable. Our implementation of this triple is simply a concatenation of these three values, and each of these resulting triples is kept as a list. Separate from this list, each facility keeps a database of processes that represents a register of "acceptable" process states. Each entry consists of the name, a calculated hash value for the executable, if it is loaded or unloaded, and the desired executable load state ("may," "must," "must not"). "May" signifies that the process may be loaded optionally, whereas "must" signifies that there is no option for loading the process—it must be present. We next describe the mutual attestation protocol between two server applications using this integrity measurement technique.

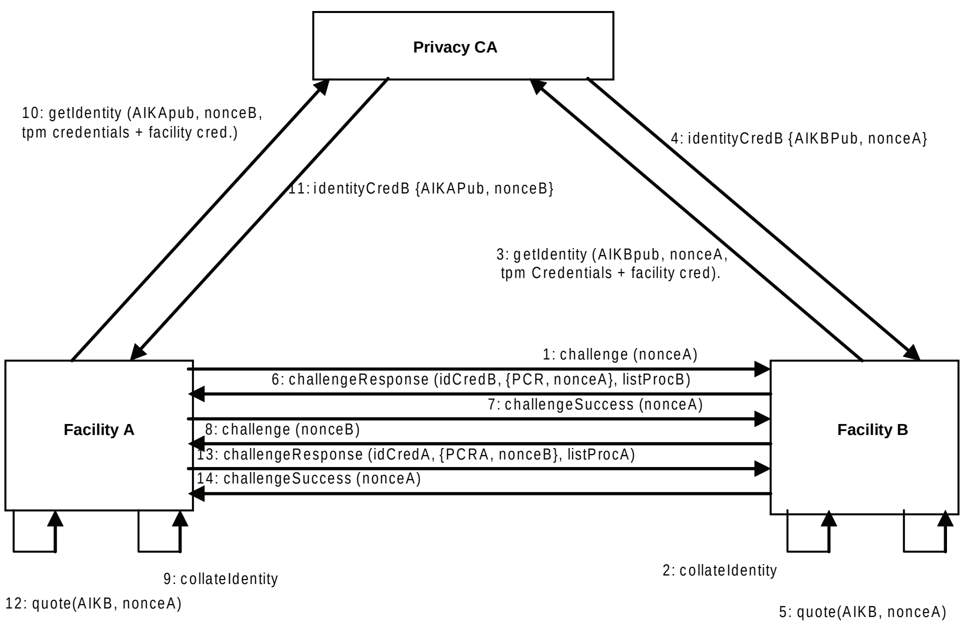
Mutual Attestation

The TrouSers FAQ (IBM Trousers, 2006) shows the existence of a dependency graph of trusted computing elements that are required to support remote attestation. Importantly, there are key elements that are missing in the current infrastructure; therefore, it is not possible to use some of the core functions provided by TCG's specifications for remote attestation, such as "CollateIdentityRequest" to collect TPM credentials or the "PCR_extend" operation. Because of this, we had to simulate these missing elements in our prototype demonstrator.

The TCG supports two remote attestation methods: one by use of a Privacy CA and the other by DAA (Yung, 2005). We chose the former in our prototype demonstration. Figure 13 illustrates the implemented mutual attestation protocol. The following explains the protocol steps when the originating Facility A wants to send a patient record to the destination Facility B.

1. Facility A sends a challenge to Facility B with an identifying nonce. We call this "nonceA."
2. Upon receiving the challenge, Facility B obtains its TPM credentials by calling the function "CollateIdentityRequest." This function also generates an AIK. The TPM credentials, facility credential, and the public part of the AIK are signed by the private part of the TPM endorsement key and then encrypted, using the public part of the Privacy CA key.
3. The encrypted blob is sent to the Privacy CA as an identity credential request.
4. The Privacy CA decrypts this blob with its private key. It verifies the signatures of the credential request, and then creates and sends an identity credential back to Facility B. This credential is a digital certificate containing the public part of the AIK, together with "nonceA," signed by the CA private key.
5. Facility B then performs an integrity measurement using the TPM "quote" function, using the AIK, "nonceA," and one or more PCR values, as input parameters. The AIK can only be used to sign the quoted PCR registers.
6. The message containing the output of the quote function, the identity credential received from the CA, and the list of recorded loaded and unloaded processes is then sent to Facility A from Facility B. On reception of this message, Facility A first checks the identity credential, verifying that the Privacy CA has signed it. This also reveals that Facility B has a TPM and that the AIK is an identity key in the TPM of Facility B. With this information, Facility A can then verify the signature of the PCR values and "nonceA." If the signature is verified successfully, then Facility A knows the PCR

Figure 13. Mutual attestation protocol



values of Facility B. Finally, Facility A does the following three verifications:
   a. All processes in the received list are in the database of acceptable processes with corresponding hash code.
   b. The expected PCR value is computed from the list of processes and compared with the PCR values of Facility B.
   c. All processes that are marked as "must" in the acceptable process database are loaded.
7. If all checks are successful, Facility A sends a challengesuccess message to Facility B. Otherwise, the attestation fails.
8. Facility B receives the challengesuccess message, and initiates the same procedure (steps 8-14 in Figure 13). At the end of these steps, both facilities are authenticated mutually and are ready to exchange medical records.

Preprocessing and Postprocessing

A patient medical record has a variety of e-tags attached to it. When a record is transferred from one facility to another, these e-tags sometimes may not be transferrable in their current form. For example, each facility may have different default policies for different categories of information, or in the case of anonymous transfers, the current e-tags must not be transferred. This requirement for processing tags indicates that there is a need for preprocessing and postprocessing of medical records, both just before sending them and just after receiving them. We defined and implemented the following preprocessing and postprocessing rules for a selection of e-tag types.

Facility default e-tags: Default facility *policies* are defined for a local facility only and must not be transferred. However, the *category* information itself must be transferred for classification purposes. That is, the originating facility will transfer e-tags together with categories but without policies. The destination facility will use its own default policies in the e-tag corresponding to the category.

Patient and family e-tags: Policies are always transferred for e-tags of type *patient* or *family*. In the case of anonymous transfer, the category value must be replaced with the anonymous identity as explained for the transfer protocol (see previous section).

E-rec specific policies: Specific policies attached to an e-rec are always transferred without any modification. As we assumed that roles, doctors, and facilities are globally identifiable, all policy expressions can be transferred and processed in every facility without any manual or automatic modification.

The developed prototype system was demonstrated to the Center for Networking Technologies for the Information Economy (CeNTIE) (2006) Enterprise System Focus Group and other stakeholders in September 2005. The demonstration has led us to the belief that our data model can be implemented using available technologies in databases. The demonstration also shows us that it is feasible to deploy mutual attestation protocols between two server applications in a distributed health care environment to enhance the trust for sharing electronic patient records.

DISCUSSIONS AND CONCLUSIONS

Previous e-consent data models were developed on a set of restricted assumptions. By widening the assumptions and examining the realistic use of e-consent within the health care system, we noted that these e-consent data models needed to be changed. Consequently, we developed a flexible data model for electronic medical records that we called the tag-based data model. The data model allows the representation of patients' medical records and incorporates their consent information. The model also allows categorization of medical records into different groups. This in turn allows the definition of default policies for each category. Unlike existing role-based access models, our data model supports both access and transfer policies on roles, categories of information, facilities, and health care practitioners.

Our experience has been that the data model allows a great deal of flexibility and autonomy for the end users and imposes a minimal set of semantic requirements on its use in specifying policies and categories. Because e-tags and e-recs are securely encrypted, indexing and searching requires the extension of the current data model to include metadata information.

Another important issue identified in previous e-consent demonstrators is the requirement for mutual trust between two interacting facilities. We believe that our mutual attestation protocol between two facilities provides a reflexive form of trust between two parties needing to support enforcement of policies. We explored the bounds of what is possible in the current Microsoft Windows environment; although the current Windows infrastructure lacks the necessary elements for remote attestation, such as chain of trust from hardware to operating system to applications.

Finally, we have developed a prototype system and demonstrated it to the CeNTIE (2006) Enterprise System Focus Group. The demonstration has led us to the belief that it is feasible to deploy such mutual attestation protocols between two server applications in a distributed health care environment to enhance the trust, while sharing electronic patient records. We also observed that the current implementation of mutual attestation protocol performance might be improved, as the protocol has to be run for each message sent or received by a facility. In future, we plan to tackle this problem by developing an application session-based mutual attestation protocol. Our preliminary study shows that such a protocol also will address the problem of discrepancy in time-of-use and time-of-attestation.

ACKNOWLEDGMENT

This work is supported by the Australian government through the Advanced Networks Program of the Department of Communications, Information Technology and the Arts. Discussions with Paul Greenfield of the prior e-consent model greatly assisted us in developing the tag-based data model.

REFERENCES

Australian Government Department of Health and Aging (DoHA) Project. (2002). *Consumer consent in electronic health data exchange—e-consent*. Retrieved March , 2006, from www.ahic.org.au/downloads/ Foundations%20for%20the%20future.pdf

Barham, P., Dragovic, B., Fraser, K., Hand, S., Harris, T., Ho, A., Neugebauer, R., Pratt, I., & Warfield, A. (2003). Xen and the art of virtualization. *19th ACM Symposium on Operating Systems Principles*, (pp. 164-177). New York: .

Brickell, E., Camenisch, J., & Chen, L. (2004). *Direct anonymous attestation* (Tech. Rep. HPL-2004-93).

Centre for Networking Technologies for the Information Economy (CeNTIE) Project. (2006). Retrieved from http://www.ict.csiro.au/page.php?cid=22

Chadwick, D., & Mundy, D. (2004). The secure electronic transfer of prescriptions. *Healthcare Computing*, 11-25.

Choudhri, A.., Kagal, L., Joshi, A.., Finin, T., & Yesha, Y. (2003). PatientService: Electronic patient record redaction and delivery in pervasive environment. *Fifth International Workshop on Enterprise Networking and Computing in Healthcare Industry* (Healthcom 2003), .

Crook, R., Ince, D., & Nuseibeh, B. (2003). Modelling access policies using roles in requirements engineering. *Information and Software Technology, 45,* 979-991.

Evered, M., & Bogeholz, S. (2004). A case study in access control requirements for a health information system. *Australasian Information Security Workshop,* 53-61.

Garfinkel, T., Pfaff, B., Chow, J., Rosenblum, M., & Boneh, D. (2003). Terra: A virtual machine-based platform for trusted computing. *19th ACM Symposium on Operating Systems Principles* (pp. 193-206). Bolton Landing, NY.

Haldar V., & Franz, M. (2004). Symmetric behavior-based trust: A new paradigm for Internet computing. *New Security Paradigms Workshop,* 79-84.

Huston, T. (2001). Security issues for implementation of e-medical records. *Communication of the ACM, 44*(9) 89-94.

IBM. Trousers. (2005). *A TSS implementation for Linux*. Retrieved from http://trousers.sourceforge.net/

Khayat, E. J., & Abdallah, A. E. (2003). A formal model for flat role-based access control. *IEEE International Conference on Computer Systems and Applications, Tunisia*.

Messerges, T. S., & Dabbish, E. A. (2003). Digital rights management in a 3G mobile phone and beyond. In *Proceedings of the 2003 ACM Workshop on Digital Rights Management* (pp. 27-38). Washington, DC.

Motta, G. H. M. B., & Furuie, S. S. (2003). A contextual role-based access control authorization model for electronic patient record. *IEEE Transactions on Information Technology in Biomedicine, 7*(3), 202-207.

Nepal, S., Zic, J., Jaccard, F., & Krahenbuehl, G. (2006a). A tag-based data model for privacy-preserving medical applications. *Proceedings of EDBT IIHA Workshop* (pp. 77-88). Munich, Germany.

Nepal, S., Zic, J., Krahenbuehl, G., & Jaccard, F. (2006b). Secure sharing of electronic patient records. *Proceedings of 1st European Conference on eHealth*. Fribourg, Switzerland.

Open Mobile Alliance (OMA), DRM Architecture, Version 2.0.6. (2004). *OMA-DRM-ARCH-V2_0_6-20040820-C.zip*. Retrieved from http://www.openmobilealliance.org

OASIS. (2005). *eXtensible access control markup language (XACML) version 2.0 3*. Retrieved from http://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-core-spec-os.pdf

O'Keefe, C. M., Greenfield, P., & Goodchild, A. (2005). A decentralised approach to electronic consent and health information access control. *Journal of Research and Practice in Information Technolog*y, 37(2), 161-178.

Paul, A., Satoshi, H., Karjoth, G., Powers, C., & Schunter, M. (2003). *Enterprise privacy authorization language (EPAL 1.1)* (IBM Tech. Rep.)

Reid, J., Cheong, I., Henricksen, M., & Smith, J. (2003a). A novel use of RBAC to protect privacy in distributed health care information systems. In R. Safavi-Naini & J. Seberry (Eds.), *Proceedings 8th Australasian Conference on Information Security and Privacy (ACISP 2003*) (pp. 403-415). Wollongong.

Reid, J., Juan, M., Nieto, G., Dawson, E., & Okamoto, E. (2003b). Privacy and trusted computing. *14th International Workshop on Database and Expert Systems Applications (DEXA'03)* (pp. 383).

Sadeghi, A., & Stüble, C. (2004). Property-based attestation for computing platforms: Caring about properties, not mechanisms. *Proceedings of the 2004 Workshop on New Security Paradigms*. (pp. 67-77). Nova Scotia, Canada: Sailer, R., Jaeger, T., Zhang, X., & Doorn, L. (2004a). Attestation-based policy enforcement for remote access. In *Proceedings of the 11th ACM Conference on Computer and Communication Security* (pp. 308-317). Washington, DC.

Sailer R., Jaeger T., Zhang X. & Doorn L. (2004a) Design and implementations of a TCG-based intergrity measurement architecture. 13th usenix Security Symposium, pp. 223-238, San Diego, CA.

Sailer, R., Jaeger, T., Zhang, X., & Doorn, L. (2004b). Design and implementation of a TCG-based integrity measurement architecture. *13th Usenix Security Symposium* (pp. 223-238). San Diego, CA.

Sandhu, R., & Zhang, X. (2005). Peer-to-peer access control architecture using trusted computing technologies.In *Proceedings of the Tenth ACM Symposium on Access Control Models and Technologies* (pp. 147-158). Stockholm, Sweden.

Shi, E., Perrig, A., & Van, D. L. (2005). BIND: A fine-grained attestation service for secure distributed systems. *IEEE Symposium on Security and Privacy* (pp. 154- 168).

Stein, L. D. (1997). The electronic medical record: Promises and threats. *Web Journal*, *2*(3).

Task Force on Medical Informatics (TFMI). (1996). Safeguard needed in transfer of patient data. *PEDIATRICS*, *98*(5), 984-986.

Trusted Computing Group (TCG). (2003). *Trusted Computing Platform Alliance (TCPA) Main Specification Version 1.1b.* Retrieved from https://www.trustedcomputinggroup.org/groups/tpm/

Yung, M. (2005). *Privacy implications of remote attestation in trusted computing* (CSIRO ICT Centre Tech. Rep. 05/146).

ENDNOTE

[1] It should be noted that access to the TPM hardware is done via a specialized software stack.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol.2 Issue 1, edited by J. Tan, pp. 14-35, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.12
Reliability and Evaluation of Health Information Online

Elmer V. Bernstam
University of Texas, Health Science Center at Houston, USA

Funda Meric-Bernstam
University of Texas, M.D. Anderson Cancer Center, USA

ABSTRACT

This chapter discusses the problem of how to evaluate online health information. The quality and accuracy of online health information is an area of increasing concern for healthcare professionals and the general public. We define relevant concepts including quality, accuracy, utility, and popularity. Most users access online health information via general-purpose search engines, therefore we briefly review Web search-engine fundamentals. We discuss desirable characteristics for quality-assessment tools and the available evidence regarding their effectiveness and usability. We conclude with advice for healthcare consumers as they search for health information online.

INTRODUCTION

The healthy, the newly diagnosed, and the chronically ill turn to the Internet for health information. In spite of some controversy regarding the number of individuals that are accessing online health information at any given time, most experts agree that the numbers are enormous. Indeed, some have estimated that on any given day, more people consult the Internet for health information than see a physician (Fox & Rainie, 2002). Furthermore, there is evidence that patients are making treatment decisions based on the information that they encounter online (Helft, Hlubocky, Gordon, Ratain, & Daugherty, 2000). Therefore, researchers, clinicians, and the general public are increasingly concerned about the reliability of health information online that is directed toward consumers of healthcare. In this chapter, we consider the problem of how to enable healthcare consumers to evaluate online health information.

DEFINITIONS: INFORMATION QUALITY VS. INFORMATION ACCURACY

One of the major impediments to research into online information quality is the lack of clear, consistent, and generally accepted definitions. In this section, we define relevant terms to provide a vocabulary for discussion.

The factual correctness (accuracy) of health information online may be difficult to assess. Indeed, even experts often disagree regarding accuracy. Therefore, review of information content by a panel of experts is considered to be the gold standard of accuracy.

Most Internet users are not healthcare experts. Therefore, they cannot judge the accuracy of online health information. Since consumers cannot assess accuracy, surrogate measures that they can assess are appealing. We refer to these surrogate measures as measures of quality and collections of these measures as quality-assessment tools. Measures of information quality are useful to the extent that they (a) can be effectively assessed by healthcare consumers and (b) correlate with outcomes of interest such as whether the information is factually incorrect or whether the information has the potential to harm health (i.e., if the advice were followed).

An example of an information quality measure is authorship (i.e., is an author identified?). The JAMA benchmarks (Silberg, Lundberg, & Musacchio, 1997) are a commonly cited quality-assessment tool. The JAMA benchmarks consist of four quality measures: authorship, attribution, disclosure, and currency. These generally came to be known as the "clarity in publishing" criteria and are similar in spirit to the type of questions one might ask about a textbook or scientific paper.

Preferably, measures of quality should be based on meaning rather than presentation of information. In contrast, some studies tested superficial characteristics such as the claim of copyright (Fallis & Fricke, 2002). In this study, Web pages that claimed copyright were more likely to be accurate than pages that did not claim copyright. The authors point out, however, that it is simple to claim copyright. Even if such superficial measures correlate with accuracy, they are easy to manipulate. Web-site developers could simply claim copyright without modifying the information displayed on the Web site. Therefore, as superficial measures become more widely used, they will become less useful.

Unfortunately, quality measures are difficult to put into practice. A systematic review of studies assessing the quality of online health information determined that although 70% of the studies found quality to be a problem, there were wide differences in the quality measures used, their operational definitions,[1] and methods in which the analyses were carried out (Eysenbach, Powell, Kuss, & Sa, 2002).

Some have argued that a high-quality Web site should display information that is accurate, easy to understand, specifically tailored to the intended audience, and pleasing to the eye. However, in order to enable communication, more precise definitions are useful. Eysenbach et al.'s (2002) review of the literature distinguished the following: (a) technical quality criteria, defined as "general, domain-independent criteria, i.e., criteria referring to how the information was presented or what metadata[2] was provided," (b) design, which "includes visual aspects of the site such as the colors used or layout," (c) readability, meaning whether the language is easy to read and is understandable, (d) accuracy, or the "degree of concordance with the best evidence or with generally accepted medical practice," and (e) completeness, which refers to the portion of relevant material covered on the Web site. These five characteristics may be independent from each other. For example, it is possible for a high-quality Web site based on technical quality criteria to display false information, or a poor-quality Web site to display accurate information.

QUALITY-ASSESSMENT TOOLS

In response to widespread concern regarding the quality and accuracy of online health information, many organizations published quality-assessment tools. Consumers are encouraged to evaluate online health information using these quality-assessment tools. In spite of this and consumers' stated concern regarding the accuracy of the information that they see on the Internet, recent studies found little evidence that consumers are considering quality when searching for information online (Eysenbach & Kohler, 2002; Meric et al., 2002).

As mentioned above, quality measures often reflect clarity-in-publishing standards adapted from traditional media such as the display of authorship, attribution, disclosures of conflicts of interest, and so forth. Similarly, a television viewer or newspaper reader may want to look for an identified author and sponsorship information. When reading a scientific paper or book chapter, one looks for references to the published literature that support assertions. It is possible that these measures are less appropriate in some domains. Specifically, complementary and alternative medicine (CAM) is an important topic commonly addressed by health-related Web sites. The accuracy of CAM information is difficult if not impossible to define. By definition, CAM information may not reflect expert medical opinion or conventional standards of care. Furthermore, few CAM Web sites reference traditional biomedical literature (Sagaram, Walji, & Bernstam, 2002; Walji et al., 2004). An alternative is to consider the potential harm of the information if the advice were followed. Clinicians assessing the potential harm of online health information are able to do so with high interobserver reliability (see next paragraph; Walji et al.).

A recent review of quality-rating tools provided by the likes of the American Medical Association and Health on the Net (HON) Foundation, among others, found that no instrument reported interobserver reliability as a measure of its validity (Gagliardi & Jadad, 2002). The purpose of quality-rating tools is to allow those looking for health information to filter out poor-quality content. However, the criteria used must also be unambiguous so that different people using the same criteria will agree upon their usage. Therefore, it is important that different reviewers can agree upon and reproduce the results. In other words, quality measures must have high interobserver reliability. Unfortunately, evaluations of interobserver reliability have not been encouraging. Even trained informaticians are not able to consistently agree on the presence or absence of commonly cited quality criteria (Bernstam, Sagaram, Walji, Johnson, & Meric-Bernstam, 2005; Sagaram, Walji, Meric-Bernstam, Johnson, & Bernstam, 2004). It seems unlikely that consumers without informatics training fare any better. Furthermore, many quality-rating tools (composed of one or more quality criteria) are not practically usable by consumers. Many tools are not publicly available, are composed of criteria that cannot be reliably assessed, or have too many criteria (more than 10 questions to be answered per site; Bernstam, Shelton, Walji, & Meric-Bernstam, 2005). Table 1 provides examples of commonly cited quality criteria, their operational definitions, and expected interobserver reliability.

In addition to questionable usability, there is also conflicting evidence regarding the effectiveness of existing quality-assessment tools to identify false or misleading information online. Some studies show a correlation between quality and accuracy (Meric et al., 2002), while others do not (Walji et al., 2004). The reason(s) for this discrepancy is not clear. One possible explanation is that the correlation is domain dependent. Therefore, studies may differ depending on the specific topic tested (e.g., breast cancer vs. complementary and alternative medicine). For a comprehensive listing of commonly cited quality-assessment tools, see Eysenbach et al. (2002).

Is There Documented Harm from Online Health Information?

In spite of widespread concern regarding the potentially harmful effects of online health information, documented evidence of physical harm is scarce. A systematic review of harm from online information found few reported cases. However, the authors caution that this may be a result of underreporting rather than actual absence of harm (Crocco, Villasis-Keever, & Jadad, 2002). Consequently, some argue that concerns have been overemphasized and benefits to users have been underappreciated (D. Hoch & Ferguson, 2005). This issue remains controversial.

SEARCHING FOR HEALTH INFORMATION (WEB SEARCH ENGINES)

Few consumers are directed to specific Web sites by their healthcare provider (Reents, 1999; Tang & Newcomb, 1998). Therefore, left to their own devices, consumers use general-purpose search engines to find health information on the Internet (Eysenbach & Kohler, 2002; Fox & Rainie, 2002). Although their search strategies are often suboptimal and they rarely look beyond the first page of results, consumers are generally satisfied with their ability to find information on the Internet (Eysenbach & Kohler; Fox & Rainie, 2000). To a large extent, search engines determine what information is seen by consumers.

Search-engine technology is complex, changes quickly, and is often proprietary. Therefore, a detailed discussion of the inner workings of search engines is beyond the scope of this chapter. However, we will present a basic overview of relevant search-engine characteristics with particular emphasis on Google (http://www.google.com), the most popular search engine in the English-speaking world (Sullivan, 2004).

In general, to use a search engine, the user types in a few terms perceived to be related to his or her information need. An information need is an expression, in the user's own language, of the information that she or he desires. For example, "Should I use tamoxifen to keep my breast cancer from coming back?" is an information need. In order to satisfy this information need, a user may issue the following query: "breast cancer tamoxifen." A result set is retrieved in response to the query. The user can click on results that seem to be interesting. If none of the results seem interesting, users generally add or subtract terms from a query (Eysenbach & Kohler, 2002). For example, if "breast cancer tamoxifen" does not lead to a successful search, then "breast cancer tamoxifen recurrence" may be a reasonable query to try.

Before Google, most search engines used text-processing techniques to identify relevant Web sites. In other words, the search engine looked for Web pages that contained the query terms. This approach worked well when the Web was small. However, with many billions of Web sites, common queries now return too many results. Therefore, identifying results that are most likely to be useful became increasingly important. The designers of Google noted that the link structure of the Web was a rich source of information. Specifically, incoming links imply importance. Important sites tend to have many incoming links. Their algorithm, known as PageRank, makes use of links, or references from one Web page to another. This approach has proven so successful that most Web search engines now use some form of link analysis (Cho & Roy, 2004). Therefore, information displayed by popular Web sites is likely to be seen by healthcare consumers. Some studies found a correlation between popularity (Google rank) and accuracy (Fricke, Fallis, Jones, & Luszko, 2005) while others did not (Meric et al., 2002).

Search engines drive traffic on the Web. Therefore, Web sites compete for position on prominent search engines. Being ranked highly by Google, for example, will dramatically increase the number of users accessing a Web site (Cho & Roy, 2004). Many search engines, including Google, provide two types of results in response to a query. Algorithmic results are generated by the general-purpose search algorithm. In contrast, sponsored results are links to Web sites that have paid a fee to the search engine and are essentially advertisements. These are likely to contain commercial offers, often related to complementary and alternative medicine (Walji, Sagaram, Meric-Bernstam, Johnson, & Bernstam, 2005).

Unfortunately, not all search engines clearly identify sponsored results. Consumers are unaware of the existence of sponsored listings and do not differentiate them from nonsponsored results, which are retrieved algorithmically (i.e., using an algorithm that does not factor payment for placement). In fact, two in five (41%) links clicked by consumers are sponsored results (Marable, 2003). The Federal Trade Commission has recently recommended that search engines clearly identify sponsored listings (Hippsley, 2002).

ADVICE TO THOSE SEEKING HEALTH INFORMATION ONLINE

In this section, we consider the advice that can be offered to users searching for health information online in the absence of usable, reliable quality-assessment tools. Unfortunately, we cannot recommend a quality-assessment tool that can be used to identify problematic health information online. Similarly, we cannot recommend a Web search engine that lists only trustworthy information. In spite of this, we can recommend simple practices that are likely to minimize exposure to inaccurate information online.

Access Sites Known to be Trustworthy

Multiple government agencies such as the National Institutes of Health (NIH, USA) and the National Health Service (NHS) of the United Kingdom maintain Web sites with extensive libraries of health information intended for consumers. MedlinePLUS, maintained by the National Library of Medicine (USA), is a specific example. There are also disease-specific organizations that maintain high-quality content on every imaginable topic. Prominent examples include the National Cancer Institute (USA) and the American Heart Association (USA). Table 2 provides a small list of trustworthy Web sites with stable URLs (uniform resource locators). However, we emphasize that the Web is a dynamic medium and therefore any list of Web sites is dated by the time that it is published in print.

Distinguish Algorithmic from Sponsored Results

As discussed above, sponsored results are basically advertisements for particular Web sites that are often financed by commercial interests. Therefore, many sponsored links lead to Web sites attempting to sell products or services. To their credit, Google clearly identifies and separates sponsored results from algorithmic results. Searchers should be aware of how they encountered the Web site that they are viewing.

Consider the Source

Common sense tells us that information displayed by Web sites with commercial interests (i.e., those that are selling products) may not be entirely unbiased. Therefore, users would do well to consider the source, just as they would when evaluating non-health information. In general, government, organization, or academic Web sites are more likely to provide impartial advice compared to company (commercial) Web sites. However, this is a generalization and there may be exceptions.

Check with a Healthcare Professional

In some ways, online health information is no different from health information provided via any other medium. When in doubt, or when making important decisions, there is still no substitute for professional assessment and judgment. Therefore, the prudent health seeker will use online resources for education, but review important decisions with their healthcare provider.

FUTURE DIRECTIONS

As of this writing, traditional quality-assessment tools do not sufficiently address the concern regarding the quality and accuracy of health information online. Clearly, self-regulation (the status quo) does not prevent or identify false, misleading, or potentially harmful online health information. Regulation, whether by government(s) or nongovernmental organizations, does not seem viable. Indeed, even voluntary standards such as the HON code of conduct are difficult to enforce (Meric et al., 2002). In this section, we consider promising new approaches with the potential to address the shortfalls of traditional quality-assessment tools.

Distributed Human Annotation

There is too much Web content for any independent organization to evaluate and validate. Since content is constantly changing, evaluation and validation must be repeated to ensure compliance over time. A potential solution is to create a framework and process by which independent evaluators can collaboratively annotate content. Evaluators include the content providers (i.e., the Web-site developers), information specialists, and healthcare professionals. Therefore, the Web-site developer provides metadata such as authorship information. The metadata is checked by a nonmedical information specialist. Finally, the Web-site content is reviewed by a healthcare professional. Once the Web site has been evaluated, it is tagged with descriptors. Therefore, users can search for a Web site that has specific characteristics (e.g., has been evaluated by a healthcare professional). The MedCIRCLE (http://www.medcircle.org) project provides a vocabulary (HIDDEL, Health Information Disclosure, Description and Evaluation Language) to allow independent evaluators to describe a Web site in a consistent manner.

The advantage of this approach is that multiple independent evaluators can effectively collaborate. Consistent self-rating by Web-site developers may be better than no rating at all. However, as discussed above, self-regulation does not always ensure compliance with quality standards. In addition, independent manual Web-site evaluation is laborious. Although a distributed approach is likely to be more effective than one that relies on a single central authority, the Web is vast and constantly changing. Therefore, any validated subset of the Web is likely to be a very small percentage of all health information available online.

Reputation-Based Systems

Reputation-based systems explicitly model trust, which is hearsay evidence of predictable behavior. There are multiple examples of reputation-based systems in common use outside of healthcare. Perhaps the best known reputation-based system is the online auction company eBay (http://www.ebay.com) that brokers transactions (auctions) between buyers and sellers. After each transaction, feedback is solicited from both parties. Users who do not behave honorably acquire a bad reputation that is visible to the community. Similar systems have been developed for e-mail networks (Golbeck & Hendler, 2004). For online health information, one could develop algorithms to identify reliable sites using link analysis given a small set of verified (known to be good) sites such as those shown in Table 2.

Reproducibility

Unlike the traditional print media, Web sites do not exist independently of each other. In other words, a user has access to multiple Web sites from the same terminal. In contrast, when reading a book or journal article on paper, one may not have access to other books or articles on the same topic. It is easy to access multiple Web sites on the same topic to see if there is a consensus regarding specific information. One might hypothesize that if information is repeated on multiple independent Web sites, it is more reliable than information found on only a single Web site. However, this heuristic has not, to our knowledge, been systematically evaluated.

Self-Correction Hypothesis

The self-correction hypothesis refers to the possibility that users can identify and correct problematic health information online without professional guidance or supervision (Fennberg, Licht, Kane, Moran, & Smith, 1996; Ferguson, 1995; D. B. Hoch, Norris, Lester, & Marcus, 1999). In a sense, this is an extension of the approach used by open-source projects such as the online encyclopedia Wikipedia (http://www.wikipedia.org) and the open-source software movement where the user community identifies and corrects errors in content or computer code. Wikipedia allows any Web user to create and modify content. A recent analysis of Wikipedia content found that Wikipedia was generally accurate compared to a traditionally curated online encyclopedia like Britannica (Giles, 2005). In the context of online health information, we do not know whether self-correction really works. However, preliminary data are promising (Esquivel, Meric-Bernstam, & Bernstam, 2006). Therefore, it is possible that, given a sufficiently large and active forum, the Web can police itself.

CONCLUSION

Many Web users are searching for health information online and express concern regarding the quality and accuracy of online health information. Although there is evidence that online information affects healthcare decisions, there is little evidence that users are considering quality when accessing health information online. Helping users identify problematic health information online remains an open problem. Currently available quality-assessment tools cannot be reliably assessed by Web users and may not be effective at identifying problematic health information online. Fortunately, this is a young and dynamic field. Promising alternatives to currently available quality-assessment tools are subjects of ongoing research.

REFERENCES

Bernstam, E. V., Sagaram, S., Walji, M., Johnson, C. W., & Meric-Bernstam, F. (2005). Usability of quality measures for online health information: Can commonly used technical quality criteria be reliably assessed? *International Journal of Medical Informatics, 74*(7-8), 675-683.

Bernstam, E. V., Shelton, D. M., Walji, M., & Meric-Bernstam, F. (2005). Instruments to assess the quality of health information on the World Wide Web: What can our patients actually use? *International Journal of Medical Informatics, 74*(1), 13-19.

Cho, J., & Roy, S. (2004). *Impact of search engines on page popularity*. Paper presented at the WWW2004, New York.

Crocco, A. G., Villasis-Keever, M., & Jadad, A. R. (2002). Analysis of cases of harm associated with use of health information on the Internet. *JAMA, 287*, 2869-2871.

Eisenberg, D. M., Kessler, R. C., Rompay, M. I. V., Kaptchuk, T. J., Wilkey, S. A., Appel, S., et al. (2001). Perceptions about complementary therapies relative to conventional therapies among adults who use both: Results from a national survey. *Annals of Internal Medicine, 135*(5), 344-351.

Esquivel, A. E., Meric-Bernstam, F., & Bernstam, E. V. (in press). Accuracy and self correction of information received from an Internet breast cancer list: Analysis of posting content. *BMJ*.

Eysenbach, G., & Kohler, C. (2002). How do consumers search for and appraise health information on the World Wide Web? Qualitative study using focus groups, usability tests, and in-depth interviews. *BMJ, 324*(7337), 573-577.

Eysenbach, G., Powell, J., Kuss, O., & Sa, E. R. (2002). Empirical studies assessing the quality of health information for consumers on the World Wide Web: A systematic review. *JAMA, 287*(20), 2691-2700.

Fallis, D., & Fricke, M. (2002). Indicators of accuracy of consumer health information on the Internet: A study of indicators relating to information for managing fever in children in the home. *J Am Med Inform Assoc, 9*(1), 73-79.

Fennberg, A., Licht, J., Kane, K., Moran, K., & Smith, R. (1996). The online patient meeting. *Journal of Neurological Science, 139*(Suppl.), 129-131.

Ferguson, T. (1995). Consumer health informatics. *Healthcare Forum Journal*, 28-33.

Fox, S., & Rainie, L. (2000). *The online healthcare revolution: How the Web helps Americans take better care of themselves*. Washington, DC: Pew Internet and American Life Project.

Fox, S., & Rainie, L. (2002). *Vital decisions: How Internet users decide what information to trust when they or their loved ones are sick*. Washington, DC: Pew Internet & American Life Project.

Fricke, M., Fallis, D., Jones, M., & Luszko, G. M. (2005). Consumer health information on the Internet about carpal tunnel syndrome: Indicators of accuracy. *Am J Med, 118*(2), 168-174.

Gagliardi, A., & Jadad, A. R. (2002). Examination of instruments used to rate quality of health information on the Internet: Chronicle of a voyage with an unclear destination. *BMJ, 324*(7337), 569-573.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature, 438*(7070), 900-901.

Golbeck, J., & Hendler, J. (2004, October). *Accuracy of metrics for inferring trust and reputation*. Paper presented at the 14th International Conference on Knowledge Engineering and Knowledge Management, Northamptonshire, United Kingdom.

Helft, P. R., Hlubocky, F. J., Gordon, E. J., Ratain, M. J., & Daugherty, C. (2000). *Hope and the media in advanced cancer patients*. Paper presented at the American Society of Clinical Oncology 36th Annual Meeting, New Orleans, LA.

Hippsley, H. (2002). *Letter to Gary Ruskin: Executive director of Commercial Alert*. Retrieved from http://www.ftc.gov/os/closings/staff/commercialalertletter.htm

Hoch, D., & Ferguson, T. (2005). What I've learned from e-patients. *PLoS Med, 2*(8), e206.

Hoch, D. B., Norris, D., Lester, J. E., & Marcus, A. D. (1999). Information exchange in an epilepsy forum on the World Wide Web. *Seizure, 8*(1), 30-34.

Marable, L. (2003). False oracles: Consumer reaction to learning the truth about how search engines work: Results of an ethnographic study. *Consumer Reports WebWatch*. Retrieved February 27, 2006, from http://64.78.25.46/view-article.cfm?id=10171&at=510

Meric, F., Bernstam, E. V., Mirza, N. Q., Hunt, K. K., Ames, F. C., Ross, M. I., et al. (2002). Breast cancer on the World Wide Web: Cross sectional survey of quality of information and popularity of Websites. *BMJ, 324*(7337), 577-581.

Reents, S. (1999). *Impacts of the Internet on the doctor-patient relationship: The rise of the Internet health consumer.* Cyber Dialogue, Inc.

Sagaram, S., Walji, M., & Bernstam, E. (2002). Evaluating the prevalence, content and readability of complementary and alternative medicine (CAM) Web pages on the Internet. *Proceedings of the AMIA Symposium*, 672-676.

Sagaram, S., Walji, M., Meric-Bernstam, F., Johnson, C., & Bernstam, E. (2004). Inter-observer agreement for quality measures applied to online health information. In *Medinfo* (Vol. 2004, pp. 1308-1312).

Silberg, W. M., Lundberg, G. D., & Musacchio, R. A. (1997). Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor—Let the reader and viewer beware. *JAMA, 277*(15), 1244-1245.

Sullivan, D. (2004). *comScore Media Metrix Search Engine ratings.* Retrieved December 4, 2004, from http://searchenginewatch.com/reports/article.php/2156431

Tang, P. C., & Newcomb, C. (1998). Informing patients: A guide for providing patient health information. *JAMIA, 5*(6), 563-570.

Walji, M., Sagaram, S., Meric-Bernstam, F., Johnson, C. W., & Bernstam, E. V. (2005). Searching for cancer-related information online: Unintended retrieval of complementary and alternative medicine information. *International Journal of Medical Informatics, 74*(7-8), 685-693.

Walji, M., Sagaram, S., Sagaram, D., Meric-Bernstam, F., Johnson, C., Mirza, N. Q., et al. (2004). Efficacy of quality criteria to identify potentially harmful information: A cross-sectional survey of complementary and alternative medicine Web sites. *J Med Internet Res, 6*(2), e21.

This work was previously published in Web Mobile-Based Applications for Healthcare Management, edited by L. Al-Hakim, pp. 181-196, copyright 2007 by IRM Press (an imprint of IGI Global).

Appendix: Tables

Table 1. Commonly cited quality criteria

Quality Criterion	Operational Definition	Page (P) or Site (S) Review	Allowed Options	Interobserver Reliability*	
				% Agreement	*K*
Disclosure of authorship	Name of the person(s) or organization(s) present that is attributed as the creator or producer of the presented information	P	Yes No	0.857	-
Disclosure of ownership	Indication of the entity that owns the information presented on the Web site (e.g., organization logo)	S	Yes No	0.571	0.121
Sources clear	Claims backed up with a source (e.g., reference, expert opinion, or bibliography)	P	Yes No	0.905	0.800
Disclosure of sponsorship	Funding source or commercial aim/intent of the Web site or organization disclosed (e.g., organization mission on "About Us" page)	S	Yes No	0.238	-
Disclosure of advertising	Clear distinction (visual or by text) between advertising (paid piece of information or banner ad) and content (does not include product listings or a "buy now" link)	P	Yes No	0.143	-
Statement of purpose	General purpose or aims behind the Web site or organization may be found on front page, or under "About Us" or "Contact Us"	S	Yes No	0.524	0.186
General disclosures	Disclosure of either authorship, ownership, sponsorship, or currency of information	P	Yes No	0.667	0.310

continued on following page

Table 1. continued

Date of creation disclosed	Date disclosed when information was produced or reported (without the phrase "date created," the given date may be assumed to be the date created, such as with news organizations)	P	Yes No	0.857	0.504
Date of last update disclosed	Date disclosed of any revision or update	P	Yes No	0.905	-
Date of creation or update disclosed	Either creation or last-update date disclosed	P	Yes No	0.905	0.696
Authors' credentials disclosed	Disclosure of authority and qualification (MD, PhD, ND, etc.) of author. Disclosure does not include "Dr." or "Professor" (NA if no author)	P	Yes No NA	0.952	-
Credentials of physicians disclosed	Disclosure of credentials of physician (MD or including area of specialization) (NA if no author, author not physician, or not known if physician)	P	Yes No NA	0.857	-
Authors' affiliation disclosed	Disclosure of author's affiliations or relationships with relevant entity (NA if no author)	P	Yes No NA	0.905	0.738
Internal search engine present	Presence or absence of a search engine (can be any type of search engine, including a product search engine)	S	Yes No	0.857	0.715
Links provided	Presence of links (internal or external but not within the page; anchor related to topic for further information)	P	Yes No	0.762	0.516

continued on following page

Table 1. continued

References provided	Presence of conventional references or citations relevant to information on the page. Link to reference also acceptable	P	Yes No	0.952	0.897
Feedback mechanism provided	Author, editor, webmaster, or other official can be contacted. Presence of e-mail address, telephone, fax, or online form	S	Yes No	0.762	0.348
Fax number provided	Presence of a fax number for contact purposes	S	Yes No	0.762	0.516
E-mail address provided	Presence of an e-mail address for contact purposes. E-mail address does not include an e-form	S	Yes No	0.857	0.488
General disclaimers provided	Presence of a general disclaimer such as "Not a substitute for professional care" or "For educational purposes only," or a link to a disclaimer	S	Yes No	0.857	0.717
Copyright notice	Presence of a copyright notice	S	Yes No	0.857	0.710
Editorial review process	Presence of claim of use of an editorial review process or the listing of an editorial review committee or medical advisory board	S	Yes No	0.952	0.644

** Interobserver reliability from Bernstam, Sagaram, et al. (2005) using precalibration values.*

K = Cohen's kappa, which cannot be calculated for some data.

Table 2. Trustworthy Web sites

	Organization	URL	Topic
General Health Information	MedlinePLUS (National Library of Medicine, NIH, USA)	http://www.medlineplus.gov	Consumer-oriented health information
	Office of Disease Prevention and Health Promotion, Department of Health & Human Services (USA)	http://www.healthfinder.gov	Consumer-oriented health information (not cancer specific)
	National Institute of Health (USA)	http://clinicaltrials.gov	Information about government- and privately funded clinical research
	NHS Direct Online (UK)	http://www.nhsdirect.nhs.uk	General health information from the British government
Disease- or Treatment-Specific Health Information	National Center for Complementary and Alternative Medicine (NIH, USA)	http://nccam.nih.gov/	Complementary and alternative medicine information
	American Heart Association (USA)	http://www.americanheart.org	Information about heart-related illnesses
	National Cancer Institute (NIH, USA)	http://www.nci.nih.gov	Information about cancer

Chapter 7.13
The Effects of Confidentiality on Nursing Self-Efficacy with Information Systems

Diane Lending
James Madison University, USA

Thomas W. Dillon
James Madison University, USA

ABSTRACT

The objective of this study is to gain an understanding of nurses' perceptions of the confidentiality of computerized charts and determine if these perceptions influence nursing self-efficacy. A questionnaire was mailed to 600 hospital nursing staff just prior to the implementation of an integrated clinical and administrative hospital-wide information system. One hundred and thirty-nine questionnaires were returned for a response rate of 23%. The measurements consisted of perceived confidentiality, self-efficacy, self-reported technology use, self-assessed computer expertise, and attitude. We found that nurses recognize confidential information and that nurses do not think that computerized records are more or less confidential than paper records. Perceived confidentiality and self-efficacy are related. Because nurses that have lower self-efficacy also think that the data found in a computer is less confidential and that there is less of a need for confidentiality for computerized data, hospitals should take steps to improve self-efficacy by enhancing confidentiality training.

INTRODUCTION

Hospitals face countless constraints to fully implement information technology for nurses (Chang, Lutes, Braswell, & Nielson, 2006). The successful implementation of clinical and administrative information systems requires that attention be directed to both managerial and organizational issues (Dixon, 1999). These issues normally include, but are not limited to, the four components of a comprehensive technology assessment: safety, effectiveness, social impact, and cost/benefits (Happ, 1991; Pillar, Jacox, & Redman, 1990). This research focuses on the perceived confidentiality of the data, a safety factor, and the self-efficacy or confidence in their own use of technology of the nurses' that interact with the information system.

The Health Insurance Portability and Accountability Act (HIPAA) of 1996 directed the secretary of the U.S. Department of Health and Human Services (HHS) to begin the process of establishing standards for health information, securing that information, and protecting the privacy of the individuals to whom that information refers (NCVHS, 2000.) Confidentiality of patient data, required by law, must be guaranteed. At the same time, this data must be easily available to the nurses and clinicians that are providing care (Priest, 1989). Neither manually produced paper records nor medical records stored within a computerized administrative and clinical system are totally secure from loss or unauthorized use, but problems of confidentiality can be anticipated and minimized if properly managed (Priest, 1989). This is emphasized in the summarized recommendations from the 2001 AMIA Spring Congress, where Yasnoff, Overhage, Humphreys, and LaVenture (2001, p. 535) write, "all stakeholders need to be engaged in coordinated activities related to public health information architectures, standards, confidentiality, best practices and research: and that informatics training is needed throughout the public health workforce."

Self-efficacy is a user's confidence that he or she has the ability to use an information system. Self-efficacy has been shown to have a strong influence on an individual's adoption of an information system. In this research, self-efficacy is used to assist in identifying the user's understanding of the privacy of confidential data and how it may affect the installation of a clinical information system.

Research is needed to determine the impact of administrative and clinical information systems on quality of care, the nurses' perceptions of the confidentiality of data and the affects of confidentiality on self-efficacy. The purpose of this study is to:

1. Develop and empirically test measures for perceptions of confidentiality.
2. Determine if the nurse's perceptions of computerized charts' confidentiality and paper charts' confidentiality differ.
3. Determine if nurses perceive that computerized charts and paper charts need the same level of confidentiality.
4. Establish that perceptions of confidentiality influence nurse's self-efficacy.

Demonstrating these relationships will help us understand if the implementation of HIPAA regulations is likely to influence nurse's acceptance of clinical information systems.

BACKGROUND

Self-Efficacy

Self-efficacy was introduced to the information systems (IS) literature by Compeau and Higgins (1991) based upon Social Cognitive Theory (Bandura, 1982). Social Cognitive Theory suggests that there are two types of forces that guide behavior: expectations relating to outcomes of the behavior, and beliefs about one's ability to perform that behavior. These beliefs about one's ability are called self-efficacy (Compeau & Higgins, 1991). Compeau and Higgins recognized that an individual's reactions to an information technology would be affected by that individual's self-efficacy or confidence that they could use the technology. They created a 10-item measure for computer self-efficacy which is widely used in the IS literature (Compeau & Higgins, 1995) and which we have used in this study.

Research has shown that self-efficacy has strong influences on many aspects of an individual's adoption of information systems. Self-efficacy has been shown to have affects on attitudes toward computer systems, including ease-of-use and usefulness (e.g., Compeau & Higgins, 1995; Venkatesh, 2000) and on behaviors such as system use and early adoption (Agarwal, Sambamurthy, & Stair, 2000.)

Henderson, Deane, and Ward (1995) included self-efficacy as part of an overall assessment of computer-related anxiety for nurses and clerical workers. They reported that nurses had poorer scores on all of the psychological variables that were thought to have some association with computer avoidance and use. This included more negative attitudes and a lower self-efficacy.

In a study of hospital personnel's adoption of a computer-based medical information system (CBMIS), management support and ease-of-use were shown to have significant effects on self-efficacy (Henry & Stone, 1995.) Self-efficacy was shown to have an effect on the staff's job satisfaction. Because job satisfaction is strongly linked to employee turnover (Hollenback & Williams, 1996), this is of major concern for hospital management. In a second study that emphasized user satisfaction of a CBMIS, Henry and Stone (1994) indicated that the end-user's sense of self-efficacy and outcome expectancy had a direct, meaningful, and positive impact on end-user satisfaction with the CBMIS.

Ammenwerth, Mansmann, Iller, and Eichstadter (2002) performed a quantitative 2-year study of factors that affect user acceptance of computer-based nursing documentation. They found two influencing factors; previous acceptance of the nursing process and the previous amount of self-confidence, a similar measure to self-efficacy.

Many variables have been shown to influence an individual's self-efficacy, including gender, age, educational level, computer ownership, computer experience, professional orientation, training, organizational support, management support, encouragement, and computer attitudes (Marakas, Yi, & Johnson, 1998). Ogletree and Williams (1990) found that computer ownership and level of computer usage predicted self-efficacy. In a study of college students, use of one technology increased self-efficacy in other technologies; for example, requiring the students to use e-mail to communicate with professors was found to increase self-efficacy in word processing and use of CD-ROM databases (Ertmer, Evenbeck, Cennamo, & Lehman, 1994). Based upon this research, we expect that computer usage and attitudes will have significant effects on nurses' beliefs about their ability to use the integrated clinical and administrative information system.

Finally, in a study looking at the introduction of a clinical and administrative information system, several factors were found to influence nurses' self-efficacy (Dillon, et al., 2003). Any familiarity with technology including home computers, answering machines, or VCRs was associated with higher self-efficacy. Additionally, self-assessed expertise in e-mail, Internet, or word processing was associated with nurses' having a higher self-efficacy toward the clinical information system (Dillon, et al., 2003.)

Confidentiality

There are a limited number of studies available that evaluate benefits or the impacts of the nurse's perception of confidentiality of the computerized record as compared to the traditional paper record. This is a particularly interesting finding because confidentiality has repeatedly been identified as a major concern when implementing a computerized record.

Curran and Curran (1991) explored privacy and confidentiality issues in electronic information systems. They reported that 76% of the nurses admitted that they used the computer to obtain patient information that was not in their assigned duty areas. This indicates that nurses may not be aware of the ethical issues related to accessing information without permission (Rittman & Gorman, 1992).

Brodt and Stronge's (1986) Nurses' Attitudes Toward Computers Questionnaire includes two items that possibly measure a nurses' attitude toward security. These were identified in a psychometric examination of the instrument by Stockton and Verhey (1995). They labeled the two items under "Computers and personal security," but did not identify confidentiality as a factor.

Privacy and confidentiality of health information were two issues reported in the Health Information Privacy Surveys of 1993 and 1996 conducted by Lou Harris and Associates (1993, 1996; Anderson, 2000; Goldman, 1998). The 1993 survey found that 27% of patient respondents believed that their health information had been made available to others without their consent. In addition, 24% of health care providers reported violations of patient's privacy, though no perceptions of privacy concerns were reported. The 1996 survey found that 18% of the general public felt that the use of patient records for medical research without the patients' permission was inappropriate.

Concerns have been raised about differences in perceptions of privacy of information based on cultural values (Milberg, Burke, Smith, & Kallman, 1995). In addition, concerns about security and confidentiality were summarized in the General Accounting Office's report to the Committee on Governmental Affairs, United States Senate, on automated medical records (General Accounting Office, 1991, p. 22), when the committee wrote, " (Because) Automated records . . . tend to be accessible to many users in many locations and these users could potentially search thousands of files with relative ease, problems with security and privacy are magnified."

More recently, privacy concerns by patients and health care providers have been addressed in the development of policy documents, issues concerning the Internet, and the increased use of bedside and portable technologies. Kluge (2000) established seven basic principles for the delivery of health care information. These principles include privacy, openness, accessibility, legitimacy, accountability, and security. Safran and Goldberg (2000) examined the impact of the Internet on electronic patient records emphasizing the changes in workflow and work patterns. Perception of privacy and confidentiality by health care workers were not included in either discussion. Fischer, Stewart, Mehta, Wax, and Lipinsky (2002) reviewed the literature for handheld computing devices and summarized issues, such as patient confidentiality, costs, and specialty specific uses of personal digital assistants.

Attitude

Attitude has long been used as a determinant for a user's intention to use or adopt an information system. Defined as a predisposition to respond in a consistently favorable or unfavorable manner, attitude concerns what the user knows or believes about the system, how much they like or dislike the system, and how they routinely act or behave regarding the system (Ajzen & Fishbein, 1980). Numerous studies have found a strong relationship between nurses' attitudes, efficacy, and systems use or adoption (e.g., Dillon et al., 2003; Dillon et al, 2005.)

In summary, the self-efficacy of nurses and their perceptions of data security and confidentiality are relatively unknown. Nursing research has investigated nurses' attitudes toward computers and effects on organizational procedures, but research is necessary to understand the nurse's self-efficacy, perceptions of the confidentiality of data from both paper and computerized systems, and the relationship for perceptions of confidentiality to efficacy.

METHOD

In order to explore if a nurse's perceptions of confidentiality influences self-efficacy, if nurses perceive that computerized and paper charts need the same level of confidentiality, and if the nurses' perceptions of computerized and paper charts' confidentiality differ we surveyed the nurses at a regional hospital center just prior to a system implementation. Our model was based upon Dillon, et al. (2003) which showed that self-reported computer use, self-assessed computer expertise, and nursing attitudes toward the information system predicted self-efficacy. We added three new factors to the model from McDowell, Dillon, Salimian, and Conklin (1998) for this study, including perceptions of (1) paper records confidentiality, (2) computerized records confidentiality, and (3) importance of confidentiality. Reliability and validity analysis were performed on all multi-item factors.

Setting

The survey was conducted at 350-bed regional hospital center in the south central United States just prior to the implementation of a new integrated clinical and administrative hospital-wide information system. All of the nursing staff at the hospital was aware that the information system was being installed. Within 30 days the new system would go "online" in most of the nursing units. The remaining units would "go live" in about 60 days after that. Prior to the new system implementation the only computerized nursing application was the ordering of medications from the hospital pharmacy. Nurses had no prior information systems experience. The hospital administration prepared the nursing staff by offering periodic training sessions. The nursing staff was kept abreast of the implementation timetable via the nurse managers that supervised each unit and information printed in a hospital-wide newsletter.

Data Collection Procedures

Surveys were distributed via home mail by the research team with the full support and assistance of the hospital. Institutional Review was approved by the hospital and the research team's university. All part-time (minimum of 12 hours per week) and full-time members of the hospital nursing staff (i.e., over 600 staff members) received the survey 30 days prior to the implementation of the integrated clinical and administrative hospital-wide information system. All participants received an addressed envelope that contained 1) a cover letter that solicited the respondents' cooperation and assured anonymity, 2) the survey instrument, 3) an entry form for a drawing ($100 gift certificate), and 4) a return envelope addressed to the research team. One hundred and thirty-nine surveys were returned prior to the drawing deadline, for a response rate of 23%. Ten of the 139 drawing-entries were selected randomly, the winning nurses were contacted, and prizes were delivered the following day. All hospital employees were informed of the drawing winners via the hospital newsletter. Table 1 shows the demographic characteristics of the respondents.

The Instrument

In order to better ascertain the factors that may affect the nurses' perception of confidentiality and self-efficacy, the questionnaire contained seven types of questions: perceptions of confidentiality with paper records, perceptions of confidentiality with computerized record, the importance of confidentiality, self-efficacy of the new information system, self-reported computer and electronic devices use, self-assessed computer expertise,

Table 1. Demographics of Survey Respondents

Variable		Number	Percentages
Work	Full Time	107	77.0%
	Part Time	32	23.0%
Sex	Female	132	95.0%
	Male	6	4.3%
	Unreported	1	.7%
Age	20-29	33	23.7%
	30-39	45	32.4%
	40-49	42	30.2%
	50-59	15	10.8%
	60 or over	2	1.4%
	Unreported	2	1.4%
Education	LPN	14	10.1%
	RN Diploma	5	3.6%
	RN Associates Degree	89	64.0%
	RN Bachelors	27	19.4%
	Masters Degree	4	2.9%

and general attitudes toward the information systems. In addition, a general comment question was provided.

MEASURES

Perceived Confidentiality

The confidentiality instrument was developed in an earlier study (McDowell et al., 1998). The current study marks the first empirical test of the instrument. The instrument began with eight forced choice items that were based on the literature review and the output from a seminar of nursing informatics graduate students. All nurses in the seminar had a wide range of experience with hospital information systems.

In the earlier study, these items were reviewed for content validity by eight nurses working in a critical care area of a regional medical center that had implemented a computer-based health care information system with a computerized patient record in the last 12 months. Nurses were asked to review item content for appropriateness in evaluating nurses' perception of confidentiality of the patient's medical record. In addition, the information systems nurse at the same medical center was asked to review the statements (McDowell et al., 1998).

Recommendations from these nine nurses were incorporated into a 10-item instrument. In each case, five identical questions were written for paper and computerized records. The questions were measured on a Likert-Scale with 1 indicating Strongly Agree and 7 indicating Strongly Disagree. We expected the questions to form three constructs, as shown in Table 2.

Reliability analysis was done with the 139 responses to see if these items associated as three separate constructs. In testing content validity with factor analysis, we found that the 10 items loaded on four factors using the principal components method with varimax rotation, as shown in Table 3. Table 4 interprets the factors and shows the Cronbach's alpha for each factor.

These factors were not exactly as expected, but were very interesting. The four items reflecting

Table 2. Confidentiality questions and expected constructs

Are Paper Records Confidential?	
If I were the patient, I would feel my traditional paper patient record was protected or secure from unauthorized use.	Papconf1
With the traditional paper patient record it is difficult for me to read information on patients that I am not currently caring for.	Papconf2
Visitors or patient's families can easily read confidential information written on various patient records (Reverse scored)	Papconf3

Are Computerized Records Confidential?	
If I were the patient, I would feel my computerized patient record was protected or secure from unauthorized use.	ISconf1
Visitors or patient's families will be able to easily view confidential information on the computer screen (Reverse scored)	ISconf2
With a computerized patient record it will be difficult for me to retrieve information on a patient that I am not currently caring for.	ISconf3

How Important is Confidentiality?	
Copies of pages from a patient's paper record are not considered confidential material.	Impconf1
Printouts of data from a computerized system that have a patient's name or number on them, but are not part of the patient record, do not need to be considered confidential information.	Impconf2
I feel it is not a breech of confidentiality if I read the chart of a patient that I am not caring for.	Impconf3
I feel it will not be a breech of confidentiality if I look up computerized data about a patient that I am not caring for.	Impconf4

need for records confidentiality did not load as a single construct. Rather, the like items loaded separately into two separate factors: Factor 1 items can be identified as "what is a breech of confidentiality?" and Factor 2 items loaded as "recognizing what was confidential?". Factor 2 also included whether it is easy for visitors to read paper confidential records, demonstrating that nurses recognize that easy-to-read records are still confidential.

Two of the three items reflecting the perception of information systems confidentiality clustered together in Factor 3. Two of the three items reflecting the perception of paper records confidentiality clustered together in Factor 4. However, Factor 4 has a very low Cronbach's alpha and should not be considered a reliable measure.

Self-Efficacy Toward the New Information System

Self-efficacy was measured using the 10 questions developed by Compeau and Higgins (1995), which are found in Table 5. A sample question was: "*I could complete the job using the <name omitted> system if there was no one around to tell me what to do as I go.*" The scale was a 10-point Likert scale ranging from 1 = Not at all Confident to 10 = Totally Confident. The self-efficacy score ranged from 1 to 100, a higher score displayed a

Table 3. Confidentiality factors found

	Components			
	1	**2**	**3**	**4**
Papconf1	.115	-8.264E-03	-6.184E-02	**.769**
Papconf2	-4.508E-02	-.135	9.454E-03	**.448**
Papconf3	-.139	**.651**	-.287	-3.277E-02
ISconf1	6.707E-02	.186	**.760**	.303
ISconf2	.109	.205	**-.752**	.382
ISconf3	-.301	.121	.217	.387
Impconf1	4.369E-02	**.773**	7.069E-02	-.144
Impconf2	.245	**.626**	.191	.4889E-02
Impconf3	**.875**	9.375E-03	9.134E-03	6.873E-02
Impconf4	**.870**	.136	1.900E-03	-7.447E-02

Table 4. Interpretation of factors

Factor	**Items**	**Cronbach's Alpha**
Factor 1: Confidentiality Breech	I feel it is not a breech of confidentiality if I read the chart of a patient that I am not caring for. I feel it will not be a breech of confidentiality if I look up computerized data about a patient that I am not caring for.	.7657
Factor 2: Recognize what is confidentiality	Copies of pages from a patient's paper record are not considered confidential material. Printouts of data from a computerized system that have a patient's name or number on them, but are not part of the patient record, do not need to be considered confidential information. Visitors or patient's families can easily read confidential information written on various patient records (Reverse scored)	.4194
Factor 3: Perceptions of computerized charts confidentiality	If I were the patient, I would feel my computerized patient record was protected or secure from unauthorized use. Visitors or patient's families will be able to easily view confidential information on the computer screen (Reverse scored)	.5564
Factor 4: Perceptions of paper charts confidentiality	If I were the patient, I would feel my traditional paper patient record was protected or secure from unauthorized use. With the traditional paper patient record it is difficult for me to read information on patients that I am not currently caring for.	.0735
Did not load on a factor	With a computerized patient record it will be difficult for me to retrieve information on a patient that I am not currently caring for.	N.A.

higher level of self-efficacy. Reliability analysis was done on the results of the survey to ensure that these items assessed a single construct. The Cronbach's alpha calculated for the 10 items was .9320, indicating excellent internal consistency. In testing content validity with factor analysis, we found that the 10 items loaded on two factors using the principal components method with varimax rotation. This indicated that the 10 items were two separate measures of self-efficacy. While most prior research reports unidimensionality of the self-efficacy construct, this finding is consistent with research such as Agarwal et al. (2000).

Self-Reported Computer and Electronic Devices Use

Two questions were asked to determine computer use. The first question requested the number of hours per week that the nurse used a home computer. The second question requested the number of years the nurse had used a computer for home or work. In addition, because self-efficacy has been related to the prior use of new technologies, information was gathered about the use of home and office electronic devices. The leading question was asked, *"Do you use any of the following at home or work?"* followed by: answering machine, cellular phone, VCR (video cassette recorder), and pager (beeper) (Settle, Dillon, & Alreck, 1999).

Self-Assessed Computer Expertise

Skill in various computer application uses was measured on a five-point Likert-scale. A sample question was "*Rate your ability to use… Word Processing*" with the available answers: beginner, novice, average, experienced, and expert. Questions were asked about computer expertise in general and word processing, Internet search, and e-mail ability.

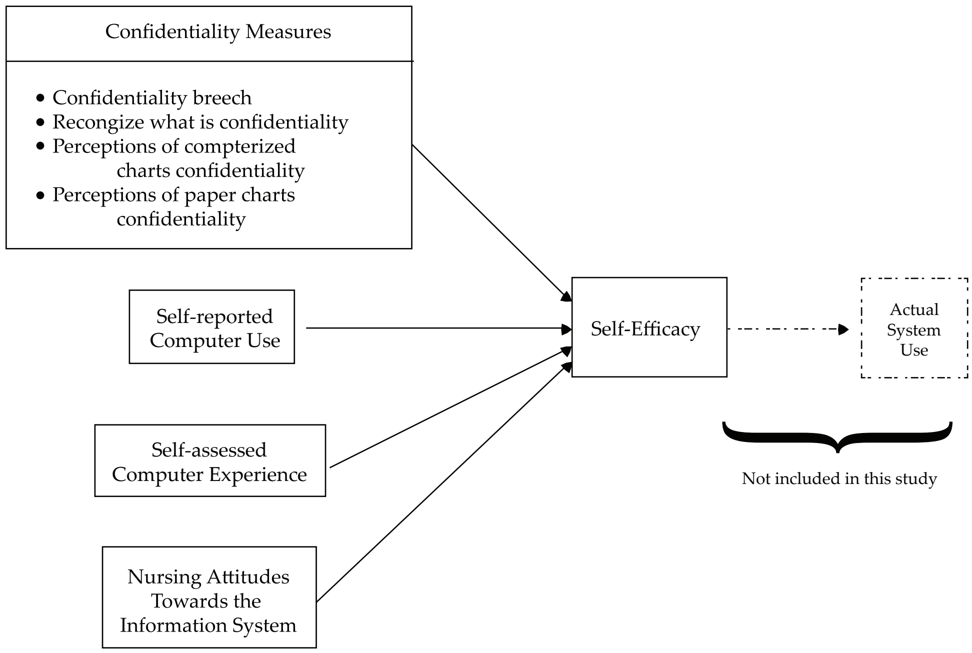
Attitudes

Attitudes toward the information systems were measured with five questions measured on a seven-point Likert scale. A sample question was: "*All things considered, using the new computerized <name omitted> patient record system is*" where 1 is Good and 7 is Bad. Reliability analysis

Table 5. Self-efficacy questions

I could complete the job using the <name omitted> system
1…if there was no one around to tell me what to do as I go
2…if I had never used a package like it before
3… if I had only the software manuals for reference
4…if I had seen someone else using it before trying it myself
5…if I could call someone for help if I got stuck
6…if someone else had helped me get started
7…if I had a lot of time to complete the job for which the software was provided
8…if I had just the built-in help facility for assistance
9…if someone showed me how to do it first
10…if I had used similar packages before this one to do the same job

Figure 1. Proposed reasearch model



was done to ensure that these items assessed a single construct. The Cronbach's alpha calculated for the five items was .9649, indicating excellent internal consistency. The five items loaded on a single factor using the principal components method with varimax rotation, supporting the unidimensionality of the attitude construct.

RESEARCH QUESTIONS AND MODEL

With the results of our reliability analysis, we slightly revised our research questions into the following questions:

- **Question 1:** Do nurses recognize what confidentiality is?
- **Question 2:** Are perceptions of computerized charts' confidentiality and paper charts' confidentiality different?
- **Question 3:** Are there differences in perceptions of breeches of confidentiality when looking at paper records or looking at computerized records for a patient the nurse is not caring for?
- **Question 4:** Is self-efficacy influenced by a nurse's perceptions of confidentiality or perceptions of the need for confidentiality? Prior research (Dillon et al., 2003) had shown that prior use of computers, computer expertise, and attitudes toward computers have a significant effect on self-efficacy. If the measures tested in research questions 1 through 3 show that perceptions of confidentiality are meaningful to nurses, can we improve the prediction of self-efficacy? In testing research question 4, we add perceptions of confidentiality to the self-efficacy model (Dillon et al., 2003) to get the model shown in Figure 1.

RESULTS

- **Question 1:** Do nurses recognize what confidentiality is?

The statements related to recognizing what confidentiality is were:

- Copies of pages from a patient's paper record are not considered confidential material
- Printout of data from a computerized system that have a patient's name or number on them, but are not part of the patient record, do not need to be considered confidential information.

Nurses strongly disagreed with both of these statements indicating that they recognize what confidential patient information is. The mean value for the first variable was a 6.56 (where 7 was highly disagree.) The mean value for the second variable was 6.12. A t-test to compare the means showed that there was a significant difference between these two means ($p < 0.001$.) This difference in means may be attributable to difference in wording between the two questions. The second question specifically indicates that the computerized information is not part of the patient record, while the first says that the paper information is part of the patient record.

- **Question 2:** Are perceptions of computerized charts' confidentiality and paper charts' confidentiality different?

Using a t-test, we compared the means of the three measures of record confidentiality. Our analysis shows that nurses consider it easier to read the records of a patient that they are not caring for if the records are on paper (a mean value of 4.75 for paper, as opposed to 3.56 for computerized records, $p < 0.001$). However, there was no significant difference in whether their perceptions of whether patients will think computerized records are more protected than paper records or whether visitors will be more able to see paper records than computer records. It may be that these second-hand perceptions are not salient to the nurses.

*Table 6. Best Model for Self-efficacy**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.644	.415	.387	13.499

Model	Sum of Squares	Df	Mean Square	F	Sign.
Regression	15910.788	6	2651.798	14.553	.000
Residual	22412.481	123	182.215		
Total	38323.259	129			

Independent Variable	Beta	Std. error	t	Sig.
Constant	19.553	11.396	1.716	.089
Word Processing expertise	4.174	1.370	3.047	.003
Attitude toward the info system	-4.131	.966	-4.277	.000
Confidentiality Factor 1	2.078	.723	2.872	.005
VCR Use	20.221	9.873	2.048	.043
E-mail expertise	3.142	1.228	2.559	.012
Answering Machine Use	7.648	3.522	2.172	.032

**dependent variable is self-efficacy*

- **Question 3:** Are there differences in perceptions of breeches of confidentiality in looking at paper records or looking at computerized records for a patient the nurse is not caring for?

Using a t-test, we compared the means of these two measures of perceptions of confidentiality breeches. There was a significant difference in these perceptions with the mean for reading paper records being 4.87 and the mean for reading computerized records 5.15 (where 7 = strongly disagree, $p < .05$.) Nurses consider that it is more of a breech of confidentiality to look at computerized patient charts than paper patient charts.

- **Question 4:** Is self-efficacy influenced by a nurse's perceptions of confidentiality or perceptions of the need for confidentiality?

We ran a multivariate analysis with self-efficacy as the dependent variable. The independent variables were the seven constructs in Figure 1: confidentiality breech, recognition of "what confidentiality is," the perception of confidentiality of paper records, the perception of confidentiality of computerized records, self-reported computer use, self-assessed computer expertise, and nursing attitudes toward the information system. Those variables that were significant in the relationship are shown in Table 6. Only one of the confidentiality factors was significant in the relationship, Factor 1, whether it was a breech of confidentiality to look at records for a patient that the nurse was not caring for. The less the nurse felt that it would be a breech of confidentiality, either on paper or from a computerized information system, the lower the nurse's self-efficacy about the new system. The adjusted R-square for this model is higher than the model without the Confidentiality factor.

While Confidentiality Factor 3, "Perceptions of computerized charts confidentiality," was not part of the best model predicting self-efficacy, a regression run between self-efficacy and Factor 3, "Perceptions of computerized charts confidentiality" was significant. Nurses that perceived computerized charts as more confidential had a higher self-efficacy than did nurses that perceived paper charts as more confidential. Table 7 shows this regression model.

DISCUSSION AND CONCLUSION

Hospitals are currently in the early stages of a move from paper to total electronic transactions and records. These new electronic information systems are closer to the patient and expand the number of clinical personnel that must interact with the new technology. By broadening the user base, there is an increasing need for an understanding of nurses' self-efficacy and perceptions of data confidentiality. The objective of this research was four fold. First, do nurses recognize what confidentiality is? Second, are perceptions of computerized charts' confidentiality and paper charts' confidentiality different? Third, are there differences in perceptions of breeches of confidentiality in looking at paper records or looking at computerized records for a patient the nurse is not caring for? And fourth, does self-efficacy influence a nurse's perceptions of confidentiality or perceptions of the need for self-efficacy?

Confidentiality Issues

Our findings contradicted those of Curran and Curran (1991) and Rittman and Gorman (1992), when we found that nurses do recognize what confidential patient information is. These findings may be the result of enhanced training programs provided by hospitals after privacy and confidentiality issues were publicized broadly by the General Accounting Office's report (1991) or the recent application of the Health Insurance Portability And Accountability Act of 1996.

*Table 7. Self-efficacy and confidentiality of computerized records**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.199	.040	.032	16.677

Model	Sum of Squares	Df	Mean Square	F	Sign.
Regression Residual Total	1507.510 36434.881 37942.391	1 131 132	1507.510 278.129	5.420	.021

Independent Variable	Beta	Std. error	t	Sig.
Constant Confidentiality Factor 3	64.230 -2.829	4.363 1.215	14.721 -2.328	.000 .021

**dependent variable is self-efficacy*

The goal of HIPAA as stated, is to "improve the portability and continuity of health insurance coverage in the group and individual markets, to combat waste, fraud, and abuse in health insurance and health care delivery, to promote the use of medical savings accounts, to improve access to long-term care services and coverage, to simplify the administration of health insurance, and for other purposes." A major function of the law is to prevent the wrongful disclosure of individually identifiable health information. Not adhering to HIPAA may result in financial penalties to the organization.

Recognition of the problems created by the accessibility of electronic patient records (Anderson, 2000; Safran & Goldberg, 2000) and changes in the laws concerning patient privacy (HIPAA, 1996) has lead to the acknowledged need and development of basic privacy and confidentiality principles (Kluge, 2000). These principles are now more fully integrated into all health care professions and organizations, enhancing the nurse's recognition of confidentiality issues (Kilbridge, 2003).

We found no significant differences in the level of confidentiality that nurses perceive in computerized and paper charts, although, nurses perceived that it was significantly easier to read the paper records of patients not in their care. We anticipated that paper-charting data would be considered more confidential because the setting of the study did not permit the nurses to use computerized data at this point in time and that the simplicity of search engines when using the Internet to retrieve general information via-keyword searches would sway respondents (Safran & Goldberg, 2000). A behavioral factor may be affecting this outcome. The computerized patient charting system is an unknown, but the nursing staff had been attending training sessions concerning the impending implementation. The realization that electronic patient records require specific identifiers, such as patient identification numbers, to access information as compared to simply reading the chart in hand may also be a factor.

Computerized charts and paper charts do need the same level of confidentiality, but the nurses surveyed in our research felt that it was more of a breech of confidentiality to access patient information from a computerized patient chart than information from a paper chart. This was in agreement with the previous findings that the nurses felt that information on paper records was easier to access (i.e., computerized patient data was more difficult to access, and thus is considered more confidential). Surprisingly, comments by those nurses that completed the survey were generally neutral when attempting to find a distinction between computerized and paper data. "Anything concerning a patient's information should be respected as confidential," was a comment made by a number of those providing comments on the survey.

Effects of Confidentiality Issues on Self-Efficacy

Previous research has identified a number of factors that affect a nurse's self-efficacy toward a computer system. For example, Dillon et al. (2003) found that familiarity with home electronic technology, such as a home computer or VCR, at least average levels of expertise with common software applications (Internet, e-mail, and word processing), and a positive attitude toward the computer system were associated with higher levels of self-efficacy. In addition, they also found that younger and better-educated nurses had a slight association to self-efficacy. A goal of this study was to extend these finding by determining if perceptions of confidentiality or perceptions of the need for confidentiality would also influence the nurse's self-efficacy of a computer system.

A nurse's perceptions of confidentiality and a nurse's perceptions of the *need* for confidentiality do influence the nurse's self-efficacy. We found that nurses that have lower self-efficacy also think that the data found in the computer is less confidential and that there is less of a *need* for computerized data to be confidential. Considering the importance of patient data confidentiality, we feel that these findings are important. Self-efficacy is a user's confidence that he or she has the ability to use an information system. Those showing a lower confidence with computerized charting are also showing a lower sense of worth for the data found in the computerized charting system. There are a number of steps that can be taken to improve the self-efficacy of the nursing staff. These include group or individualized training, availability of "practice" technology, and the accessibility of peer mentors.

In conclusion, self-efficacy refers to an individual's belief that he or she possesses the requisite skills and abilities to accomplish a given task. Self-efficacy determines the individual's level of persistence to learn a task and is influenced by the perceptions of the confidentiality for computerized patient data. Improvements of the nurses' self-efficacy toward an information system can be induced in several ways. The most obvious are to include tasks in training that foster self-confidence regarding their ability to use the system and enhance their understanding data confidentiality.

Limitations of the Study

Like any research, this study has limitations. First, the study was done at a single hospital in a single area. There may be something particular about this hospital or this area which affects the generalizability of the results. Second, the survey methodology used only allows us to find associations. We imply that confidentiality influences self-efficacy, but in actuality, we can only show that the two are associated. A time-series experiment would be needed to show causation. Finally, all measures are based upon perceptions prior to implementation of the system. We capture perceived confidentiality, attitudes, and self-efficacy after training, but prior to use. Future research could profitably study these measures after implementation.

REFERENCES

Agarwal, R., Sambamurthy, V., & Stair, R.M. (2000). Research report: The evolving relationship between general and specific computer self-efficacy – an empirical assessment. *Information Systems Research*, *11*(4), 418-430.

Ajzen, I. & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice Hall.

Ammenwerth, E., Mansmann, U., Iller, C., & Eichstadter, R. (2002). Factors affecting and affected by user acceptance of computer-based nursing documentation: Results of a two-year study. *Journal of the American Medical Informatics Association*, *10*, 69-84.

Anderson, J.R.(2000). Security of the distributed electronic patient record: A case-based approach to identifying policy issues. *International Journal of Medical Informatics*, *60*(2), 111-118.

Bandura, A.(1982). Self-efficacy mechanism in human agency. *American Psychologist*, *37*(2), 122-147.

Brodt, A., & Stronge, J.H. (1986). Nurses' attitudes toward computerization in a mid-western community hospital. *Computers in Nursing*, *4*, 82-86.

Chang, K., Lutes, K.D., Braswell, M.L., & Nielson, J.K. (2006). Nurses' perceptions of using a pocket pc for shift reports and patient care. *International Journal of Healthcare Information Systems and Informatics*, *1*(1), 52-63.

Compeau, D.R., & Higgins, C.A. (1991, December16-18). A social cognitive theory perspective on individual reactions to computing technology. In J. DeGross et al. (Eds.), *Proceedings of the Twelfth International Conference on Information Systems* (pp. 187-198).

Compeau, D.R., & Higgins, C.A.(1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly*, *19*(2), 189-211.

Curran, M., & Curran, K. (1991). The ethics of information. *JONA*, *21*(1), 47-49.

Dillon, T.W., Crews, T.R., & Blankenship, R. (2005). Nursing attitudes and images of electronic patient record systems. *CIN: Computers, Informatics, Nursing*, *23*(3), 139-145.

Dillon, T.W., Lending, D., Crews, T.R., & Blankenship, R. (2003). Nursing self-efficacy of an integrated clinical and administrative information system. *CIN: Computers, Informatics, Nursing*, *21*(4), 198-205.

Dixon, R.D. (1999). The behavioral side of information technology. *International Journal of Medical Informatics*, *56*, 117-123.

Ertmer, P.A., Evenbeck, E., Cennamo, K.S., & Lehman, J.D. (1994). Enhancing self-efficacy for computer technologies through the use of positive classroom experiences. *Educational Technology Research and Development*, *42*(3), 45-62.

Fischer, S., Stewart, T.E., Mehta, S., Wax, R., & Lapinsky, S.E. (2002). Handheld computing in medicine. *Journal of the American Medical Informatics Association*, *10*, 139-149.

General Accounting Office. (1991, January). Medical ADP systems automated medical records hold promise to improve patient care. *Report to the Chairman, Committee on Governmental Affairs, U. S. Senate* (Rep. No. IMTEC-91-5).

Goldman, J. (1998). Protecting privacy to improve health care. *Health Affairs*, *17*(6), 47-60.

Happ, B. (1991). Technology assessment and computers in nursing. *CARING Newsletter*, *6*(1), 3.

Health Insurance Portability and Accountability Act Of 1996, Public Law 104-191, 104th Congress. Retrieved March 15, 2007, from http://aspe.hhs.gov/admnsimp/pl104191.htm#1177.

Henderson, R.D., Deane, F.P., & Ward, M.J. (1995). Occupational differences in computer-related anxiety: Implications for the implementation of a computerized patient management information system. *Behaviour & Information Technology*, *14*(1), 23-31.

Henry, J.W., & Stone, R.W. (1994). A structural equation model of end-user satisfaction with a computer-based medical information system. *Information Resources Management Journal*, *7*(3), 21-33.

Henry, J.W., & Stone, R.W. (1995). Computer self-efficacy and outcome expectancy: The effects on the end-user's job satisfaction. *Computer Personnel*, *16*(4), 15-34.

Hollenback, J.R., & Williams, C.R. (1996). Turnover functionality vs. turnover frequency: A note on work attitudes and organizational effectiveness. *Journal of Applied Psychology*, *73*(4), 647-656.

Kilbridge, P. (2003). The cost of HIPAA compliance. *The New England Journal of Medicine*, *348*(15), 1423-1424.

Kluge, E.W. (2000). Professional codes for electronic HC record protection: Ethical, legal, economic, and structural issues. *International Journal of Medical Informatics*, *60*(2), 85-96.

Lou Harris and Associates, Health Information Privacy Survey, 1993. Harris, New York, 1992, 1993, 1995, 1996. (Reported in Goldman, 1998).

Marakas, G.M., Yi, M.Y., & Johnson, R.D. (1998). The multilevel and multifaceted character of computer self-efficacy: Toward clarification of the construct and an integrative framework for research. *Information Systems Research*, *9*(2), 126-163.

McDowell, D., Dillon, T.W., Salimian, F., & Conklin, D. (1998). A preliminary study: Capturing perceived accuracy and confidentiality of traditional and computerized patient records. In *Proceedings of the Association for Information Systems* (pp. 780-783).

Milberg, S., Burke,S., Smith, J., & Kallman, E. (1995). Values, privacy, personal information and regulatory approaches. *Communications of the ACM*, *38*(12), 65-74.

NCVHS. (2000, August 21). The National Committee on Vital and Health Statistics. Retrieved March 15, 2007, from www.ncvhs.hhs.gov.

Ogletree, S.M., & Williams, S.W. (1990). Sex and sex-typing effects on computer attitudes and aptitude. *Sex Roles*, *23*(11/12), 703-712.

Pillar, B., Jacox, A.K., & Redman, B.K. (1990). Technology, its assessment, and nursing. *Nursing Outlook*, *38*(1), 16-19.

Priest, S. (1989). *Understanding computer resources a health care perspective*. Owings Mills, MD: National Health Publishing.

Rittman, M.R., & Gorman, R.H. (1992). Computerized databases: Privacy issues in the development of the nursing minimum data set. *Computers in Nursing*, *10*(1), 14-18.

Safran, C., & Goldberg, H. (2000). Electronic patient records and the impact of the internet. *International Journal of Medical Informatics*, *60*(2), 77-83.

Settle, R.B., Dillon, T.W., & Alreck, P.L. (1999). Acceptance of the phone-based interface for automatic call direction. *Behaviour and Information Technology*, *18*(2), 97-108.

Stockton, A.H., & Verhey, M.P. (1995). A psychometric examination of the stronge-brodt nurses' attitudes toward computers questionnaire. *Computers in Nursing*, *13*(3), 109-113.

Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research*, *11*(4), 342-365.

Yasnoff, W.A., Overhage, J.M., Humphreys, B.L., & LaVenture, M. (2001). A national agenda for health informatics. *Journal of the American Medical Informatics Association*, *8*(6), 535-545.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 2, Issue 3, edited by J. Tan, pp. 49-64, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.14
Standardization in Health and Medical Informatics

Josipa Kern
Andrija Stampar School of Public Health, Zagreb University Medical School, Croatia

ABSTRACT

Standard is a thing or quality or specification by which something may be tested or measured. The development of standards is organized on a global, international level, existing also on a national level, well harmonized with an international one. International developers are organizations working on this matter, like the International Organization for Standardization (ISO) or the European Committee for Standardisation (Comité Européen de Normalisation—CEN). Standards in health and medical informatics enable access to patient health records to read or to add some new data relevant to other healthcare providers taking care of a patient. Bad medical informatics can lead to patient deaths, and standardization in the field can prevent this from happening.

INTRODUCTION

When things go well, often it is because they conform to standards (International Organization for Standardization [ISO], 2005). In the Oxford Dictionary of Modern English, there is a lot of explanation of what the word standard means, but, in the context of the first sentence, the best meaning is the following: "standard is a thing or quality or specification by which something may be tested or measured." A personal computer is a standardized computer. This means that all of its components are made according to strictly defined specification. Consequently, it does not matter who produces the components or where they are produced.

Industry put in the first demand for standards. Standardization is especially important for electronics, and for ICT and its application in different areas. Nowadays, the developing of standards is organized on a global, international level, but it exists also on the national level, which is well harmonized with the international one.

Developers of standards are organizations and groups working on this matter. The leading standard developer in the world is the International Organization for Standardization. ISO is a nongovernmental organization that was established February 23, 1947. Its mission is to promote the development of standardization and related activities in the world with a view to facilitating the international exchange of goods and services, and to developing cooperation in the spheres of intellectual, scientific, technological, and economic activity (International Organization for Standardizaton, 2005). ISO collaborates with its partner in international standardization, the International Electrotechnical Commission (IEC), a nongovernmental body whose scope of activities complements ISO's. The ISO and the IEC cooperate on a joint basis with the International Telecommunication Union (ITU), part of the United Nations organization whose members are governments. The ISO standard can be recognized by the ISO logo, ISO prefix, and the designation "International Standard."

The European developer of standards is the European Committee for Standardisation (Comité Européen de Normalisation, CEN). It was founded in 1961 by the national standards bodies in the European Economic Community and EFTA countries. CEN promotes voluntary technical harmonization in Europe in conjunction with worldwide bodies and its partners in Europe, and the conformity assessment of products and their certification (Comité Européen de Normalisation, 2005). CEN cooperates with the European Committee for Electrotechnical Standardization (CENELEC) and the European Telecommunications Standards Institute (ETSI). A product of this cooperation is the European standard, which can be recognized by the prefix EN. Any added prefix to the existing one, for both the ISO and CEN standards, means that this standard is the result of cooperation with another standardization group or organization. The prefix ENV in European standardization means that this standard is not yet a full standard (it is under development by CEN).

ISO and CEN have technical committees working in specific areas. ISO/TC215, established in 1998, and CEN/TC251, established in 1991, are corresponding technical committees working on standardization in health and medical informatics. Both standardization bodies cooperate and mutually exchange their standards.

There are also a variety of other organizations and groups developing standards, either cooperating with ISO and CEN or acting as administrators and coordinators in standardization. For example, there are Health Level 7 (HL7); Digital Imaging and Communications in Medicine (DICOM); the American National Standards Institute (ANSI), a nonprofit organization that administers and coordinates the U.S. voluntary standardization and conformity assessment system, and so forth.

BACKGROUND

Definition

A standard is a set of rules and definitions that specify how to carry out a process or produce a product, or more precisely, a standard is a document established by consensus and approved by a recognized body that provides, for common and repeated use, guidelines or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context.

The main role of a standard is raising the levels of quality, safety, reliability, efficiency, and interchangeability, and consequently lowering costs (International Organization for Standardization, 2005).

Standard Creation Process

There are several phases in the process of standardization. The first phase of this process is characterized by demand for a standard. There must be someone who needs a standard. Most standards are prepared at the request of industry. The European Commission can also request the standards bodies to prepare standards in order to implement European legislation. This standardization is mandated by the Commission through the Standing Committee of the Directive in support of the legislation. Groups of users can also ask for a standard in the field of their interest. The second phase of standardization is the developing of standards by following specifications based on the needs defined in the first phase. Experts of a specific field work on related standards. After the standard has been approved by the standardization body, it becomes a prototype and goes through testing and evaluation. Positive test results imply the dissemination of standards, and they start to take effect. It should be highlighted that a standard is dynamic, and it changes time after time. Most standards require periodic revision. Several factors combine to render a standard out of date: technological evolution, new methods and materials, and new quality and safety requirements. To take account of these factors, the general rule has been established that all standards should be reviewed at intervals of not more than a predefined time period. On occasion, it is necessary to revise a standard earlier (Hammond & Cimino, 2001).

Standardization in Health and Medical Informatics: Why and What

Standardization has been a major factor in companies' financial and clinical success, enabling faster implementation, greater quality control, and significant cost savings (Ball, Cortes-Comerer, Costin, Hudson, & Augustine, 2004). The contribution of the standardization process in healthcare terminology initiated by CEN/TC251 and supported now by the work of CEN/TC215/WG3 to this new approach can be summarized as the practical realization of ontology (Rodrigues, Trombert Paviot, Martin, Vercherin, & Samuel, 2002). The standard CEN ENV 12924 contains a security categorization model for information systems in healthcare, distinguishing six categories plus some refinements. For each category, it specifies the required protection measures (Louwerse, 2002). Standards support interoperability and electronic-health-record (EHR) communication. They support cooperative work among health agents when it is necessary to share health-care information about patients in a meaningful way. Examples of requirements for EHR are provided in four themes: EHR functional requirements; ethical, legal, and security requirements; clinical requirements; and technical requirements. The main logical building blocks of an EHR use the terminology of CEN/TC251 ENV13606 (Lloyd & Kalra, 2003; Maldonado, Crespo, Sanchis, & Robles, 2004; Marley, 2002). Many specific medical records, like medical records of patients suffering from beta-thalassaemia, which are inevitably complex and grow in size very fast, are based also on ENV 13606 (Deftereos, Lambrinoudakis, Andriopoulos, Farmakis, & Aessopos, 2001). The wider electronic exchange of clinical information between heterogeneous information systems in the delivery of diabetes care demands a common structure in the form of a message standard and close cooperation with CEN/TC251 (Vaughan et al., 2000).

Table 1. Working groups of ISO/TC215

ISO/TC215 WG 1 Health records and modeling coordination
ISO/TC215 WG 2 Messaging and communication
ISO/TC215 WG 3 Health concept representation
ISO/TC215 WG 4 Security
ISO/TC215 WG 5 Health cards
ISO/TC215 WG 6 Pharmacy and medicines business

Table 2. Standards in health and medical informatics given by ISO/TC215

ISO/IEEE 11073-10101:2004	Health informatics—Point-of-care medical device communication—Part 10101: Nomenclature
ISO/IEEE 11073-10201:2004	Health informatics—Point-of-care medical device communication—Part 10201: Domain information model
ISO/IEEE 11073-20101:2004	Health informatics—Point-of-care medical device communication—Part 20101: Application profiles—Base standard
ISO/IEEE 11073-30201:2004	Health informatics—Point-of-care medical device communication—Part 30200: Transport profile -- Cable connected
ISO/IEEE 11073-30300:2004	Health informatics—Point-of-care medical device communication—Part 30300: Transport profile -- Infrared wireless
ISO/TR 16056-1:2004	Health informatics—Interoperability of telehealth systems and networks—Part 1: Introduction and definitions
ISO/TR 16056-2:2004	Health informatics—Interoperability of telehealth systems and networks—Part 2: Real-time systems
ISO/TS 16058:2004	Health informatics—Interoperability of telelearning systems
ISO/TS 17090-1:2002	Health informatics—Public key infrastructure—Part 1: Framework and overview
ISO/TS 17090-2:2002	Health informatics—Public key infrastructure—Part 2: Certificate profile
ISO/TS 17090-3:2002	Health informatics—Public key infrastructure—Part 3: Policy management of certification authority
ISO/TS 17117:2002	Health informatics—Controlled health terminology—Structure and high-level indicators
ISO/TR 17119:2005	Health informatics—Health informatics profiling framework
ISO/TS 17120:2004	Health informatics—Country identifier standards
ISO 17432:2004	Health informatics—Messages and communication—Web access to DICOM persistent objects
ISO 18104:2003	Health informatics—Integration of a reference terminology model for nursing
ISO/TR 18307:2001	Health informatics—Interoperability and compatibility in messaging and communication standards -- Key characteristics
ISO/TS 18308:2004	Health informatics—Requirements for an electronic health record architecture
ISO 18812:2003	Health informatics—Clinical analyser interfaces to laboratory information systems -- Use profiles
ISO/TR 21089:2004	Health informatics—Trusted end-to-end information flows
ISO 21549-1:2004	Health informatics—Patient healthcard data—Part 1: General structure
ISO 21549-2:2004	Health informatics—Patient healthcard data—Part 2: Common objects
ISO 21549-3:2004	Health informatics—Patient healthcard data—Part 3: Limited clinical data
ISO/TS 21667:2004	Health informatics—Health indicators conceptual framework
ISO/TR 21730:2005	Health informatics—Use of mobile wireless communication and computing technology in healthcare facilities—Recommendations for the management of unintentional electromagnetic interference with medical devices
ISO 22857:2004	Health informatics—Guidelines on data protection to facilitate trans-border flows of personal health information

ISO/TS is a normative document representing technical consensus within an ISO committee.
ISO/TR is an informative document containing information of a different kind from that normally published in a normative document (Beolchi, 2003).
IEEE is the Institute of Electrical and Electronic Engineers, USA.

CURRENT STATUS OF STANDARDIZATION IN HEALTH AND MEDICAL INFORMATICS

All the standards developers work through their working technical committees in health and medical informatics and a number of working groups specialized in a specific area.

What are Specific Areas of Work and Results of the ISO/TC215?

Table 1 shows working groups acting in the ISO/TC215. Table 2 shows a list of standards given by this technical committee.

Table 3. CEN/TC251 working groups

CEN/TC251 WG 1 Communications: Information models, messaging, and smart cards
CEN/TC251 WG 2 Terminology
CEN/TC251 WG 3 Security, safety, and quality
CEN/TC251 WG 4 Technology for interoperability (devices)

Table 4. Standards in health and medical informatics given by CEN/TC251

EN 14484:2003 I	nternational transfer of personal health data covered by the EU data protection directive—High level security policy
EN 14485:2003 G	uidance for handling personal health data in international applications in the context of the EU data protection directive
EN 1828:2002 C	ategorial structure for classifications and coding systems of surgical procedures
EN ISO 18104:2003 I	ntegration of a reference terminology model for nursing (ISO 18104:2003)
EN ISO 18812:2003 C	linical analyser interfaces to laboratry information systems—Use profiles (ISO 18812:2003)
EN ISO 21549-1:2004	Patient healthcard data—Part 1: General structure (ISO 21549-1:2004)
EN ISO 21549-2:2004	Patient healthcard data—Part 2: Common objects (ISO 21549-2:2004)
EN ISO 21549-3:2004	Patient healthcard data—Part 3: Limited clinical data (ISO 21549-3:2004)
EN 12251:2004 S	ecure user identification for healthcare - Identification and authentication by passwords—Management and security
EN 12252:2004	Digital imaging—Communication, workflow and data management (which endorses all of DICOM as a European Standard)

What are Specific Areas of Work and Results of the CEN/TC251?

The work carried out by CEN/TC251 is mentioned in Table 3.

CEN/TC251 has been operating for 10 years. By October 2004, it had created 10 full standards or EN (Table 4).

FUTURE TRENDS IN STANDARDIZATION IN HEALTH AND MEDICAL INFORMATICS

There is no doubt that candidates for standardization in medical informatics will be core data sets for healthcare speciality groups, decision-support algorithms and clinical guidelines, and vocabulary. The identification of patients, content, and structure of electronic patient records, and messages being communicating in the healthcare system are also candidates for international standards. Standard formats need to be defined for special kinds of data like images, signals and waveforms, sound and voice, and video, including motion video. Data security, the security of objects and communication channels, and data archiving, especially in case of a catastrophe of any kind, should be standardized. Some of these standards exist or have been under way, but standardization is a continuous process, depending on the development of information and communication technology, and therefore all standards need to be improved and adapted to new technology coming day after day.

CONCLUSION

Standards in health and medical informatics are means to enabling better healthcare. Healthcare is supposed to be better if healthcare providers can access data in a patient's health record, and can read it or add some new data relevant for other healthcare providers taking care of the patient. Information technology used in diagnostics produces data about a patient as images, signals, or waves, as well as classic alphanumeric data. The format of such data should be standardized, usable, and readable on any instrument of the same kind wherever it is in the world. Medical language should be standardized, and coding systems should be universal. The transfer of medical data should be secure, as well as data storage and communication. Some patient data should be portable by health cards, especially for patients suffering from chronic diseases. Bad medical informatics can kill a patient, and standardization in the field can help make this not happen.

REFERENCES

Ball, M. J., Cortes-Comerer, N., Costin, M., Hudson, K., & Augustine, B. (2004). HCA Inc.: Standardization in action. *Journal of Healthcare Information Management, 18*(2), 59-63.

Beolchi, L. (Ed.). (2003). *European telemedicine glossary of concepts, technologies, standards and users* (5th ed.). Brussels, Belgium: European Commission.

Comité Européen de Normalisation (CEN, European Committee for Standardisation). (n.d.). Retrieved February 11, 2005, from http://www.cenorm.be/cenorm/index.htm

Deftereos, S., Lambrinoudakis, C., Andriopoulos, P., Farmakis, D., & Aessopos, A. (2001). A Java-based electronic healthcare record software for beta-thalassaemia. *Journal of Medical Internet Research, 3*(4), E33.

Hammond, W. E., & Cimino, J. J. (2001). Standards in medical informatics. In E. H. Shortliffe & L. E. Perreault (Eds.), *Medical informatics: Computer applications in health care and biomedicine* (pp. 212-255). New York: Springer.

International Organization for Standardization. (ISO). (n.d.). *Why standards matter?* Retrieved February 11, 2005, from http://www.iso.org/iso/en/aboutiso/introduction/index.html#one

Lloyd, D., & Kalra, D. (2003). EHR requirements. *Studies in Health Technology and Informatics, 96*, 231-237.

Louwerse, K. (2002). Demonstration results for the standard ENV 12924. *Studies in Health Technology and Informatics, 69*, 111-139 (pp. 229-237).

Maldonado, J. A., Crespo, P., Sanchis, A., & Robles, M. (2004). Pangaea: A mediator for the integration of distributed electronic healthcare records. In M. Fieschi, E. Coiera, & Y. C. J. Li (Eds.), *Medinfo 2004* (p. 1738). Amsterdam: IOS Press.

Marley, T. (2002). Standards supporting interoperability and EHCR communication: A CEN TC251 perspective. *Studies in Health Technology and Informatics, 87,* 72-77.

Rodrigues, J. M., Trombert Paviot, B., Martin, C., Vercherin, P., & Samuel, O. (2002). Co-ordination between clinical coding systems and pragmatic clinical terminologies based on a core open system: The role of ISO/TC215/WG3 and CEN/TC2511/WG2 standardisation? *Studies in Health Technology and Informatics, 90*, 401-405.

Vaughan, N. J., Cashman, S. J., Cavan, D. A., Gallego, M. R., Kohner, E., Benedetti, M. M., et al. (2000). A detailed examination of the clinical terms and concepts required for communication by electronic messages in diabetes care. *Diabetes, Nutrition and Metabolism, 13*(4), 201-209.

KEY TERMS

ANSI (American National Standards Institute): The private, nonprofit membership organization responsible for approving official American national standards.

CEN (Comité Européen de Normalisation, European Committee for Standardisation): The European authority for standards.

CEN/TC251: CEN technical committee for standardization in health and medical informatics.

EN: European standard made by CEN.

European Standardization: Activity of the European authority for standards.

International Standardization: Activity of the world authority for standards.

ISO (International Organization for Standardization): World authority for standards. It is also the international standard made by the organization.

ISO/TC215: ISO technical committee for standardization in health and medical informatics.

Standard: A set of rules and definitions that specify how to carry out a process or produce a product.

Standardization: Process of producing of standards.

This work was previously published in Handbook of Research on Informatics in Healthcare and Biomedicine, edited by A.A. Lazakidou, pp. 44-50, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 7.15
Biomedical Data Mining Using RBF Neural Networks

Feng Chu
Nanyang Technological University, Singapore

Lipo Wang
Nanyang Technological University, Singapore

INTRODUCTION

Accurate diagnosis of cancers is of great importance for doctors to choose a proper treatment. Furthermore, it also plays a key role in the searching for the pathology of cancers and drug discovery. Recently, this problem attracts great attention in the context of microarray technology. Here, we apply radial basis function (RBF) neural networks to this pattern recognition problem. Our experimental results in some well-known microarray data sets indicate that our method can obtain very high accuracy with a small number of genes.

BACKGROUND

Microarray is also called gene chip or DNA chip. It is a newly appeared biotechnology that allows biomedical researchers monitor thousands of genes simultaneously (Schena, Shalon, Davis, & Brown, 1995). Before the appearance of microarrays, a traditional molecular biology experiment usually works on only one gene or several genes, which makes it difficult to have a "whole picture" of an entire genome. With the help of microarrays, researchers are able to monitor, analyze and compare expression profiles of thousands of genes in one experiment.

On account of their features, microarrays have been used in various tasks such as gene discovery, disease diagnosis, and drug discovery. Since the end of the last century, cancer classification based on gene expression profiles has attracted great attention in both the biological and the engineering fields. Compared with traditional cancer diagnostic methods based mainly on the morphological appearances of tumors, the method using gene expression profiles is more objective, accurate, and reliable. More importantly, some types of cancers have subtypes with very similar appearances that are very hard to be classified by traditional methods. It has been proven that gene expression has a good capability to clarify this previously muddy problem.

Thus, to develop accurate and efficient classifiers based on gene expression becomes a problem of both theoretical and practical importance. Recent approaches on this problem include artificial neural networks (Khan *et al.*, 2001), support vector machines (Guyon, Weston, Barnhill, & Vapnik, 2002), k-nearest neighbor (Olshen & Jain, 2002), nearest shrunken centroids (Tibshirani, Hastie, Narashiman, & Chu, 2002), and so on.

A solution to this problem is to find out a group of important genes that contribute most to differentiate cancer subtypes. In the meantime, we should also provide proper algorithms that are able to make correct prediction based on the expression profiles of those genes. Such work will benefit early diagnosis of cancers. In addition, it will help doctors choose proper treatment. Furthermore, it also throws light on the relationship between the cancers and those important genes.

From the point of view of machine learning and statistical learning, cancer classification using gene expression profiles is a challenging problem. The reason lies in the following two points. First, typical gene expression data sets usually contain very few samples (from several to several tens for each type of cancers). In other words, the training data are scarce. Second, such data sets usually contain a large number of genes, for example, several thousands. That is, the data are high dimensional. Therefore, this is a special pattern recognition problem with relatively small number of patterns and very high dimensionality. To provide such a problem with a good solution, appropriate algorithms should be designed.

In fact, a number of different approaches such as k-nearest neighbor (Olshen and Jain, 2002), support vector machines (Guyon *et al.*,2002), artificial neural networks (Khan *et al.*, 2001) and some statistical methods have been applied to this problem since 1995. Among these approaches, some obtained very good results. For example, Khan *et al.* (2001) classified small round blue cell tumors (SRBCTs) with 100% accuracy by using 96 genes. Tibshirani *et al.* (2002) successfully classified SRBCTs with 100% accuracy by using only 43 genes. They also classified three different subtypes of lymphoma with 100% accuracy by using 48 genes. (Tibshirani, Hastie, Narashiman, & Chu, 2003)

However, there are still a lot of things can be done to improve present algorithms. In this work, we use and compare two gene selection schemes, i.e., principal components analysis (PCA) (Simon, 1999) and a t-test-based method (Tusher, Tibshirani, & Chu, 2001). After that, we introduce an RBF neural network (Fu & Wang, 2003) as the classification algorithm.

MAIN THRUST

After a comparative study of gene selection methods, a detailed description of the RBF neural network and some experimental results are presented in this section.

Microarray Data Sets

We analyze three well-known gene expression data sets, i.e., the SRBCT data set (Khan *et al.*, 2001), the lymphoma data set (Alizadeh *et al.*, 2000), and the leukemia data set (Golub *et al.*, 1999).

The lymphoma data set (http://llmpp.nih.gov/lymphoma) (Alizadeh *et al.*, 2000) contains 4026 "well measured" clones belonging to 62 samples. These samples belong to following types of lymphoid malignancies: diffuse large B-cell lymphoma (DLBCL, 42 samples), follicular lymphoma (FL, nine samples) and chronicle lymphocytic leukemia (CLL, 11 samples). In this data set, a small part of data is missing. A k-nearest neighbor algorithm was used to fill those missing values (Troyanskaya *et al.*, 2001).

The SRBCT data set (http://research.nhgri.nih.gov/microarray/Supplement/) (Khan *et al.*, 2001) contains the expression data of 2308 genes. There are totally 63 training samples and 25 testing samples. Five of the testing samples are not SRBCTs. The 63 training samples contain 23 Ewing family of tumors (EWS), 20 rhabdomyosarcoma (RMS), 12 neuroblastoma (NB), and eight Burkitt lymphomas (BL). And the 20 testing samples contain six EWS, five RMS, six NB, and three BL.

The leukemia data set (http://www-genome.wi.mit.edu/cgi-\\bin /cancer/publications) (Golub *et al.*, 1999) has two types of leukemia, i.e., acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL). Among these samples, 38 of them are for training; the other 34 blind samples are for testing. The entire leukemia data set contains the expression data of 7,129 genes. Different with the cDNA microarray data, the leukemia data are oligonucleotide microarray data. Because such expression data are raw data, we need to normalize them to reduce the systemic bias induced during experiments. We follow the normalization procedure used by Dudoit, Fridlyand, and Speed (2002). Three preprocessing steps were applied: (a) thresholding with floor of 100 and ceiling of 16000; (b) filtering, exclusion of genes with $max/min<5$ or $(max-min)<500$. *max* and *min* refer to the maximum and the minimum of the gene expression values, respectively; and (c) base 10 logarithmic transformation. There are 3571 genes survived after these three steps. After that, the data were standardized across experiments, i.e., minus the mean and divided by the standard deviation of each experiment.

Methods for Gene Selection

As mentioned in the former part, the gene expression data are very high-dimensional. The dimension of input patterns is determined by the number of genes used. In a typical microarray experiment, usually several thousands of genes take part in. Therefore, the dimension of patterns is several thousands. However, only a small number of the genes contribute to correct classification; some others even act as "noise". Gene selection can eliminate the influence of such "noise". Furthermore, the fewer the genes used, the lower the computational burden to the classifier. Finally, once a smaller subset of genes is identified as relevant to a particular cancer, it helps biomedical researchers focus on these genes that contribute to the development of the cancer. The process of gene selection is ranking genes' discriminative ability first and then retaining the genes with high ranks.

As a critical step for classification, gene selection has been studied intensively in recent years. There are two main approaches, one is principal component analysis (PCA) (Simon, 1999), perhaps the most widely used method; the other is a t-test-based approach which has been more and more widely accepted. In the important papers (Alizadeh *et al.*, 2000; Khan *et al.*, 2001), PCA was used. The basic idea of PCA is to find the most "informative" genes that contain most of the information in the data set. Another approach is based on t-test that is able to measure the difference between two groups. Thomas, Olsen, Tapscott, and Zhao. (2001) recommended this method. Tusher *et al.* (2001) and Pan (2002) also proposed their method based on t-test, respectively. Besides these two main methods, there are also some other methods. For example, a method called Markov blanket was proposed by Xing, Jordan, and Karp

(2001). Li, Weinberg, Darden, and Pedersen (2001) applied another method which combined genetic algorithm and K-nearest neighbor.

PCA (Simon, 1999) aims at reducing the input dimension by transforming the input space into a new space described by principal components (PCs). All the PCs are orthogonal and they are ordered according to the absolute value of their eigenvalues. The k-th PC is the vector with the k-th largest eigenvalue. By leaving out the vectors with small eigenvalues, the input space's dimension is reduced.

In fact, the PCs indicate the directions with largest variations of input vectors. Because PCA chooses vectors with largest eigenvalues, it covers directions with largest variations of vectors. In the directions determined by the vectors with small eigenvalues, the variations of vectors are very small. In a word, PCA intends to capture the most informative directions (Simon, 1999).

We tested PCA in the lymphoma data set (Alizadeh *et al*., 2000). We obtained 62 PCs from the 4026 genes in the data set by using PCA. Then, we ranked those PCs according to their eigenvalues (absolute values). Finally, we used our RBF neural network that will be introduced in the latter part to classify the lymphoma data set.

At first, we randomly divided the 62 samples into two parts, 31 samples for training and the other 31 samples for testing. We then input the 62 PCs one by one to the RBF network according to their eigenvalue ranks starting with the PC ranked one. That is, we first used only a single PC that is ranked 1 as the input to the RBF network. We trained the network with the training data and subsequently tested the network with the testing data. We repeated this process with the top two PCs, then the top three PCs, and so on. Figure 1 shows the testing error. From this result, we found that the RBF network can not reach 100% accuracy. The best testing accuracy is 93.55% that happened when 36 or 61 PCs were input to the classifier. The classification result using the t-test-based gene selection method will be shown in the next section, which is much better than PCA approach.

The t-test-based gene selection measures the difference of genes' distribution using a t-test based scoring scheme, i.e., t-score (TS). After that, only the genes with the highest TSs are to be put into our classifier. The TS of gene *i* is defined as follows (Tusher *et al*., 2001):

$$TS_i = \max\left\{ \left| \frac{\bar{x}_{ik} - \bar{x}_i}{d_k s_i} \right|, k = 1, 2, \ldots K \right\}$$

where:

$$\bar{x}_{ik} = \sum_{j \in C_k} \bar{x}_{ij} / n_k$$

$$\bar{x}_i = \sum_{j=1}^{n} x_{ij} / n$$

$$s_i^2 = \frac{1}{n-K} \sum_k \sum_{j \in C_k} \left(x_{ij} - \bar{x}_{ik} \right)^2$$

$$d_k = \sqrt{1/n_k + 1/n}$$

There are K classes. max $\{y_k, k = 1,2,..K\}$ is the maximum of all y_k, $k = 1,2,..K$. C_k refers to class k that includes n_k samples. x_{ij} is the expression value of gene i in sample j. is the mean expression value in class k for gene i. n is the total number of samples. $\bar{x}_i$ is the general mean expression value for gene i. s_i is the pooled within-class standard deviation for gene i. Actually, the t-score used here is a t-statistics between a specific class and the overall centroid of all the classes.

To compare the t-test-based method with PCA, we also applied it to the lymphoma data set with the same procedure as what we did by using PCA. This method obtained 100% accuracy with only the top six genes. The results are shown in Figure 1. This comparison indicated that the t-test-based method was much better than PCA in this problem.

An RBF Neural Network

An RBF neural network (Haykin, 1999) has three layers. The first layer is an input layer; the second

Figure 1. Classification results of using PCA and the t-test-based method as gene selection methods

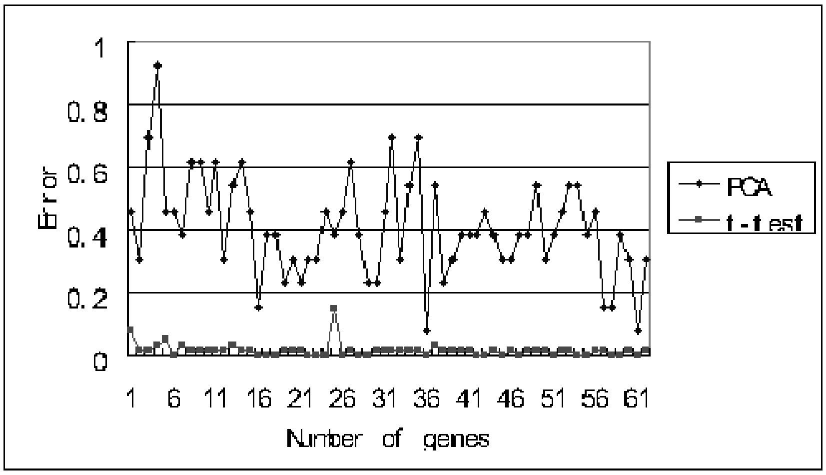
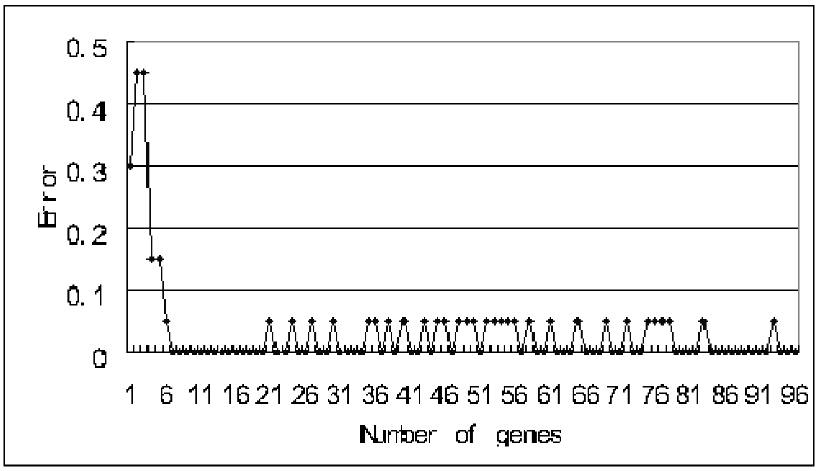


Figure 2. The testing result in the SRBCT data set



layer is a hidden layer that includes some radial basis functions, also known as hidden kernels; the third layer is an output layer. An RBF neural network can be regarded as a mapping of the input domain X onto the output domain Y. Mathematically, an RBF neural network can be described as follows:

$$y_m(x) = \sum_{i=1}^{N} w_{mi} G(\|x - t_i\|) + b_m,$$

$i=1,2,\ldots,N;\ m=1,2,\ldots M$

Here $\|\cdot\|$ stands for the Euclidean norm. M is the number of outputs. N is the number of hidden kernels. $y_m(x)$ is the output m corresponding to the input *x*. t_i is the position of kernel *i*. w_{mi} is the weight between the kernel *i* and the output *m*. b_m is the bias on the output *m*. b_m is the kernel function. Usually, an RBF neural network uses Gaussian kernel functions as follows:

$$G(\|x-t_i\|) = \exp(-\frac{\|x-t_i\|^2}{2\sigma_i^2})$$

where σ_i is the radius of the kernel *i*.

The main steps to construct an RBF neural network include: (a) determining the positions of all the kernels (t_i); (b) determining the radius of each kernel (); and (c) calculating the weights between each kernel and each output node.

In this paper, we use a novel RBF neural network proposed by Fu and Wang (Fu and Wang, 2003), which allows for large overlaps of hidden kernels belonging to the same class.

Results

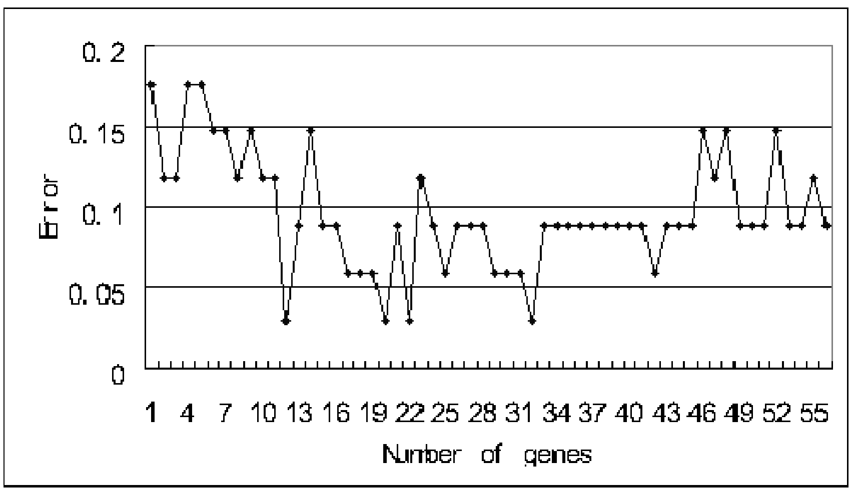
In the SRBCT data set, we first ranked the entire 2308 genes according to their TSs (Tusher *et al.*, 2001). Then we picked out 96 genes with the highest TSs. We applied our RBF neural network to classify the SRBCT data set. The SRBCT data set contains 63 samples for training and 20 blind samples for testing. We input the selected 96 genes one by one to the RBF network according to their TS ranks starting with the gene ranked one. We repeated this process with the top two genes, then the top three genes, and so on. Figure 2 shows the testing errors with respect to the number of genes. The testing error decreased to 0 when the top seven genes were input into the RBF network.

In the leukemia data set, we chose 56 genes with the highest TSs (Tusher *et al.*, 2001). We followed the same procedure as in the SRBCT data set. We did classification with 1 gene, then two genes, then three genes and so on. Our RBF neural network got an accuracy of 97.06%, i.e. one error in all 34 samples, when 12, 20, 22, 32 genes were input, respectively.

FUTURE TRENDS

Until now, the focus of work is investigating the information with statistical importance in microarray data sets. In the near future, we will try to incorporate more biological knowledge into our algorithm, especially the correlations of genes.

Figure 3. The testing result in the leukemia data set



In addition, with more and more microarray data sets produced in laboratories around the world, we will try to mine multi-data-set with our RBF neural network, i.e., we will try to process the combined data sets. Such an attempt will hopefully bring us a much broader and deeper insight into those data sets.

CONCLUSION

Through our experiments, we conclude that the t-test-based gene selection method is an appropriate feature selection/dimension reduction approach, which can find more important genes than PCA can.

The results in the SRBCT data set and the leukemia data set proved the effectiveness of our RBF neural network. In the SRBCT data set, it obtained 100% accuracy with only seven genes. In the leukemia data set, it made only one error with 12, 20, 22, and 32 genes, respectively. In view of this, we also conclude that our RBF neural network outperforms almost all the previously published methods in terms of accuracy and the number of genes required.

REFERENCES

Alizadeh, A.A. et al. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, *403*, 503-511.

Dudoit, S., Fridlyand, J., & Speed, J. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of American Statistics Association*, *97*, 77-87.

Fu, X., & Wang, L. (2003). Data dimensionality reduction with application to simplifying RBF neural network structure and improving classification performance. *IEEE Trans. Syst., Man, Cybernetics. Part B: Cybernetics*, *33*, 399-409.

Golub, T.R. et al. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, *286*, 531-537.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*, 389-422.

Haykin, S. (1999). *Neural network, a comprehensive foundation* (2nd ed.). New Jersey, U.S.A: Prentice-Hall, Inc.

Khan, J.M. et al. (2001). Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature Medicine*, *7*, 673-679.

Li, L., Weinberg, C.R., Darden, T.A., & Pedersen, L.G. (2001). Gene selection for sample classification based on gene expression data: Study of sensitivity to choice of parameters of the GA/KNN method. *Bioinformatics*, *17*, 1131-1142.

Olshen, A.B., & Jain, A.N. (2002). Deriving quantitative conclusions from microarray expression data. *Bioinformatics*, *18*, 961-970.

Pan, W. (2002). A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. *Bioinformatics*, *18*, 546-554.

Schena, M., Shalon, D., Davis, R.W., & Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, *270*, 467-470.

Thomas, J.G., Olsen, J.M., Tapscott, S.J. & Zhao, L.P. (2001). An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. *Genome Research*, *11*, 1227-1236.

Tibshirani, R., Hastie, T., Narashiman, B., & Chu, G. (2002). Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc. Natl. Acad. Sci. USA*, *99*, 6567-6572.

Tibshirani, R., Hastie, T., Narashiman, B., & Chu, G. (2003). Class predication by nearest shrunken centroids with applications to DNA microarrays. *Statistical Science, 18,* 104-117.

Troyanskaya, O., Cantor, M, & Sherlock, G., *et al.* (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics, 17,* 520-525.

Tusher, V.G., Tibshirani, R., & Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA, 98,* 5116-5121.

Xing, E.P., Jordan, M.I., & Karp, R.M. (2001). Feature selection for high-dimensional genomic microarray data. *Proceedings of the Eighteenth International Conference on Machine Learning* (pp. 601-608). Morgan Kaufmann Publishers, Inc.

KEY TERMS

Feature Extraction: Feature extraction is the process to obtain a group of features with the characters we need from the original data set. It usually uses a transform (e.g. principal component analysis) to obtain a group of features at one time of computation.

Feature Selection: Feature selection is the process to select some features we need from all the original features. It usually measures the character (e.g. t-test score) of each feature first, then, chooses some features we need.

Gene Expression Profile: Through microarray chips, an image that describes to what extent genes are expressed can be obtained. It usually uses red to indicate the high expression level and uses green to indicate the low expression level. This image is also called a gene expression profile.

Microarray: A Microarray is also called a gene chip or a DNA chip. It is a newly appeared biotechnology that allows biomedical researchers monitor thousands of genes simultaneously.

Principal Components Analysis: Principal components analysis transforms one vector space into a new space described by principal components (PCs). All the PCs are orthogonal to each other and they are ordered according to the absolute value of their eigenvalues. By leaving out the vectors with small eigenvalues, the dimension of the original vector space is reduced.

Radial Basis Function (RBF) Neural Network: An RBF neural network is a kind of artificial neural network. It usually has three layers, i.e., an input layer, a hidden layer, and an output layer. The hidden layer of an RBF neural network contains some radial basis functions, such as Gaussian functions or polynomial functions, to transform input vector space into a new non-linear space. An RBF neural network has the universal approximation ability, i.e., it can approximate any function to any accuracy, as long as there are enough hidden neurons.

T-Test: T-test is a kind of statistical method that measures how large the difference is between two groups of samples.

Testing a Neural Network: To know whether a trained neural network is the mapping or the regression we need, we test this network with some data that have not been used in the training process. This procedure is called testing a neural network.

Training a Neural Network: Training a neural network means using some known data to build the structure and tune the parameters of this network. The goal of training is to make the network represent a mapping or a regression we need.

This work was previously published in Encyclopedia of Data Warehousing and Mining, edited by J. Wang, pp. 106-111, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 7.16
Mining BioLiterature:
Toward Automatic Annotation of Genes and Proteins

Francisco M. Couto
Universidade de Lisboa, Portugal

Mário J. Silva
Universidade de Lisboa, Portugal

ABSTRACT

This chapter introduces the use of Text Mining in scientific literature for biological research, with a special focus on automatic gene and protein annotation. This field became recently a major topic in Bioinformatics, motivated by the opportunity brought by tapping the BioLiterature with automatic text processing software. The chapter describes the main approaches adopted and analyzes systems that have been developed for automatically annotating genes or proteins. To illustrate how text-mining tools fit in biological databases curation processes, the chapter presents a tool that assists protein annotation. Besides the promising advances of Text Mining of BioLiterature, many problems need to be addressed. This chapter presents the main open problems in using text-mining tools for automatic annotation of genes and proteins, and discusses how a more efficient integration of existing domain knowledge can improve the performance of these tools.

INTRODUCTION

Bioinformatics aims at understanding living systems using biological information. The facts discovered in biological research have been mainly published in the scientific literature (BioLiterature) since the 19th century. Extracting knowledge from such a large amount of unstructured information is a painful and hard task, even to an expert. A solution could be the creation of a database where authors would deposit all the facts published in BioLiterature in a structured form. Some generic databases, such as UniProt, collect and distribute biological information (Apweiler et al., 2004). However, different communities have different needs and views on specific topics, which change over time. As a result, researchers do not look only for the facts, but also for their evidence. Before a researcher considers a fact as relevant to his work, he checks the evidence presented by the author, because facts are normally valid only in a specific context. This explains why Molecular Biology knowledge continues to be mainly published in BioLiterature. Another solution is Text Mining, which aims at automatically extracting knowledge from natural language texts (Hearst, 1999). Text-mining systems can be used to identify the following types of information: entities, such as genes, proteins and cellular components; relationships, such as protein localization or protein interactions; and events, such as experimental methods used to discover protein interactions. Bioinformatics tools to collect more information about the concepts they analyze also use Text Mining. For example, information automatically extracted from the BioLiterature can improve gene expression clustering (Blaschke, Oliveros, & Valencia, 2004).

Text Mining of BioLiterature has been studied since the last decade (Andrade & Valencia, 1998). The interest in the topic has been steadily increasing, motivated by the vast amount of publications that curators have to read in order to update biological databases, or simply to help researchers keep up with progress in a specific area. Text Mining can minimize these problems mainly because BioLiterature articles are quite often publicly available. The most widely used BioLiterature repository is MEDLINE, which provides a vast collection of abstracts and bibliographic information. For example, in 2003, about 560,000 citations have been added to MEDLINE. Reading 10 of these documents per day, it would take around 150 years to read all the documents added in 2003. Moreover, the number of new documents added per year increased by more than 20,000 from 2000 to 2003. Hence, text-mining systems could have a great impact in minimizing this effort by automatically extracting information that can be used for multiple purposes and could not possibly be organized by other means.

This chapter starts by providing broad definitions used in Text Mining and describes the main approaches. Then, it summarizes the state-of-the-art of this field and shows how text-mining systems can be used to automatically annotate genes or proteins. Next, the chapter describes a tool designed for assisting protein annotation. Finally, the chapter discusses future and emerging trends and presents concluding remarks.

TEXT MINING

Text Mining aims at automatically extracting knowledge from unstructured text. Usually the text is organized as a collection of documents, or corpus.

TextMining = NLP + DataMining

Data Mining aims at automatically extracting knowledge from structured data. (Hand, Mannila, & Smyth, 2000). Thus, Text Mining is a special case of Data Mining, where input data is text instead of structured data. Normally, text-mining systems generate structured representations of the text, which are then analyzed by Data Mining tools. The simplest representation of a text is a vector with the number of occurrences of each word in the text (called a bag-of-words). This representation can be easily created and manipulated, but ignores all the text structure. Text-mining systems may also use Natural Language Processing (NLP) techniques to represent and process text more effectively. NLP is a broad research area that aims at analyzing spoken, handwritten, printed, and electronic text for different purposes, such as speech recognition or translation (Manning & Schütze, 1999). The most popular NLP techniques used by text-mining systems include: tokenization, morphology analysis, part-of-speech tagging, sense disambiguation, parsing, and anaphora resolution.

Tokenization aims at identifying boundaries in the text to fragment it into basic units called tokens. The first step in a text-mining system is to identify the tokens. The most commonly used token is the word. In most languages, white-space characters can be considered an accurate boundary to fragment the text into words. This problem is more complex in languages without explicitly delimiters, such as Chinese (Wu & Fung, 1994). Morphology analysis aims at grouping the words (tokens) that are variants of a common word, and therefore are normally used with a similar meaning (Spencer, 1991). This involves the study of the structure and formation of words. A common type of inflectional variants results from the tense on verbs. For example, "binding" and "binds" are inflectional variants of "bind." Some other word variants result from prefixing, suffixing, infixing, or compounding.

Part-of-speech tagging aims at labeling each word with its semantic role, such as article, noun, verb, adjective, preposition, or pronoun (Baker, 1989). This involves the study of the structure and formation of sentences. The tagging is a classification of words according to their semantic role and to their relations to each other in a sentence. Sense disambiguation selects the correct meaning of a word in a given piece of text. For example, "compound" has two different senses in the expressions "compound the ingredients" and "chemical compound." Normally, the part-of-speech tags are used as a first step in sense disambiguation (Wilks & Stevenson, 1997)

Parsing aims at identifying the syntactic structure of a sentence (Earley, 1970). The syntactic structure of a sequence of words is composed by a set of other syntactic structures related to smaller sequences, except for the part-of-speech tags that are syntactic structures directly linked to words. Normally, the syntactic structure of a sentence is represented by a syntax tree, where leafs represent the words and internal nodes the syntactic structures. Algorithms to identify the complete syntactic structure of a sentence are in general inaccurate and time-consuming, given the combinatorial explosion in long sentences. An alternative is shallow-parsing, which does not attempt to parse complex syntactic structures. Shallow-parsing only splits sentences into phrases, in other words, subsequences of words that represent a grammatical unit, such as noun phrase or verb phrase. Anaphora (or co-reference) resolution aims at determining different sequences of words referring to the same entity. For example, in the sentence "The enzyme has an intense activity, thus, this protein should be used", the noun phrases "The enzyme" and "this protein" refer to same entity.

Some of the NLP techniques described above can be implemented using algorithms also used in Data Mining. For example, part-of-speech taggers can use Hidden Markov Models (HMMs) to estimate the probability of a sequence of part of speech assignments (Smith, Rindflesch, & Wilbur, 2004). Not all NLP techniques improve the performance of a given text-mining system. As a result, designers of text-mining systems have to select which NLP techniques would be useful to achieve their ultimate goal.

After creating a structured representation of texts, text-mining systems can use the following approaches for extracting knowledge (Leake, 1996):

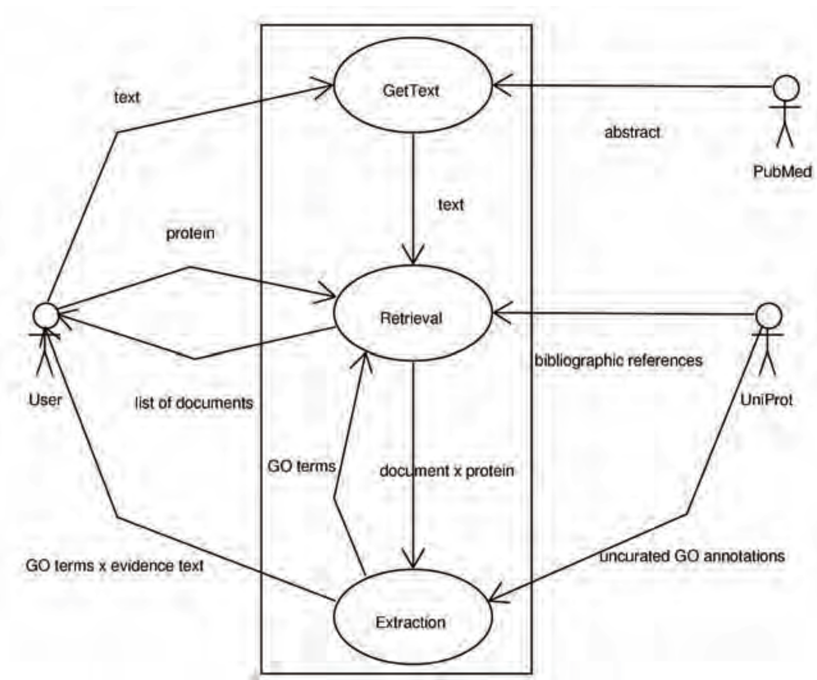
rule-based or case-based

The rule-based approach relies on rules inferred from patterns identified from the text by an expert. The rules represent in a structured form the knowledge acquired by experts when performing the same task. The expert analyzes a subpart of the text and identifies common patterns in which the relevant information is expressed. These patterns are then converted to rules to identify the relevant information in the rest of the text. The main bottleneck of this approach is the manual process of creating rules and patterns. Besides being time-consuming, in most cases, this manual process is unable to devise from a

Table 1. Categorization of some recent text-mining systems designed for automatic annotation of genes and proteins; for each system, the table indicates the mining approach taken, the proportion of NLP techniques used and the proportion of manual intervention needed to generate rules, patterns, or training sets

System	Mining	NLP	Manual
Andrade and Valencia (1998)	Rule-based	nil	nil
Pérez et al. (2004)	Rule-based	nil	nil
Corney et al. (2004)	Rule-based	Low	High
Müller et al. (2004)	Rule-based	Low	Medium
Kim and Park (2004)	Rule-based	Medium	Medium
Koike et al. (2005)	Rule-based	High	Medium
Palakal et al. (2003)	Case-based	Medium	Low
Chiang and Yu (2003)	Case-based	Medium	Low

Figure 1. UML use case diagram of GOAnnotator



subpart of the text the set of rules that encompass all possible cases.

The case-based approach relies on a predefined set of texts previously annotated by an expert, which is used to learn a model for the rest of the text. Cases contain knowledge in an unprocessed form, and they only describe the output expected by the users for a limited set of examples. The expert analyzes a subpart of the text (training set) and provides the output expected to be returned by the text-mining system for that text. The system uses the training set to create a probabilistic model that will be applied to the rest of the text. The main bottleneck of this approach is the selection and creation of a training set large enough to enable the creation of a model accurate for the rest of the text.

The manual analysis of text requires less expertise in the case-based approach than in the rule-based approach. In the rule-based approach, the expert has to identify how the relevant information is expressed in addition to the expected output. However, rule-based systems can use this expertise to achieve high precision by selecting the most reliable rules and patterns.

STATE OF THE ART

The main problem in BioLiterature mining is coping with the lack of a standard nomenclature for describing biologic concepts and entities. In BioLiterature, we can often find different terms referring to the same biological concept or entity (synonyms), or the same term meaning different biological concepts or entities (homonyms). Genes, whose name is a common English word, are frequent, which makes it difficult to recognize biological entities in the text.

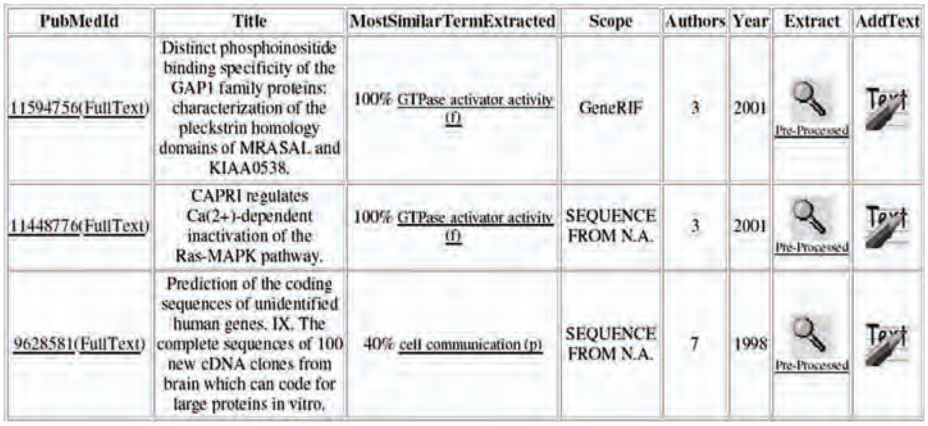
Recent advances in Text Mining of BioLiterature already achieved acceptable levels of accuracy in identifying gene and protein names in the text. However, the extraction of relationships, such as functional annotations, is still far from being solved. Recent surveys report these advances by presenting text-mining tools that are not only run in different corpora but also perform different tasks (Hirschman, Park, Tsujii, Wong, & Wu, 2002; Blaschke, Hirschman, & Valencia 2002; Dickman, 2003; Shatkay & Feldman, 2003).

On the other hand, recent challenging evaluations compared the performance of different approaches in solving the same tasks using the same corpus. For example, the 2002 KDD Cup (bio-text task) consisted on identifying which biomedical articles contained relevant experimental results about Drosophila (fruit fly), and the genes (transcripts and proteins) involved (Yeh, Hirschman, & Morgan, 2003). The best submission out of 32 obtained 78% F-measure in the article decision, and 67% F-measure in the gene decision.

A similar challenging evaluation was the 2004 TREC genomics track, which consisted on identifying relevant documents and documents with relevant experimental results about the mouse (Hersh et al., 2004). The first task was a typical Information Retrieval task. A list of documents and a list of topics were given as input. The goal was to identify the relevant documents for each topic. The best submission out of 47 obtained 41% precision. The second subtask comprised the selection of documents with relevant experimental information. The best submission out of 59 obtained 27% F-measure. In addition to document selection, the task also comprised automatic annotations of genes. The best submission out of 36 obtained 56% F-measure.

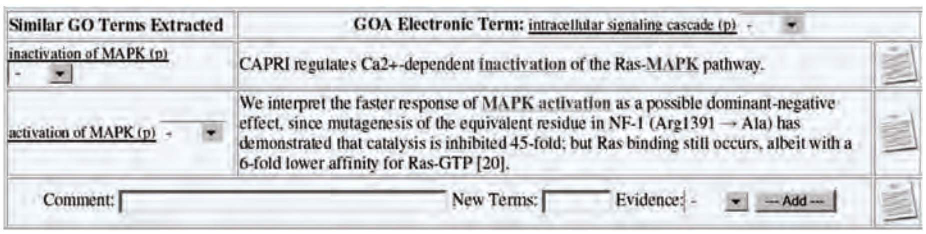
Another challenging evaluation was BioCreAtIvE (Hirschman, Yeh, Blaschke, & Valencia, 2005). This evaluation comprised two tasks. The first aimed at identifying genes and proteins in BioLiterature. The best submission out of 40 obtained 83% F-measure. The second task addressed the automatic annotation of human proteins, and involved two subtasks. The first subtask required the identification of the texts that provided the evidence for extracting each annotation. From 21 submissions, the highest precision was 78% and

Figure 2. A list of documents related to the protein "Ras GTPase-activating protein 4" provided by the GOAnnotator. The list is sorted by the similarity of the most similar term extracted from each document. The curator can invoke the links in the "Extract" column to see the extracted terms together with the evidence text. By default, GOAnnotator uses only the abstracts of scientific documents, but the curator can replace or add text (links in the "AddText" column).



PubMedId	Title	MostSimilarTermExtracted	Scope	Authors	Year	Extract	AddText
11594756(FullText)	Distinct phosphoinositide binding specificity of the GAP1 family proteins: characterization of the pleckstrin homology domains of MRASAL and KIAA0538.	100% GTPase activator activity (f)	GeneRIF	3	2001	Pre-Processed	Text
11448776(FullText)	CAPRI regulates Ca(2+)-dependent inactivation of the Ras-MAPK pathway.	100% GTPase activator activity (f)	SEQUENCE FROM N.A.	3	2001	Pre-Processed	Text
9628581(FullText)	Prediction of the coding sequences of unidentified human genes. IX. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro.	40% cell communication (p)	SEQUENCE FROM N.A.	7	1998	Pre-Processed	Text

Figure 3. For each uncurated annotation, GOAnnotator shows the similar GO terms extracted from a sentence of the selected document. If any of the sentences provides correct evidence for the uncurated annotation, or if the evidence supports a GO term similar to that present in the uncurated annotation, the curator can use the "Add" option to store the annotation together with the document reference, the evidence codes and additional comments.



Similar GO Terms Extracted	GOA Electronic Term: intracellular signaling cascade (p) -
inactivation of MAPK (p) -	CAPRI regulates Ca2+-dependent inactivation of the Ras-MAPK pathway.
activation of MAPK (p) -	We interpret the faster response of MAPK activation as a possible dominant-negative effect, since mutagenesis of the equivalent residue in NF-1 (Arg1391 → Ala) has demonstrated that catalysis is inhibited 45-fold; but Ras binding still occurs, albeit with a 6-fold lower affinity for Ras-GTP [20].
Comment:	New Terms: Evidence: - --- Add ---

the highest recall was 23%. The second subtask consisted on automatic annotation of proteins. From 18 submissions, the highest precision was 34% and the highest recall was 12%.

AUTOMATIC ANNOTATION

One of the most important applications of text-mining systems is the automatic annotation of genes and proteins. A gene or protein annotation consists of a pair composed by the gene or protein and a description of its biological role. Normally, descriptions use terms from a common ontology. The Gene Ontology (GO-Consortium, 2004) provides a structured controlled vocabulary that can be applied to different species (GO-Consortium, 2004). GO has three different aspects: molecular function, biological process, and cellular component. To comprehend a gene or protein activity is also important to know the biological entities that interact with it. Thus, the annotation of a gene or protein also involves identifying interacting chemical substances, drugs, genes and proteins.

Text-mining systems that automatically annotate genes or proteins can be categorized according to: the mining approach taken (rule-based or case- based), the NLP techniques applied, and the amount of manual intervention required.

Rule-Based Systems

AbXtract was one of the first text-mining systems attempting to characterize the function of genes and proteins based on information automatically extracted from BioLiterature (Andrade & Valencia, 1998). The system assigns relevant keywords to protein families based on a rule comprising the frequency of the keywords in the abstracts related to the family. In addition to using a rule-based approach, AbXtract relies in only one rule that does not require human intervention. A similar approach is taken by the system proposed by Pérez, Perez-Iratxeta, Bork, Thode, and Andrade (2004), which annotates genes with keywords extracted from abstracts based on mappings between different ontologies.

An example of a system based on a large number of rules is BioRAT (Corney, Buxton, Langdon, & Jones, 2004). Given a query, BioRAT finds documents and highlights the most relevant facts in their abstracts or full texts. However, the rules are exclusively derived from patterns inserted by the user. Textpresso is another rule-based system that finds documents and marks them up with terms from a built-in ontology (Müller, Kenny, & Sternberg, 2004). The system assigns to each entry of the ontology regular expressions that capture how the entry can be expressed in BioLiterature. Textpresso is not so dependent on the user as BioRAT, since many of the regular expressions are automatically generated to account for regular forms of verbs and nouns.

BioIE is a system that takes more advantage of NLP techniques. It extracts biological interactions from BioLiterature and annotates them with GO terms (Kim & Park, 2004). The system uses morphology, sense disambiguation, and rules with syntactic dependencies to identify GO terms in the text. BioIE uses 1,312 patterns to match interactions in the sentences, so it also requires substantial manual intervention. Koike, Niwa, and Takagi (2005) propose a similar system that annotates gene, protein, and families with GO terms extracted from texts. The system uses morphology, part-of-speech tagging, shallow parsing, and simple anaphora resolution. To extract the relationships, it uses both automatically generated and manually inserted rules.

Case-Based Systems

A text-mining system using the case-based approach was proposed by Palakal, Stephens, Mukhopadhyay, and Raje (2003). The system extracts relationships between biological objects (e.g. protein, gene, cell cycle). The system uses sense disambiguation, and a probabilistic model to find directional relationships. The model is trained using examples of sentences expressing a relationship.

MeKE is another system that extracts protein functions from BioLiterature using sentence alignment (Chiang & Yu, 2003). MeKE also uses sense disambiguation. The system uses a statistical classifier that identifies common patterns in examples of sentences expressing GO annotations. The classifier uses these patterns to decide if a given sentence expresses a GO annotation.

Discussion

The systems described above show how Text Mining can help curators in the annotation process. Most rely on domain knowledge manually inserted by curators (see Table 1). Domain knowledge improves precision, but it cannot be easily extended to work on other domains and demands an extra effort to keep the knowledge updated as BioLiterature evolves. This approach is time-consuming and makes the systems too specific to be extended to new domains. Thus, an approach to avoid this process is much needed.

GOAnnotator

This section illustrates how text-mining can be integrated in a biological database curation process, by describing GOAnnotator, a tool for assisting the GO annotation of UniProt entries (Rebholz-Schuhmann, Kirsch, & Couto, 2005). GOAnnotator links the GO terms present in the uncurated annotations with evidence text automatically extracted from the documents linked to UniProt entries.

Figure 1 presents the data flow involved in the processing steps of GOAnnotator and in its interaction with the users and external sources. Initially, the curator provides a UniProt accession number to GOAnnotator. GOAnnotator follows the bibliographic links found in the UniProt database and retrieves the documents. Additional documents are retrieved from the GeneRIF database (Mitchell et al., 2003). Curators can also provide any other text for mining. GOAnnotator then extracts from the documents GO terms similar to the GO terms present in the uncurated annotations.

In GOAnnotator the extraction of GO terms is performed by FiGO, a tool that receives text and returns the GO terms detected (Couto, Silva, & Coutinho, 2005). FiGO is rule-based, does not use any NLP technique and does not require manual intervention. FiGO assigns a confidence value to each GO term that represents the terms' likelihood of being mentioned in the text. The confidence value is the product of two parameters. The first, called local evidence context (LEC), is used to measure the likelihood that words in the text are part of a given GO term. The second parameter is the inverse of their frequency in GO. GO terms are similar if they are in the same lineage or if they share a common parent in the GO hierarchy. FiGO uses the semantic similarity measure of (Lin, 1998) to compute the degree of similarity between two GO terms.

GOAnnotator ranks the documents based on the extracted GO terms from the text and their similarity to the GO terms present in the uncurated annotations (see Figure 2). Any extracted GO term is an indication for the topic of the document, which is also taken from the UniProt entry.

GOAnnotator displays a table for each uncurated annotation with the GO terms that were extracted from a document and found similar to the GO term present in the uncurated annotation (see Figure 3). The sentences from which the GO terms were extracted are also displayed. Words that have contributed to the extraction of the GO terms are highlighted. GOAnnotator gives the curators the opportunity to manipulate the confidence and similarity thresholds to modify the number of predictions.

FUTURE TRENDS

The performance of text-mining tools that automatically annotate genes or proteins is still not acceptable by curators. Gene or protein annotation is more subjective and requires more expertise than simply finding relevant documents and recognizing biological entities in texts. Moreover, an annotation tool can only perform well when it is using the correct documents and the correct entities. Errors in the retrieval of documents or in the recognition of entities will be the cause of errors in the annotation task.

Existing tools that retrieve relevant documents do not always provide what the curators want. On the contrary, curators spend a large amount of their time finding the right documents. This is probably the main reason why many curators are still not using text-mining tools for gene or protein annotation. Another reason is that quite often the full texts are not electronically available. Curators need additional information that is not usually present in the abstracts, such as the type of experiments applied and the species from which proteins originate. Finally, another reason is that most text-mining tools depend on domain knowledge manually inserted by curators, which is also very time-consuming.

Text-mining tools acquire domain knowledge from the curators in the form of rules or cases. The identification of rules requires more effort to the curators than the evaluation of a limited set of cases. However, a single rule can express knowledge not contained in a large set of cases. Neither source of knowledge subsumes the other: the knowledge represented by a rule is normally not well-represented by any set of cases, and it is difficult to identify a set of rules representing all knowledge expressed by a set of cases.

Couto, Martins, and Silva (2004) proposed an approach to obtain the domain knowledge that does not require human intervention. Instead of obtaining the domain knowledge from curators, they propose acquiring it from publicly available databases that already contain curated data. Text-mining systems could consider these databases as training sets from which rules, patterns, or models can be automatically generated. Besides avoiding direct human intervention, these automated training sets are usually much larger than individually generated training sets. Another advantage is that the tools' training data does not become outdated as public databases can be tracked for updates as they evolve.

SUMMARY

Bioinformatics aims at understanding living systems by inferring knowledge from biological information, such as DNA and protein sequences. The role of Text Mining in Bioinformatics is to automatically extract knowledge from BioLiterature. This field is new and has evolved over the last decade, motivated by the opportunity brought by tapping the large amount of information that has been published in BioLiterature with automatic text processing software.

Researchers will tend to use databases to store and find facts, but the evidence substantiating them will continue to be described as unstructured text. As a result, text-mining tools will continue to have an important role in Bioinformatics. Recent advances in Text Mining of BioLiterature are already promising, but many problems remain. In our opinion, the future of text-mining tools for gene or protein annotation will mainly depend on a better use of NLP techniques, and in an efficient integration of existing domain knowledge available in biological databases and ontologies.

REFERENCES

Andrade, M., & Valencia, A. (1998). Automatic extraction of keywords from scientific text: Application to the knowledge domain of protein families. *Bioinformatics, 14*(7), 600-607.

Apweiler, R., Bairoch, A., Wu, C., Barker, W., Boeckmann, B., Ferro, S., et al. (2004). UniProt: The universal protein knowledgebase. *Nucleic Acids Research, 32* (Database issue), D115-D119.

Baker, C. (1989). *English syntax*. MIT Press.

Blaschke, C., Hirschman, L., & Valencia, A. (2002). Information extraction in molecular biology. *Briefings in BioInformatics, 3*(2), 154-165.

Blaschke, C., Oliveros, J., & Valencia, A. (2004). Mining functional information associated to expression arrays. *Functional and Integrative Genomics, 1*(4), 256-268.

Chiang, J., & Yu, H. (2003). MeKE: Discovering the functions of gene products from biomedical literature via sentence alignment. *Bioinformatics, 19*(11), 1417-1422.

Corney, D., Buxton, B., Langdon, W., & Jones, D. (2004). BioRAT: Extracting biological information from full-length papers. *Bioinformatics, 20*(17), 3206-3213.

Couto, F., Martins, B., & Silva, M. (2004). Classifying biological articles using Web resources. *Proceedings of the 2004 ACM Symposium on Applied Computing* (pp. 111-115).

Couto, F., Silva, M., & Coutinho, P. (2005). Finding genomic ontology terms in text using evidence content. *BMC Bioinformatics*, *6*(Suppl 1), S21.

Dickman, S. (2003). Tough mining: The challenges of searching the scientific literature. *PLoS Biology, 1*(2), E48.

Earley, J. (1970). An efficient context-free parsing algorithm. *Communications of ACM, 13,* 94-102.

GO-Consortium (2004). The gene ontology (GO) database and informatics resource. *Nucleic Acids Research, 32*(Database issue), D258-D261.

Hand, D., Mannila, H., & Smyth, P. (2000). *Principles of Data Mining.* MIT Press.

Hearst, M. (1999). Untangling text data mining. *Proceedings of the 37th ACL Meeting of the Association for Computational Linguistics* (pp. 3-10).

Hersh, W., Bhuptiraju, R., Ross, L., Johnson, P., Cohen, A., & Kraemer, D. (2004). TREC 2004 genomics track overview. *Proceedings of the 13th Text REtrieval Conference (TREC 2004)* (pp. 14-24).

Hirschman, L., Park, J., Tsujii, J., Wong, L., & Wu, C. (2002). Accomplishments and challenges in literature data mining for biology. *Bioinformatics, 18*(12), 1553-1561.

Hirschman, L., Yeh, A., Blaschke, C., & Valencia, A. (2005). Overview of BioCrEAtIvE: critical assessment of information extraction for biology. *BMC Bioinformatics*, *6*(Suppl 1), S1.

Kim, J., & Park, J. (2004). BioIE: Retargetable information extraction and ontological annotation of biological interactions from literature. *Journal of Bioinformatics and Computational Biology, 2*(3), 551-568.

Koike, A., Niwa, Y., & Takagi, T. (2005). Automatic extraction of gene/protein biological functions from biomedical text. *Bioinformatics, 21*(7), 1227-1236.

Leake, D. (1996). *Case-based reasoning: Experiences, lessons, and future directions.* AAAI Press/MIT Press.

Lin, D. (1998). An information-theoretic definition of similarity. *Proceedings of the 15th International Conference on Machine Learning* (pp. 296-304).

Manning, C., & Schütze, H. (1999). *Foundations of statistical natural language processing.* The MIT Press.

Mitchell, J., Aronson, A., Mork, J., Folk, L., Humphrey, S., & Ward, J. (2003). Gene indexing: characterization and analysis of NLM's GeneRIFs. Paper presented at the AMIA 2003 Annual Symposia, Washington, DC.

Müller, H., Kenny, E., & Sternberg, P. (2004). Textpresso: An ontology-based information retrieval and extraction system for biological literature. *PLOS Biology, 2*(11), E309.

Palakal, M., Stephens, M., Mukhopadhyay, S., & Raje, R. (2003). Identification of biological relationships from text documents using efficient computational methods. *Journal of Bioinformatics and Computational Biology, 1*(2), 307-342.

Pérez, A., Perez-Iratxeta, C., Bork, P., Thode, G., & Andrade, M. (2004). Gene annotation from scientific literature using mappings between keyword systems. *Bioinformatics, 20*(13), 2084-2091.

Rebholz-Schuhmann, D., Kirsch, H., & Couto, F. (2005). Facts from text - is text mining ready to deliver? *PLoS Biology, 3*(2), e65.

Shatkay, H. & Feldman, R. (2003). Mining the biomedical literature in the genomic era: An overview. *Journal of Computational Biology, 10*(6), 821-855.

Smith, L., Rindflesch, T., & Wilbur, W. (2004). MedPost: A part-of-speech tagger for biomedical text. *Bioinformatics, 20*(14), 2320-2321.

Spencer, A. (1991). *Morphological theory.* Oxford: Blackwell.

Wilks, Y., & Stevenson, M. (1997). The grammar of sense: Using part-of-speech tags as a first step in semantic disambiguation. *Journal of Natural Language Engineering, 4*(3).

Wu, D. & Fung, P. (1994). Improving Chinese tokenization with linguistic filters on statistical lexical acquisition. *Proceedings of the 4th ACL Conference on Applied Natural Language Processing* (pp. 13-15).

Yeh, A., Hirschman, L., & Morgan, A. (2003). Evaluation of text data mining for database curation: Lessons learned from the KDD challenge cup. *Bioinformatics, 19*(1), i331-i339.

This work was previously published in Advanced Data Mining Technologies in Bioinformatics, edited by H. Hsu, pp. 283-295, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.17
Preparing Clinical Text for Use in Biomedical Research

John P. Pestian
Cincinnati Children's Hospital Medical Center, University of Cincinnati, USA

Lukasz Itert
Nicolaus Copernicus University, Torun, Poland

Charlotte Andersen
Cincinnati Children's Hospital Medical Center, University of Cincinnati, USA

Wlodzislaw Duch
Nicolaus Copernicus University, Torun, Poland

ABSTRACT

Approximately 57 different types of clinical annotations construct a patient's medical record. These annotations include radiology reports, discharge summaries, and surgical and nursing notes. Hospitals typically produce millions of text-based medical records over the course of a year. These records are essential for the delivery of care, but many are underutilized or not utilized at all for clinical research. The textual data found in these annotations is a rich source of insights into aspects of clinical care and the clinical delivery system. Recent regulatory actions, however, require that, in many cases, data not obtained through informed consent or data not related to the delivery of care must be made anonymous (as referred to by regulators as harmless), before they can be used. This article describes a practical approach with which Cincinnati Children's Hospital Medical Center (CCHMC), a large pediatric academic medical center with more than 761,000 annual patient encounters, developed open source software for making pediatric clinical text harmless without losing its rich meaning. Development of the software dealt with many of the issues that often arise in natural language processing, such as data collection, disambiguation, and data scrubbing.

INTRODUCTION

Hospitals typically produce millions of text-based medical records over the course of a year. These records are essential for the delivery of care but underutilized or not utilized at all for clinical research. Digitized clinical data are a rich lode of possibilities for advances in biomedical research, because, in aggregate, they contain information about the variation in the delivery and quality of care.

Inherent in such research, however, is the use of data without the patient's consent. Recognizing this problem, the United States Department of Health and Human Services (HHS) has issued rules defining Protected Health Information (PHI) as part of the Health Insurance Portability and Accountability Act of 1996 (HIPAA) (Annas, 2002). In order for researchers to access such data, either they must have the patient's consent, or, as in most retrospective cases, the data must be made harmless, and the governing board must provide a waiver.

The HHS provides guidance for making health-care data harmless (HIPAA Standards for Privacy of Individually Identifiable Health Information: An Introduction to the Consent Debate, 2002). Data can be made harmless through three steps: (1) deidentification (i.e., the removal or modification of data fields that could identify a patient, such as name and social security number); (2) rendering the data ambiguous by ensuring that every data record in a public data set has a non-unique set of characterizing data (Berman, 2002a; Bouzelat, Quantin, & Dusserre, 1996; Quantin et al., 1998); and (3) data scrubbing (i.e., the removal or transformation of those tokens in text that can be used to identify persons or that contain information that is incriminating or otherwise private) (Berman, 2003; Sweeney, 1996). Although each of these methods has the potential to render the medical record harmless for its use by natural language processing investigators, attempts to design a fully anonymous system continue.

This article describes how Cincinnati Children's Hospital Medical Center (CCHMC), a large pediatric academic medical center with more than 761,000 pediatric patient encounters per year, has taken a practical approach to this challenge by developing, evaluating, and implementing the Encryption Broker (EB) software. The EB has a number of uses. First, it is essential for the ongoing development of a large pediatric corpus for pediatric natural language processing research and decision support (Pestian, Itert, & Duch, 2004). This corpus serves as an artificial intelligence training set for classifying text into the appropriate clinical domain, such as rheumatology or neonatology. Without the EB, these data could not be retrieved from the electronic portion of the medical records. Second, the EB ensures that research-needing text conforms to federal regulations. It does so through data disambiguation algorithms, deidentification, and data scrubbing.

The EB has another role. A key strategy of the organization is personalized medicine research that requires genomic and clinical delivery data to predict or prevent disease or to personalize treatment. This research requires substantial amounts of knowledge to be gleaned automatically from these data in real time. To do so, machine-learning systems that conceptually map the data into some ontology are required. The EB provides natural language scientists with large repositories of harmless clinical text for developing these systems.

The EB is recognized by CCHMC's Risk Management group as a tool to gather clinical text without violating HIPAA regulations. This approval is institution-specific; each institution using the EB is responsible for seeking its own internal certification. The EB essentially acts as a broker for investigators who wish to do retrospective analysis of clinical text and potentially makes it easier to receive approval for these purposes. CCHMC makes the EB software, the associated decision rules, and the related data files fully available for academic purposes by e-mailing joan.taylor@cchmc.org. The remaining sections of this article discuss methods and challenges for making these data harmless, CCHMC's approach, and the evaluation of this methodology.

LITERATURE REVIEW

It is beyond the scope of this article to describe fully the rich history of research in the areas of natural language processing; this review highlights those areas that have contributed to developing the conceptual approach underpinning the research presented: word sense disambiguation and data scrubbing.

WORD SENSE DISAMBIGUATION

Examining tokens in their context and determining exactly what sense is being used is the task of Word Sense Disambiguation (WSD). WSD is a difficult task and, as such, receives considerable theoretical and practical attention. To disambiguate (i.e., OR to operating room) requires an understanding of the surrounding tokens. In other words, "You shall know a word by the company it keeps" (Firth, 1957). There are two ways to do this. One is a supervised approach that integrates rule-based information into the semantic analysis. The other is an unsupervised stand-alone approach, where sense disambiguation is performed independent of and prior to compositional semantic analysis.

For this research, integrated rule-to-rule approach was used, because raw clinical notations are heavily packed with jargon, and unsupervised methods are traditionally used with well-formed text. Ng and Zelle (1997) note:

For each token to be disambiguated, the appropriate inference knowledge must be handcrafted. It is difficult to come up with a comprehensive set of the necessary disambiguation knowledge. Also, as the amount of disambiguation knowledge grows, manual maintenance and further expansion become increasingly complex. Thus, it is difficult to scale up manual knowledge acquisition to achieve wide coverage for real-world sentences.

This summary points out the limitations and provides future research guidance. That is, since rule-to-rule WSD requires substantial effort at some point, it will be necessary to integrate this work into a stand-alone unsupervised machine learning system.

Determining the optimal window size for token analysis is another important task. The linguistic tools used for WSD can be divided into two general classes: collocation and co-occurrence. Co-location, a quantifiable position-specific relationship between two lexical items, encodes local

Figure 1. Trigram analysis neighborhood

Analysis Neighborhood

Notation	B_{T3} B	$_{T2}$ B	$_{T1}$	τ	A_{T1} A	$_{T2}$ A	$_{T3}$
Example	John	was	born	FT	with	no	complications

Where Btn = tokens before the Evaluation token, ET = evaluation token and ATN equals tokens after the evaluation token.

Table 1. Output example

Before	After
Fred Thompson* is an 8 y/o AAM with a hx of asthma. He presented in the ED with a laceration on his R radius approx 3 in. long.	John Johnson is an eight-year-old African American male with a history of asthma. He presented in the emergency department with a laceration on his right radius approximately 3 inches long.

** Fred Thompson is not the patient's name.*

lexical and grammatical information that often can accurately isolate a given sense (Jurafsky & Martin, 2000). In collocation, the assumption is that some tokens often are found together (e.g., emergency room or breast milk).

Co-occurrence data focus on the frequency of the same token within a particular range of tokens while ignoring its position. For example, "John's parents were in the *emergency room* while the *emergency room* physician treated John." Co-occurrence focuses on the fact that *emergency room* occurred twice. Collocation focuses on the fact that *emergency* is located next to *room*.

These tools enable selection of specific domain tokens from a larger generalized corpus (Jurafsky & Martin, 2000). This study formally uses local collocations to disambiguate terms. In particular, +/- three tokens around the target token (**t**) were analyzed. This window of tokens is referred to as a trigram. This strategy was based on previous research that notes:

[L]ocal co-location provides the most important source of disambiguation knowledge, although the accuracy of disambiguation achieved by the combined knowledge sources exceeds that obtained by using any one of the knowledge sources alone. That local collocation is the most predictive seems to agree with past observation that humans need a narrow window of only a few tokens to perform WSD. (Ng & Zelle, 1997)

DATA SCRUBBING

The literature describes many forms of data scrubbing. Scientists use data-scrubbing methods to de-identify pathology data (Berman, 2003), threshold cryptographic protocols (Berman, 2002b), automate record hash coding and linkages for epidemiological follow-up data confidentiality (Quantin et al., 1998), object-oriented software components (Herting & Barnes, 1998), cryptographic framework for document objects resulting from multiparty collaborative transactions (Goh, 2000), use personal identifiers while retaining confidentiality in child abuse cases (Kruse, Ewigman, & Tremblay, 2001), and describe data hiding techniques (Chao, Hsu, & Miaou, 2002).

Although research in the area of deidentification has been active over the last few years, scholars are still undecided as to whether it is possible to fully de-identify data. For example, Sweeney and Dreiseitl conclude that most data can be reidentified by linking or matching the data to other databases or by looking at unique characteristics found in the fields and records of the database itself (Dreiseitl, Vinterbo, & Ohno-Machado, 2001; Sweeney, 1997a). Sweeny, after reviewing a number of data-scrubbing systems, concludes that removing all explicit identifiers from medical data does not guarantee anonymity; rather, complementary policies will be necessary (Sweeney, 1997b). Others, however, regard those processes as too onerous to yield any practical consideration (Fisher, Baron, DJ, Barett, & Bubolz, 1990). Until the optimal set of strategies is found, each institution must address problems with deidentification as it finds best.

METHODS

A patient's medical record is comprised of approximately 57 different types of documents (Zweigenbaum, Jacquemart, Grabar, & Habert, 2001). These documents contain both structured data (e.g., computerized order entry data) and unstructured data (e.g., clinical dictations). Some data are confidential; others are a matter of public record. Computerized or handwritten notes include birth and death records, discharge summaries, imaging reports, short problem descriptions, and letters (Friedman, 1997; Grefenstette, 1994; Sager, Friedman & Lyman, 1987; Zweigenbaum & Menelas, 1994). The content of these documents has a great deal of variation not only between the documents but also within the documents

Table 2. Disambiguation rule: Pseudo-code and rule coding

<table>
<tr><th>Pseudo-Code</th><th>Rules File Coding*</th></tr>
<tr><td>Evaluation Token = ALL
If ALL is all upper case and preceded by HISTORY OF, RULE OUT, RULING OUT, H/O, B-CELL, T-CELL, FOR, HIGH RISK, #REFRACTORY, PROBABLE, WITH, T-CELL, PRE-B, RELAPSED, then change ALL to Acute Lymphocytic Leukemia.
If ALL is upper case and followed by LOW RISK, then ALL= Acute Lymphocytic Leukemia.
#All others stay as ALL.</td><td>%ALL;|||PRE_ISM(out,h/o,b-cell,t-cell,for,risk,refractory,probable, with,t-cell,pre-b,relapsed)||||;Acute Lymphocytic Leukemia
%ALL;||PRE_INC_PHR(history,of)|||||;Acute Lymphocytic Leukemia
%ALL;||PRE_INC_PHR(low,risk)|||||;Acute Lymphocytic Leukemia
ALL;|||FINAL()||||;ALL</td></tr>
<tr><td>Evaluation Token = mm
If mm is preceded by moist, dry, pale, sticky, tacky, then change mm to mucus membranes.
If mm is followed by moist, dry, tacky, pale, sticky, then change mm to mucus membranes.
If mm is immediately preceded by a number (i.e., 100, 7.1, 13-14, etc.), then change mm to millimeters.
If mm is followed by clinic, repair, sac, workup, surgery then change mm to Myelomeningocele.
If mm is immediately preceded by, diagnosis of, known, h/o, history of, s/p, secondary to, with, then change mm to Myelomeningocele
All others stay as mm</td><td>mm;|||PRE_ISM(moist,dry,pale,sticky,tacky)||||;mucus membranes
mm;||||POST_ISM(moist,dry,tacky,pale,sticky)|||;mucus membranes
mm;|||NUM()||||;millimeters
mm;||||POST_ISM(clinic,repair,sac,workup,surgery)|||;Myelomeningocele
#ADDED:
mm;|ANY_ISM(known,vp,shunt,thoracic,secondary, spina,bifida)||||||;Myelomeningocele
#
mm;||IS(diagnosis)|IS(of)||||;Myelomeningocele
mm;|||ISM(known,h/o,s/p,with)||||;Myelomeningocele
mm;||IS(history)|IS(of)||||;Myelomeningocele
mm;||IS(secondary)|IS(to)||||;Myelomeningocele
mm;|||FINAL()||||;mm</td></tr>
</table>

**Note: Full technical documentation is provided online at http://info.cchmc.org.*

themselves (Biber & Finegan, 1994). This study concentrates on unstructured clinical text found in discharge summaries, radiology reports, surgical reports, and pathology reports.

The minimum regulatory standards for making PHI harmless require removal of up to 16 specific pieces of information (Madsen, Masys, & Miller, 2003). In the case of unstructured text, simply removing or encrypting these identifiers will disrupt the ability to understand the PHI and its meaning, thus rendering it useless for natural language processing research.

The remaining sections of this article outline the methods for collecting data, development of rules, three stages of software development, and the evaluation of the software.

DATA COLLECTION

From 2000 to 2002, CCHMC's division of Biomedical Informatics developed the Discovery System (DS), a centralized research repository (Pestian, Aronow, & Davis, 2002). The DS is populated regularly with new and updated clinical, research, and administrative data generated by the medical center. Substantial amounts of these clinical data are text from such specialties as pathology and radiology and from discharge summaries and surgical notes. The DS combined with other data are used for studying genotypic prediction of pharmacological responses and microarray expression of newborn hearing testing, sepsis onset in intensive care patients, the onset and severity of juvenile rheumatoid arthritis,

Table 3. Descriptive statistics

Description	Total
Total tokens in data set	19,924,949
Total sentences in data set	1,263,271
Average tokens/sentence (standard deviation)	15.33 (9.93)
Total paragraphs in data set	173,933
Average number of sentences per paragraph (standard deviation)	7.42 (20.44)
Total unique tokens in data set	129,282
Total trigrams in data set	20,291,335
Total unique trigram in data set	5,118,035

Table 4. Evaluation descriptive statistics

Description	Total
Tokens	133,210
Sentences	10,240
Average tokens per sentence (SD)	13 (9)
Correct number of changes (% of Total Tokens)	1420 (99.2)
Incorrect number of changes (% of Total Tokens)	110 (0.08)
Correct number of non-changes (% of Total Tokens)	132,670 (98.93)
Incorrect number of non-changes (% of Total Tokens)	770 (0.58)

quality assurance, financial reporting, and other activities.

Access to the data for research is governed by HIPAA regulations and controlled by the organization's Institutional Review Board (IRB). Prospective studies receive approval before the study begins. Access to retrospective data also must receive approval from the IRB. Requests for text that are not part of a formal research study sanctioned by the IRB can only be approved only after the data have been made harmless by using the EB or some other method.

DATA CLEANSING METHOD

The data-cleansing algorithm relies on two steps in order to render the unstructured clinical text harmless and preprocess it for use. The first step is to disambiguate the unstructured clinical text that is dense with jargon and acronyms. The second step is data scrubbing or deidentification. Each of these steps is described in subsequent paragraphs.

WORD SENSE DISAMBIGUATION

"All grammars leak" (Sapir, 1921). This is because people are always stretching and bending the rules to meet their communicative needs (Manning & Schutze, 1999). It should be no surprise that extensive jargon and acronyms have leaked into clinical text. The language of clinicians, though fundamental to patient care, lacks the structure and clarity necessary for natural language analysis. For example, in a clinical text, the token FT can be an abbreviation for *full-term*, *fort* (as in Fort Sumter), *feet or foot*, *field test*, *full-time*, or *family therapy*. Until these text data are disambiguated, there is no certainty that data scrubbing is accurate.

To resolve the ambiguities found in the text, a series of clinical disambiguation rules were made. The first step for developing these rules was to create a reference dataset that contained known ambiguous terms, clinical acronyms, and abbreviations. After developing a dataset of known acronyms and abbreviations, clinical experts reviewed the text for ambiguous terms. Ambiguous terms were added to the dataset. This review was done three times until the experts believed that most ambiguous terms were included in the dataset.

This reference dataset was then used to create a dataset of trigrams. Software was developed to extract from the all the data the trigrams for each ambiguous term. Clinical experts then reviewed these trigrams to create the disambiguation rules. Figure 1 presents a schematic of this approach. In the figure, one term, *FT,* is being evaluated by looking at the three tokens before *FT* and the three tokens after *FT.* The experts then reviewed all the trigrams and developed the disambiguation rules, using a majority/minority approach. That is, all instances of a specific term (i.e., *FT*) remain as a specific term (i.e., *FT*), unless an evaluation parameter is met. For example, one rule is *If FT* is proceeded by *born*; then FT = *Full-Term*. When developing rules, the use of expert(s) with subject matter knowledge substantially decreases the likelihood of disambiguation error. For example, based on the trigram in Figure 1, one may develop the rule: *if FT* is followed by *with*; then FT = *Full-Term*. Keeping in mind that FT can be an acronym for, say, *family therapy* or *foot*, this rule would be incorrect. Examples such as *patient attended family therapy with Dr. Johnson* or *laceration on right foot with decrease pedal pulses* would also meet the described conditions, thus voiding the rule.

DATA SCRUBBING

Once the data were disambiguated, they were reviewed for the presence of any of the 16 possible Protected Health Information (PHI) data elements. Limited PHI was found in the unstructured text fields. What were found were the patient and physician names and, rarely, a date of service; all other PHI was located in other structured database fields and could be eliminated by excluding those fields from the original query. Next, systematic bias was introduction into the data as a method of encryption; all female names were changed to *Jane*, all male names were changed to *John*, and all surnames were changed to *Johnson*. Table 1 provides an example of how the input data were changed.

TOKEN EVALUATION

The token evaluation criteria are based on the n-gram approach where n = the number of tokens to be evaluated before and after the token under consideration. The default value is NGRAM = 3, or a trigram. Thus,

τ;||| ||||;δ

is the syntax to evaluate a particular token, where τ represents the token under consideration, and δ represents its replacement. In this case, because NGRAM = 3, the series of pipes (|) symbolize the trigram of three tokens preceding and the three tokens following τ; seven pipes yield six points to evaluate τ. Bigrams would have five pipes, and so forth. In this case, the positions are |1|2|3|4|5|6|.

Numbers 1, 2, and 3 are positions of tokens preceding τ; numbers 4, 5, and 6 are positions of tokens that occur after the τ. Consider the following sentence.

The patient stayed in OR for one hour.

Each token is assigned a position:

Patient/1/ stayed/2/ in/3/ for/4/ one/5/ hour/6/

τ = *OR* it is excluded from the position assignment.

Next is the syntax for the rule to evaluate the abbreviation *OR* based on its collocation to *IN* using one of the more than 40 predefined conditions and placed at the NGRAM position. Detailed software documentation is included.

or;|||CONDITION(in)|||;operating room

If the condition is fulfilled, then the abbreviation *OR* will be replaced with *operating room* in text. If the condition is not fulfilled, then the next condition is considered. This occurs until the last condition is evaluated via an exit criterion.

A typical rule for the mentioned example could look like this:

or;|||IS(in)|||;operating room
or;||||IS(for)||;operating room
or;||||FINAL()|||;or

The first condition evaluates if the token *in* is before *operating room*. If this condition is not satisfied, the second condition is analyzed. If *OR* is followed by *for*, then *OR* is replaced with *operating room*, but if this is not true, the last condition says that the token should remain as *OR*. There are more than 40 predefined conditions (e.g., IS, PRE_NUM, POST_NUM) that can be used for testing. By default, all abbreviations are converted to lower case. Table 2 shows the pseudo code and the corresponding syntax.

EVALUATION

Evaluation of the EB consisted of randomly selecting encrypted sentences and pairing them with the original sentences. Clinical experts then reviewed these data and classified each token into one of four categories: a correct replacement, an incorrect replacement, a correct miss, and an incorrect miss. Proportions were then computed.

RESULTS

Processing scripts were written in Perl 5.0. Processing took place on a Sun Microsystems E6500, using 12 900-Mhz processors with 24 GB RAM.

DATA COLLECTION

All 2002 clinical texts were extracted from the DS. Table 3 provides the descriptive statistics for these data.

DISAMBIGUATION

Tokens of 915 candidate-ambiguous terms included all approved hospital acronyms, unapproved acronyms, and terms that were found. Clinical experts reviewed 715,518 trigrams that included these ambiguous terms. From this review 1,146 distinct rules for resolving ambiguity of the tokens were developed. Each rule was based on reviewing an average of 781 trigrams for a particular ambiguous token. These rules were then added to the EB's rules file to enable ambiguity resolution during data parsing.

DATA SCRUBBING

A review of the text fields found that the PHI present in the text clinical annotations was the patient's name, physician's name, and various dates. This finding made the algorithm for data scrubbing rather straightforward; by the introduction of systematic bias, data could be changed without compromising their meanings. The software's algorithm changed all male names to John, all female names to Jane, all surnames to Johnson, and all dates to 01/01/2005. This version of the software does not deal with neutral names (e.g., Pat). Future versions will.

EVALUATION RESULTS

The EB was evaluated by randomly selecting 348 records (a 0.05, 95% CI) from the original data and pairing these data with the corresponding data output from the EB. Table 4 shows the results of this comparison. A total of 10,240 (paired) sentences were reviewed by clinical experts. Ninety-eight percent of the time, the EB correctly changed a token; equally important, 99% of the time, when a token should not be changed, it was not. Of those tokens that were incorrectly changed (0.58%), a clear pattern emerged. The majority of these errors were related to ambiguous names. For example, the token *may* can mean the given name *May*, the month of *May*, or the command that *he may play sports in two weeks*. Errors in the output were found when any of the supporting files were not kept current.

SUMMARY

Protecting health information always has been a responsibility of healthcare organizations. Now that HIPAA regulations require additional levels of accountability, healthcare organizations must be creative when rendering such data harmless for research purposes. This approach shows that this is possible, but it has taken considerable effort, expense, and resources to develop and to evaluate the appropriate software. For example, to develop the first set of rules, the process includes collecting data, manually reviewing more than 700,000 trigrams to develop more than 1,000 disambiguation rules.

An important next step will be to determine the possibility of migrating from a hand-crafted rules approach to rules that are made based on supervised or unsupervised machine learning algorithms. A recent paper by Liu et al. (2004) best describes this discussion: “Supervised WSD is suitable only when we have enough sense-tagged instances with at least a few dozens of instances for each sense." Here, sense-tagged refers to ambiguous tokens that have been clarified via various methods like collocation or co-occurrence. "The combination of collocations and neighboring tokens are appropriate selections for the context. For terms with biomedical unrelated senses, a large window size such as the whole paragraph should be used, while for general English words a moderate window size between four and ten should be used" (Liu, Teller, & Friedman, 2004). Thus suggesting that the optimal method by be a combination of hand-crafter rules, and machine learning.

Other questions remain unresolved. First, how generalizable are disambiguation rules? That is, is the jargon used by clinicians in one part of the country or in one hospital, for that matter, different from the jargon used in another part of the country or another hospital? Second, how generalizable are disambiguation rules from the pediatric population to adult populations? While it is conjectured that there is little differences, certain differences will be inherent in the populations (i.e., adults will not be diagnosed with atrial septal defects; likewise children will not have coronary artery bypass grafts procedures). Third, how will a patient's longitudinal records be linked with this approach?

REFERENCES

Annas, G. J. (2002). Medical privacy and medical research — Judging the new federal regulations. *N Engl J Med, 346*(3), 216-220.

Berman, J. J. (2002a). Confidentiality issues for medical data miners. *Artif Intell Med, 26*(1-2), 25-36.

Berman, J. J. (2002b). Threshold protocol for the exchange of confidential medical data. *BMC Med Res Methodol, 2*(1), 12.

Berman, J. J. (2003). Concept-match medical data scrubbing. *Arch Pathol Lab Med, 127*(6), 680-686.

Biber, D., & Finegan, E. (1994). Intra-textual variation within medical research articles. In N. Ooostdijk, & P. de Haan (Eds.), *Corpus-based research into language* (Vol. 12) (pp. 201-222). Amsterdam: Rodopi.

Bouzelat, H., Quantin, C., & Dusserre, L. (1996). Extraction and anonymity protocol of medical files. In *Proceedings of the AMIA Annual Fall Symposium* (pp. 323-327).

Chao, H. M., Hsu, C. M., & Miaou, S. G. (2002). A data-hiding technique with authentication, integration, and confidentiality for electronic patient records. *IEEE Trans Inf Technol Biomed, 6*(1), 46-53.

Dreiseitl, S., Vinterbo, S., & Ohno-Machado, L. (2001). Disambiguation data: Extracting information from anonymized sources. In *Proceedings of the AMIA Symposium* (pp. 144-148).

Firth, J. E. (1957). A synopsis of linguistic theory 1930-1955. In F. R. Palmer (Ed.), *Selected papers of J. R. Firth 1952-1959* (pp. 1-32). London: Oxford Philogical Society.

Fisher, E. S., Baron, J., DJ, M., Barett, J., & Bubolz, T. (1990). Overcoming potential pitfalls in the use of Medicare data for epidemiologic research. *Am J Public Health, 80*(12), 1487-1490.

Friedman, C. (1997). Towards a comprehensive medical natural language processing system: Methods and issues. *Journal of the American Medical Informatics Association, 4*(Suppl), 595-599.

Goh, A. (2000). Cryptographic framework for document-objects resulting from multiparty collaborative transactions. *Stud Health Technol Inform, 77,* 1069-1073.

Grefenstette, G. (1994). *Explorations in automatic thesaurus discovery*. London: Kluwer.

Herting, R. L., Jr., & Barnes, M. R. (1998). Large scale database scrubbing using object oriented software components. In *Proceedings of the AMIA Symposium* (pp. 508-512).

HIPAA standards for privacy of individually identifiable health information: An introduction to the consent debate. (2002). *J Health Law, 35*(3), 387-394.

Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing.* Upper Saddle River, NJ: Prentice Hall.

Kruse, R. L., Ewigman, B. G., & Tremblay, G. C. (2001). The zipper: A method for using personal identifiers to link data while preserving confidentiality. *Child Abuse Negl, 25*(9), 1241-1248.

Liu, H., Teller, V., & Friedman, C. (2004). A multi-aspect Comparison study of supervised word sense disambiguation. *Journal of the American Medical Informatics Association*.

Madsen, E., Masys, D. R., & Miller, R. A. (2003). HIPPA possums. *Journal of the American Medical Informatics Association, 10*(3), 294.

Manning, C. D., & Schutze, H. (1999). *Foundations of statistical natural language processing.* Cambridge: MIT Press.

Ng, H. T., & Zelle, J. (1997). Corpus-based approach to semantic interpretation in NLP. *AI, 18*(4), 45-54.

Pestian, J., Aronow, B., & Davis, K. (2002). *Design and data collection in the discovery system*. Paper presented at the International Conference on Mathematics and Engineering Techniques in Medicine and Biological Science.

Pestian, J. P., Itert, L., & Duch, W. (2004, May). *Development of a pediatric text-corpus for part-of-speech tagging.* Paper presented at the Intelligent Information Systems, Poland.

Quantin, C., Bouzelat, H., Allaert, F. A., Benhamiche, A. M., Faivre, J., & Dusserre, L. (1998). Automatic record hash coding and linkage for epidemiological follow-up data confidentiality. *Methods Inf Med, 37*(3), 271-277.

Sager, N., Friedman, C., & Lyman, M. (Eds.). (1987). *Medical information processing — Computer management of narrative data*. Reading, MA: Addison Wesley.

Sapir, E. (1921). *Language: An introduction to the study of speech*. New York: Harcort Brace.

Sweeney, L. (1996). Replacing personally-identifying information in medical records, the scrub system. In J. J. Cimino (Ed.), *Proceedings, Journal of the American Medical Informatics Association* (Vol. 1996, pp. 333-337). Washington, DC: Hanley & Belfus, Inc.

Sweeney, L. (1997a). Guaranteeing anonymity when sharing medical data, the datafly system. In *Proceedings of the AMIA Annual Fall Symposium* (pp. 51-55).

Sweeney, L. (1997b). Weaving technology and policy together to maintain confidentiality. *J Law Med Ethics, 25*(2-3), 98-110, 182.

Zweigenbaum, P., Jacquemart, P., Grabar, N., & Habert, B. (2001). *Building a text corpus for representing the variety of medical language*. Paper presented at the Medinfo.

Zweigenbaum, P., & Menelas, C. (1994). Menelas: An access system for medical records using natural language. *Compt Methods Programs Biomed, 45*, 117-120.

This work was previously published in Journal of Database Management, Vol. 17, Issue 2, edited by K. Siau, pp. 1-11, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.18
Knowledge Discovery in Biomedical Data Facilitated by Domain Ontologies

Amandeep S. Sidhu
Curtin University of Technology, Australia

Paul J. Kennedy
University of Technology Sydney, Australia

Simeon Simoff
University of of Western Sydney, Australia

Tharam S. Dillon
Curtin University of Technology, Australia

Elizabeth Chang
Curtin University of Technology, Australia

ABSTRACT

In some real-world areas, it is important to enrich the data with external background knowledge so as to provide context and to facilitate pattern recognition. These areas may be described as data rich but knowledge poor. There are two challenges to incorporate this biological knowledge into the data mining cycle: (1) generating the ontologies; and (2) adapting the data mining algorithms to make use of the ontologies. This chapter presents the state-of-the-art in bringing the background ontology knowledge into the pattern recognition task for biomedical data.

INTRODUCTION

Data mining is traditionally conducted in areas where data abounds. In these areas, the task of the data mining is to identify patterns within the data, which may eventually become knowledge. To this end, the data mining methods used, such as cluster analysis, link analysis and classification and regression, typically aim to reduce the amount of information (or data) to facilitate this pattern recognition. These methods do not tend to contain (or bring to the problem) specific domain specific information. In this way, they may be termed "knowledge-empty." However, in some real-world areas, it is important to enrich the data with external background knowledge so as to provide context and to facilitate pattern recognition. These areas may be described as data rich but knowledge poor. External background information that may be used to enrich data and to add context information, and facilitate data mining is in the form of ontologies, or structured vocabularies. So long as the original data can be linked to terms in the ontology, the ontology may be used to provide the necessary knowledge to explain the results and even generate new knowledge.

In accelerating quest for disease biomarkers, the use of high-throughput technologies, such as DNA microarrays and proteomics experiments, has produced vast datasets identifying thousands of genes whose expression patterns differ in diseased vs. normal samples. Although many of these differences may reach statistical significance, they are not biologically meaningful. For example, reports of mRNA or protein changes of as little as two-fold are not uncommon, and although some changes of this magnitude turn out to be important, most are attributes to disease-independent differences between the samples. Evidence gleaned from other studies linking genes to disease is helpful, but with such large datasets, a manual literature review is often not practical. The power of these emerging technologies—the ability to quickly generate large sets of data—has challenged current means of evaluating and validating these data. Thus, one important example of a data rich but knowledge poor area is biological sequence mining. In this area, there exist massive quantities of data generated by the data acquisition technologies. The bioinformatics solutions addressing these data are a major current challenge. However, domain specific ontologies such as gene ontology (GO Consortium, 2001), MeSH (Nelson & Schopen, 2004) and protein ontology (Sidhu & Dillon, 2005a, 2006a) exist to provide context to this complex real world data.

There are two challenges to incorporate this biological knowledge into the data mining cycle: (1) generating the ontologies; and (2) adapting the data mining algorithms to make use of the ontologies. This chapter presents the state-of-the-art in bringing the background ontology knowledge into the pattern recognition task for biomedical data. These methods are also applicable to other areas where domain ontologies are available, such as text mining and multimedia and complex data mining.

GENERATING OONTOLOGIES: CASE OF PROTEIN ONTOLOGY

This section is devoted to the practical aspects of generating ontologies. It presents the work on building the protein ontology (Sidhu et al., 2006a; Sidhu & Dillon, 2005a, 2006b; Sidhu et al., 2005b) in the section "Protein Ontology (PO)." It then compares the structures of the protein ontology and the well established gene ontology (GO Consortium, 2001) in the section "Comparing PO and GO."

Protein Ontology (PO)

Advances in technology and the growth of life sciences are generating ever increasing amounts of data. High-throughput techniques are regularly used to capture thousands of data points in an experiment. The results of these experiments normally end up in scientific databases and publications. Although there have been concerted efforts to capture more scientific data in specialist databases, it is generally acknowledged that only

20% of biological knowledge and data is available in a structured format. The remaining 80% of biological information is hidden in the unstructured scientific results and texts. Protein ontology (PO) (Sidhu et al., 2006a; Sidhu et al., 2006b; Sidhu et al., 2005a; Sidhu et al., 2005b) provides a common structured vocabulary for this structured and unstructured information and provides researchers a medium to share knowledge in proteomics domain. It consists of concepts, which are data descriptors for proteomics data and the relations among these concepts. Protein ontology has (1) a hierarchical classification of concepts represented as classes, from general to specific; (2) a list of attributes related to each concept, for each class; and (3) a set of relations between classes to link concepts in ontology in more complicated ways then implied by the hierarchy, to promote reuse of concepts in the ontology. Protein ontology provides description for protein domains that can be used to describe proteins in any organism. Protein ontology framework describes: (1) protein sequence and structure information, (2) protein folding process, (3) cellular functions of proteins, (4) molecular bindings internal and external to proteins and (5) constraints affecting the final protein conformation. Protein ontology uses all relevant protein data sources of information. The structure of PO provides the concepts necessary to describe individual proteins, but does not contain individual protein themselves. Files using Web ontology language (OWL) format based on PO acts as instance store for the PO. PO uses data sources include new proteome information resources like PDB, SCOP, and RESID as well as classical sources of information where information is maintained in a knowledge base of scientific text files like OMIM and from various published scientific literature in various journals. PO database is represented using OWL. PO database at the moment contains data instances of following protein families: (1) prion proteins, (2) B.Subtilis, (3) CLIC and (4) PTEN. More protein data instances will be added as PO is more developed. The complete class hierarchy of protein ontology (PO) is shown in Figure 1. More details about PO is available at the Web site: http://www.proteinontology.info/

Semantics in protein data is normally not interpreted by annotating systems, since they are not aware of the specific structural, chemical and cellular interactions of protein complexes. Protein ontology framework provides specific set of rules to cover these application specific semantics. The rules use only the relationships whose semantics are predefined to establish correspondence among terms in PO. The set of relationships with predefined semantics is: {SubClassOf, PartOf,

Figure 1. Class hierarchy of protein ontology

- ProteinOntology
    - AtomicBind
    - Atoms
    - Bind
    - Chains
    - Family
    - ProteinComplex
        - ChemicalBonds
            - CISPeptide
            - DisulphideBond
            - HydrogenBond
            - ResidueLink
            - SaltBridge
        - Constraints
            - GeneticDefects
            - Hydrophobicity
            - ModifiedResidue
        - Entry
            - Description
            - Molecule
            - Reference
        - FunctionalDomains
            - ActiveBindingSites
            - BiologicalFunction
                - PathologicalFunctions
                - PhysiologicalFunctions
            - SourceCell
        - StructuralDomains
            - Helices
                - Helix
                    - HelixStructure
            - OtherFolds
                - Turn
                    - TurnStructure
            - Sheets
                - Sheet
                    - Strands
        - Structure
            - ATOMSequence
            - UnitCell
    - Residues
    - SiteGroup

AttributeOf, InstanceOf, and ValueOf}. The PO conceptual modelling encourages the use of strictly typed relations with precisely defined semantics. Some of these relationships (like SubClassOf, InstanceOf) are somewhat similar to those in RDF schema but the set of relationships that have defined semantics in our conceptual PO model is small so as to maintain simplicity of the system. The following is a description of the set of predefined semantic relationships in our common PO conceptual model.

- **SubClassOf:** The relationship is used to indicate that one concept is a subclass of another concept, for instance: SourceCell SubClassOf FunctionalDomains. That is any instance of SouceCell class is also instance of FunctionalDomains class. All attributes of FunctionalDomains class (_FuncDomain_Family, _FuncDomain_SuperFamily) are also the attributes of SourceCell class. The relationship SubClassOf is transitive.
- **AttrributeOf:** This relationship indicates that a concept is an attribute of another concept, for instance: _FuncDomain_Family AttributeOf Family. This relationship also referred as PropertyOf, has same semantics as in object-relational databases.
- **PartOf:** This relationship indicates that a concept is a part of another concept, for instance: Chain PartOf ATOMSequence indicates that Chain describing various residue sequences in a protein is a part of definition of ATOMSequence for that protein.
- **InstanceOf:** This relationship indicates that an object is an instance of the class, for instance: ATOMSequenceInstance_10 InstanceOf ATOMSequence indicates that ATOMSequenceInstance_10 is an instance of class ATOMSequence.
- **ValueOf:** This relationship is used to indicate the value of an attribute of an object, for instance: "Homo Sapiens" ValueOf OrganismScientific. The second concept, in turn has an edge, OrganismScientific AttributeOf Molecule, from the object it describes.

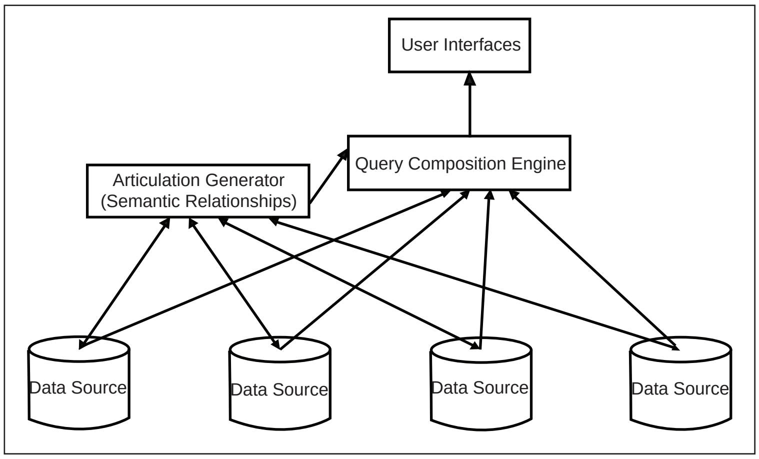
Comparing PO and GO

Gene ontology (GO Consortium, 2001) defines a structured controlled vocabulary in the domain of biological functionality. GO initially consisted of a few thousand terms describing the genetic workings of three organisms and was constructed for the express purpose of database interoperability; it has since grown to a terminology of nearly 16,000 terms and is becoming a de facto standard for describing functional aspects of biological entities in all types of organisms. Furthermore, in addition to (and because of) its wide use as a terminological source for database-entry annotation, GO has been used in a wide variety of biomedical research, including analyses of experimental data (GO Consortium, 2001) and predictions of experimental results (GO Consortium & Lewis, 2004)). Characteristics of GO that we believe are most responsible for its success: community involvement; clear goals; limited scope; simple, intuitive structure; continuous evolution; active curation; and early use.

It is clear that organisms across the spectrum of life, to varying degrees, possess large numbers of gene products with similar sequences and roles. Knowledge about a given gene product (i.e., a biologically active molecule that is the deciphered end product of the code stored in a gene) can often be determined experimentally or inferred from its similarity to gene products in other organisms. Research into different biological systems uses different organisms that are chosen because they are amenable to advancing these investigations. For example, the rat is a good model for the study of human heart disease, and the fly is a good model to study cellular differentiation. For each of these model systems, there is a database employing curators who collect and store the body of biological knowledge for that organism. This enormous amount of data can potentially add insight to related molecules found in other organisms. A reliable wet-lab biological experiment performed in one organism can be used to deduce attributes of an analogous (or related) gene product in another organism, thereby reducing the need to reproduce experiments in each individual organism (which would be expensive, time-consuming, and, in many organisms, technically impossible). Mining of scientific text and literature is done to generate list of keywords that is used as GO terms. However, querying heterogeneous, independent databases in order to draw these inferences is difficult: The different database projects may use different terms to refer to the same concept and the same terms to refer to different concepts. Furthermore, these terms are typically not formally linked with each other in any way. GO seeks to reveal these underlying biological functionalities by providing a structured controlled vocabulary that can be used to describe gene products, and shared between biological databases. This facilitates querying for gene products that share biologically meaningful attributes, whether from separate databases or within the same database.

Challenges faced while developing GO from unstructured and structured data sources are addressed while developing PO. Protein ontology is a conceptual model that aim to support consistent and unambiguous knowledge sharing and that provide a framework for protein data and knowledge integration. PO links concepts to their interpretation, that is, specifications of their meanings including concept definitions and relationships to other concepts. Apart from semantic relationships defined in "Protein Ontology (PO)," PO also model relationships like sequences. By itself semantic relationships described in "Protein Ontology (PO)" does not impose order among the children of the node. In applications using protein sequences, the ability of expressing the order is paramount. Generally protein sequences are a collection of chains of sequence of residues, and that is the format protein sequences have been represented unit now using various data representations and data mining techniques for bioinformatics. When we are defining sequences for semantic heterogeneity of protein data sources using PO we are not only considering traditional representation of protein sequences but also link protein sequences to protein structure, by linking chains of residue sequences to atoms defining three-dimensional structure. In this section we will describe how we used a special semantic relationship like *Sequence(s)* in protein ontology

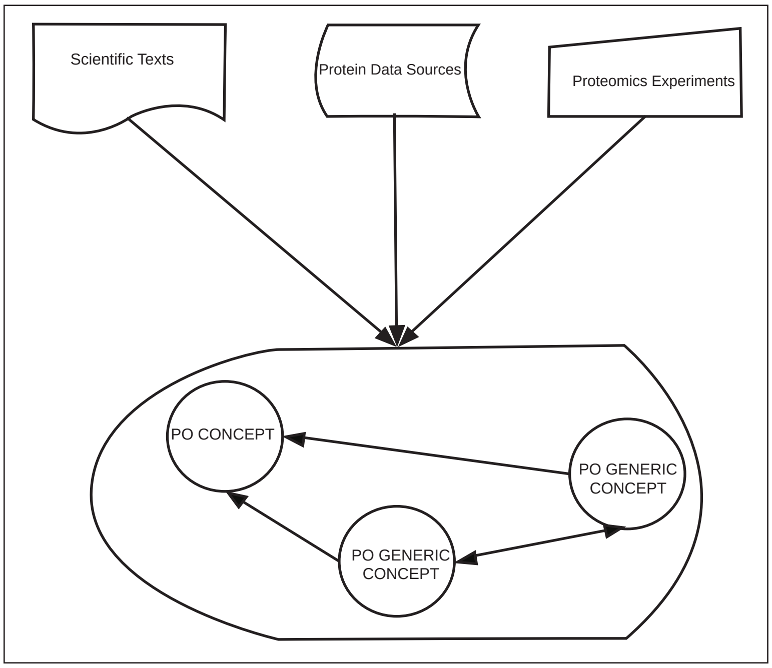
Figure 2. Semantic interoperability framework for PO



to describe complex concepts defining structure, structural folds and domains and chemical bonds describing protein complexes. PO defines these complex concepts as *Sequences* of simpler generic concepts defined in PO. These simple concepts are *Sequences* of object and data type properties defining them. A typical example of *Sequence* is as follows. PO defines a complex concept of *ATOMSequence* describing three dimensional structure of protein complex as a combination of simple concepts of *Chains*, *Residues*, and *Atoms* as: *ATOMSequence Sequence (Chains Sequence (Residues Sequence (Atoms)))*. Simple concepts defining ATOMSequence are defined as: *Chains Sequence (ChainID, ChainName, ChainProperty)*; *Residues Sequence (ResidueID, ResidueName, ResidueProperty)*; and *Atoms Sequence (AtomID, Atom, ATOMResSeqNum, X, Y, Z, Occupancy, TempratureFactor, Element)*. Semantic interoperability framework used in PO is depicted Figure 2.

Therefore, PO reflects the structure and relationships of protein data sources. PO removes the constraints of potential interpretations of terms in various data sources and provides a structured vocabulary that unifies and integrates all data and knowledge sources for proteomics domain (Figure 3). There are seven subclasses of protein ontology (PO), called generic classes that are used to define complex concepts in other PO classes: *Residues*, *Chains*, *Atoms*, *Family*, *AtomicBind*, *Bind*, and *SiteGroup*. Concepts from these generic classes are reused in various other PO classes for definition of class specific concepts. Details and properties of residues in a protein sequence are defined by instances of *Residues* class. Instances of chains of residues are defined in *Chains* class. All the three dimensional structure data of protein atoms is represented as instances of *Atoms* class. Defining *Chains*, *Residues* and *Atoms* as individual classes has the benefit that any special properties or changes affecting a particular chain, residue and atom can be easily added. Protein *Family* class

Figure 3. Unification of protein data and knowledge



represents protein super family and family details of proteins. Data about binding atoms in chemical bonds like hydrogen bond, residue links, and salt bridges is entered into ontology as an instance of *AtomicBind* Class. Similarly the data about binding residues in chemical bonds like disulphide bonds and CIS peptides is entered into ontology as an instance of *Bind* class. All data related to site groups of the active binding sites of proteins is defined as instances of *SiteGroup* class. In PO the notions classification, reasoning, and consistency are applied by defining new concepts or classes from defined generic concepts or classes. The concepts derived from generic concepts are placed precisely into class hierarchy of protein ontology to completely represent information defining a protein complex.

As such PO can be used to support automatic semantic interpretation of data and knowledge sources, thus providing a basis for sophisticated mining of information.

CLUSTERING FACILITATED BY DOMAIN ONTOLOGIES

In this section we demonstrate how to modify clustering algorithms in order to utilise the structure of ontology. In the section “Challenges with Clustering Data Enriched with Ontological Information” we present the differences between clustering data items with associated ontological information compared to clustering data items without this information. In “A Distance Function for Clustering Ontologically Enriched Data,” we show how these differences must be met in a clustering algorithm. Finally, “Automatic Cluster Identification and Naming” describes an automatic method of naming and describing the clusters found with the domain ontology.

Challenges with Clustering Data Enriched with Ontological Information

Many algorithms exist for clustering data (Duda, Hart, & Stork, 2001; Theodoridis & Koutroumbas, 1999). However, one of the primary decisions to be made when applying cluster analysis to data, and before choosing a specific algorithm, is the way of measuring distances between data items. Generally this involves defining some distance or similarity measure between data items defined in terms of their attributes.

Just as many clustering algorithms have been defined over a wide variety of data types, so to has a large set of potential similarity and distance functions been devised for comparing data items (Theodoridis & Koutroumbas, 1999). In general, a similarity function measures the degree to which two items are similar to one another. Conversely, a distance function measures how two data items are dissimilar. The choice of distance function for data items is often orthogonal to the particular clustering algorithm used as many clustering algorithms take as input a distance matrix, which contains the results of applying a distance function to each combination of data items. The distance matrix is a square symmetric matrix with each cell *i*, *j* measuring the distance between data items *i* and *j*. The particular distance function used with data items is generally dependent on the type of data being compared. For example, the distance between vectors of real valued data is often defined with the Euclidean distance function, whereas more elaborate functions are required for the sequence data types often found in biomedical datasets.

Thus, the first question we must address when devising a distance function for data enriched with information from ontologies is: what form does the data take? Details will, of course, depend on the particular ontology applied to the data. However, we can make some general comments and apply them in an example of comparing genes based on the associated gene ontology terms. In this example, the "knowledge-poor" or raw data items consist solely of gene names, for example, AA458965 or AA490846, using the GenBank accession codes. These gene names are essentially class labels with no knowledge embedded in them. Hence, there is no useful way to compare them on their own. Ontological information from the gene ontology may be associated with each gene by using the gene ontology database or with the use of a search engine such as SOURCE (Diehn et al., 2003). In our example, the gene ontology associations are shown in Table 1.

Two characteristics regarding the enriched data are apparent. First, there are different numbers of terms from the gene ontology associated with each gene. For the first gene there are four terms whilst the second has six associations. In general, this will be the norm for associations. Second, we do not seem to have accomplished much from the data enrichment. The associated terms can still be regarded as individual class labels for a very large number of classes (more than 16,000). The terms only have meaning in their relationships within the ontology hierarchy.

Thus, algorithms to cluster ontology enriched data items (1) must be able to handle different numbers of terms associated with data items; and (2) must be able to compare terms based on relationships in the ontology.

A Distance Function for Clustering Ontologically Enriched Data

Given the requirements for the clustering of ontologically enriched data developed in the last section, what kind of similarity measure or distance measure is appropriate? Standard measures like Euclidean distance are not applicable because the data contains different numbers of attributes and there is no natural way to define a distance between classes.

One possible approach is suggested by an analogy to comparison of documents in the field of computational linguistics. A common approach in this field is to transform a free form document into a sparse vector of word counts where each position in the vector refers to a different word in the corpus (see, e.g., Chapter 10 of Shawe-Taylor & Cristianini, 2004). This simplified knowledge representation of the text document ignores relationships between words. In the same way that this representation views a document as a vector of word counts, the ontologically enriched data items may be thought of as a vector of occurrences of gene ontology terms. We could devise a long sparse binary vector with each position referring to the presence or absence of an association with each of the thousands of gene ontology terms to the data item. The problem with this knowledge representation is that most of the gene ontology terms apply to only a very few genes in the database. This means that very few similarities could be found between the vectors for different data items. The solution to this difficulty lies in incorporating the relationships within the ontology into the knowledge representation.

Referring back to the example of the two genes in the last section, there is another characteristic of the enriched data that are not, at first, apparent. We can retrieve further enriched data for the genes by tracing back up the gene ontology

Table 1. Enriched data. First column lists "knowledge-poor" data in the form of GenBank identifiers. Second column lists associated Gene Ontology term identifiers for each gene.

AA458965	GO:0005125, GO:0005615, GO:0006955, GO:0007155
AA490846	GO:0004872, GO:0005515, GO:0007160, GO:0007229, GO:0008305, GO:0016021

Table 2. Enriched data with background associations: First column lists "knowledge-poor" data in the form of GenBank identifiers, rows of second column show gene ontology terms at successive distances from the directly associated terms

AA458965	0: GO:0005125, GO:0005615, GO:0006955, GO:0007155
	1: GO:0005102, GO:0006952, GO:0007154, GO:0050874
	2: GO:0004871, GO:0005488, GO:0007582, GO:0009607, GO:0009987
	3: GO:0050896, 2 x GO:0003674, 2 x GO:0008150
	4: GO:0007582
	5: GO:0008150
AA490846	0: GO:0004872, GO:0005515, GO:0007160, GO:0007229, GO:0008305, GO:0016021
	1: GO:0004871, GO:0005488, GO:0007155, GO:0007166, GO:0043235
	2: 2 x GO:0003674, GO:0007154, GO:0007165, GO:0043234
	3: GO:0007154, GO:0005575, GO:0009987
	4: GO:0009987, GO:0008150
	5: GO:0008150

Note: Some terms are seen multiple times at the same distance or at further distances

hierarchy. In the gene ontology, parent terms are more general concepts of child terms. For example, for the gene AA458965 the term GO:0006952 (defense response) can be derived from the term GO:0006955 (immune response) by following the is-a relationship in the ontology. This allows us to retrieve more general terms describing the genes. These more general terms give a sort of background knowledge for the genes. As we trace back terms higher in the hierarchy we successively build up more general background knowledge for the genes. The complete set of associations for the genes in our example is shown in Table 2. It should be clear that the terms associated with the genes differ in importance. Terms that are lower in the hierarchy are more specific to the data items and should be treated as more significant for comparisons between data items. Conversely, terms that are far from the original terms (in terms of distance up the hierarchy) are more general and should play a less significant role in comparison of data items. Furthermore, different child terms may have the same parent term or terms. This means that as we trace back up the ontology hierarchy we may draw in the same term more than once. Consequently, the background knowledge of terms may, and usually will, have duplicated terms.

This observation suggests a method of applying a similar knowledge representation to that used in the field of computational linguistics. Rather than using a binary vector to represent the presence or absence of an association between data item and gene ontology term, we use a real value measure or weighting of the degree of significance of the term to the data item. Terms directly associated with each data item, for example those listed in Table 1, receive a weight value of 1, terms indirectly associated with the data item (i.e., higher in the hierarchy) are given a lower weighting and terms that cannot be reached from terms associated with the data item are assigned 0. This leads to a less sparse vector where comparisons may be made.

A straightforward method of deriving the distance between terms using a weighting scheme like this is to adapt a similarity measure called the Tanimoto measure (Theodoridis & Koutroumbas, 1999). The Tanimoto measure defines a measure of similarity between sets:

$$\frac{n_{X \cap Y}}{n_X + n_Y - n_{X \cap Y}} = \frac{n_{X \cap Y}}{n_{X \cup Y}}$$

where X and Y are the two sets being compared and n_X, n_Y and $n_{X \cap Y}$ are the number of elements in the sets X, Y and $X \cap Y$ respectively.

However, in the current situation, the "sets" being compared are the gene ontology terms for the two genes. As there may be duplicated terms in the lists associated with each data item we adapt the Tanimoto measure to give similarities between bags rather than sets.

Also, as the terms higher in the ontology are less significant in terms of comparison than the more specific terms towards the bottom, we weight the contribution of terms by the distance from the descendent gene ontology term directly associated with the gene. In effect, this results in a "weighted" cardinality of the bag of gene ontology terms. Furthermore, as we are interested in a distance rather than a similarity, we subtract the similarity from 1. The final distance function used, then, is:

$$D_{X,Y} = 1 - \frac{n'_{X \cap Y}}{n'_X + n'_Y - n'_{X \cap Y}} = 1 - \frac{n'_{X \cap Y}}{n'_{X \cup Y}}$$

where X and Y are the bags of terms being compared and n'_X, n'_Y and $n'_{X \cap Y}$ are the weighted cardinalities of the bags X, Y and $X \cap Y$ respectively given by:

$$n'_X = \sum_{i \in X} c^{d_i}$$

where X is the bag of gene ontology terms, d_i is the distance of term of X with index i from its associated descendent in the original set of gene ontology terms for the gene, and c is the weight constant. The weighted cardinality of the other bags is similarly defined.

The more general gene ontology terms provide a context for the understanding of the lower level terms directly associated with genes. The c weight constant allows variation of the importance of the "context" to the comparison. A value of $c = 0$ means that higher level are ignored. A value of 1 considers all terms equally irrespective of their position in the hierarchy and regards the very general terms as overly significant. The c parameter may be viewed as a sort of "constant of gravity" for the clusters. The higher the value of c, the more that distantly related genes are gathered into a cluster. A choice of $c = 0.9$ gives reasonable results.

A similar graph-based approach for determining similarity based on gene ontology relationships to our described above is given in Lee, Hur, and Kim (2004). That approach involves transformation of the gene ontology from a directed acyclic graph into a tree structure and encoding of gene ontology accession codes to map into the tree.

Our similarity function contains several assumptions about ontologies. It treats distances between levels in the ontology as the same. This means that terms that are the same distance away from the terms directly associated to data items have the same effect on the similarity measure. This may not necessarily reflect the knowledge encoded in the ontology. The level of fan-out from a parent to child in the ontology may be an indication of the concentration of knowledge in the ontology. For example, when the fan-out from parent to child is large, this may indicate that the parent concept has been investigated more or is understood better than parents with less fan-out. This and other measures could conceivably be incorporated into the similarity function.

Automatic Cluster Identification and Naming

Once clusters have been identified, the ontology can facilitate inference of cluster descriptions. The descriptions say how data items in the cluster are similar to one another and different to other clusters using the vocabulary of the ontology.

Cluster descriptions are inferred for each cluster using the method shown in the pseudo code in Figure 4. At lines 1 and 2 the algorithm starts with an empty set of definitions and a list of all terms directly associated with the data items in the given cluster. The ontology hierarchy is traversed upwards replacing terms with their parent (more general) terms. Terms are replaced (line 12) only if the parent term is *not* associated with data items in another cluster (or is one of any of the ancestor terms in another cluster). At lines 8 and 14 the algorithm chooses a term to be added to the class description. Line 8 is the case when the top of the hierarchy is reached and line 14 is the case when no parent terms could be found that referred only to the cluster of interest. The output of the algorithm is a list of terms for a cluster that describe in the most general way possible the data items in the cluster (but not so general that it describes another cluster).

Insight into structure within clusters can be gained by examining which data items are associated with terms in the cluster description. It can happen that a subset of the data items in a cluster may have a description that is more concise than the description for all the data items in the cluster. This may be an indication of poor clustering of the data items.

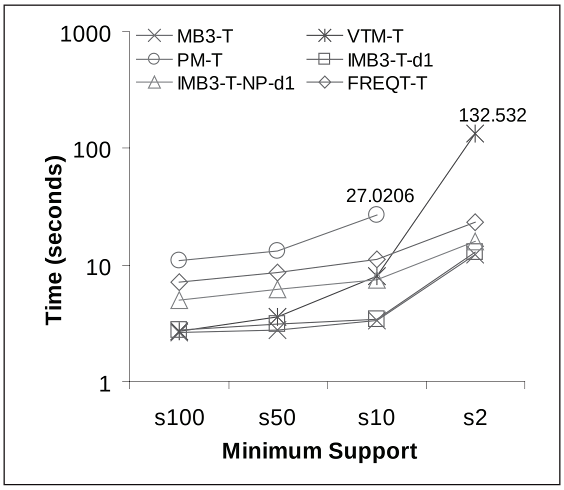
CASE STUDY

This section presents a case study of enriching bio-medical data with the protein ontology. The case study discusses the results of six data mining algorithms on PO data. The protein ontology database is created as an instance store for various protein data using the PO format. PO provides technical and scientific infrastructure to allow evidence based description and analysis of relationships between proteins. PO uses data sources like PDB, SCOP, OMIM and various published scientific literature to gather protein data. PO database is represented using OWL. PO database at the moment contains data instances of following protein families: (1) prion proteins, (2) B.Subtilis, (3) CLIC and (4) PTEN. More protein data instances will be added as PO is more de-

Figure 4. Pseudo code for cluster identification and naming

```
1    definitions = { }
2    working = terms directly associated with the data items in the cluster
3    while there are terms in working
4        new_working = { }
5        for each term in working
6            parents = parent terms of term
7            if there are no parents
8                add term to definitions
9            else
10               for each parent_term in parents
11                   if parent_term is associated only with this cluster
12                       add parent_term to new_working
13                   else
14                       add term to definitions
15       working = new_working
16   end while
17   definitions is the set of terms describing the cluster.
```

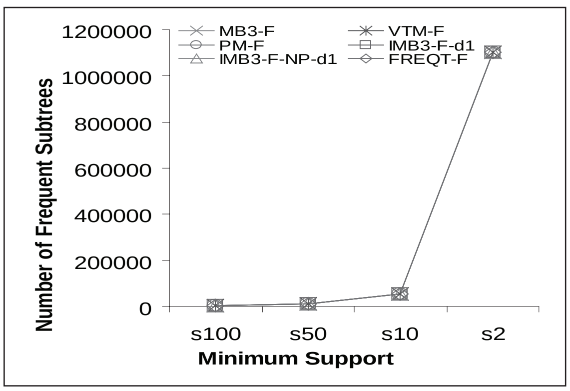
Figure 5. Time performance for prion dataset of PO data



veloped. The PO instance store at moment covers various species of proteins from bacterial and plant proteins to human proteins. Such a generic representation using PO shows the strength of PO format representation.

We used some standard hierarchical and tree mining algorithms (Tan & Dillon, in press) on the PO database. We compared MB3-Miner (MB3), X3-Miner (X3), VTreeMiner (VTM) and PatternMatcher (PM) for mining embedded subtrees and IMB3-Miner (IMB3), FREQT (FT) for mining induced subtrees of PO data. In these experiments we are mining prion proteins dataset

Figure 6. Number of frequent subtrees for prion dataset of PO data



described using protein ontology framework, represented in OWL. For this dataset we map the OWL tags to integer indexes. The maximum height is 1. In this case all candidate subtrees generated by all algorithms would be induced subtrees. Figure 5 shows the time performance of different algorithms. Our original MB3 has the best time performance for this data.

Quite interestingly, with prion dataset of PO the number of frequent candidate subtrees generated is identical for all algorithms (Figure 6). Another observation is that when support is less than 10, PM aborts and VTM performs poorly. The rationale for this could be because the utilized join approach enumerates additional invalid subtrees. Note that original MB3 is faster than IMB3 due to additional checks performed to restrict the level of embedding.

CONCLUSION

We discussed the two challenges to incorporate this biological knowledge into the data mining cycle: generating the ontologies, and adapting the data mining algorithms to make use of the ontologies. We present protein ontology (PO) framework, discuss semantic interoperability relationships between its concepts, and compare its structure with gene ontology (GO). We also demonstrate how to modify clustering algorithms in order to utilize the structure of GO. The results of six data mining algorithms on PO data are discussed, showing the strength of PO in enriching data for effective analysis.

REFERENCES

Diehn, M., Sherlock, G., Binkley, G., Jin, H., Matese, J. C., Hernandez-Boussard, T., et al. (2003). SOURCE: A unified genomic resource of functional annotations, ontologies, and gene expression data. *Nucleic Acids Research, 31*(1), 219-223.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (2nd ed). New York: John Wiley & Sons.

GO Consortium. (2001). Creating the gene ontology resource: Design and implementation. *Genome Research, 11*, 1425-1433.

GO Consortium & S. E. Lewis. (2004). Gene ontology: Looking backwards and forwards. *Genome Biology, 6*(1), 103.1-103.4.

Lee, S. G., Hur, J. U., & Kim, Y. S. (2004). A graph-theoretic modeling on GO space for biological interpretation of gene clusters. *Bioinformatics, 20*(3), 381-388.

Nelson, S. J., Schopen, M., et al. (2004, September 7-11).The MeSH translation maintenance system: Structure, interface design, and implementation. In M. Fieschi (Ed.), *Proceedings of the 11th World Congress on Medical Informatics*, San Francisco (pp. 67-69).

Shawe-Taylor, J., & Cristianini, N. (2004). *Kernel methods for pattern analysis*. Cambridge: Cambridge University Press.

Sidhu, A. S., & Dillon, T. S. (2006a). Protein ontology: Data integration using protein ontology. In Z. Ma & J. Y. Chen (Eds.), *Database modeling in biology: Practices and challenges*. New York: Springer Science, Inc.

Sidhu, A. S., & Dillon, T. S. (2006b). *Protein ontology project: 2006 Updates*. Invited paper presented at Data Mining and Information Engineering 2006, Prague, Czech Republic. WIT Press.

Sidhu, A. S., & Dillon, T. S. (2005a). Ontological foundation for protein data models. In *Proceedings of the First IFIP WG 2.12 & WG 12.4 International Workshop on Web Semantics (SWWS 2005) in conjunction with On The Move Federated Conferences (OTM 2005)*, Agia Napa, Cyprus (LNCS). Springer-Verlag.

Sidhu, A. S., & Dillon, T. S (2005a). Protein ontology: Vocabulary for protein data. In *Proceedings of the 3rd IEEE International Conference on Information Technology and Applications (IEEE ICITA 2005)*, Sydney, Australia (pp. 465-469).

Tan, H., & Dillon. T.S. (in press). IMB3-miner: Mining induced/embedded subtrees by constraining the level of embedding. In *Proceedings of PAKDD 2006*.

Theodoridis, S., & Koutroumbas, K. (1999). *Pattern recognition*. San Diego: Academic Press.

This work was previously published in Knowledge Discovery and Data Mining: Challenges and Realities, edited by X. Zhu and I. Davidson, pp. 189-201, copyright 2007 by Information Science Publishing (an imprint of IGI Global).

Chapter 7.19
A Haplotype Analysis System for Genes Discovery of Common Diseases

Takashi Kido
HuBit Genomix, Inc., Japan

ABSTRACT

This chapter introduces computational methods for detecting complex disease loci with haplotype analysis. It argues that the haplotype analysis, which plays a major role in the study of population genetics, can be computationally modeled and systematically implemented as a means for detecting causative genes of complex diseases. In this chapter, the author provides a review of issues on haplotype analysis and proposes the analysis system which integrates a comprehensive spectrum of functions on haplotype analysis for supporting disease association studies. The explanation of the system and some real examples of the haplotype analysis will not only provide researchers with better understanding of current theory and practice of genetic association studies, but also present a computational perspective on the gene discovery research for the common diseases.

INTRODUCTION

In recent years, much attention has been focusing on finding causative genes for common diseases in human genetics. (Badano & Katsanis, 2002; Daly, 2001; Fan & Knapp, 2003; Gabriel et al., 2002) These findings of causative genes of common diseases including diabetes, hypertension, heart disease, cancer, and mental illness are expected to be opening doors for realizing new diagnoses and drug discoveries. A promising approach for the gene discovery on common diseases is to statistically examine genetic association between the risk of common diseases and DNA variations in human populations. While single nucleotide polymorphisms (SNPs), the most common genetic variation, are widely used for this genetic association study, haplotypes, the combination of closely linked SNPs on a chromosome, has been shown to have pivotal roles in the study of the genetic basis of disease (Clark, 2004; Niu, 2004; Schaid, 2004). The main purpose of this report is to provide a comprehensive review of haplotype analysis in genetic association studies on complex, common diseases and provide the computational framework which enables us to carry out successful high-throughput genome-wide association study.

In addition to the review of the recent developments of haplotype analysis, the author presents the design, implementation, and application of a haplotype analysis system for supporting genome-wide association study. While there are some useful tools or programs available for haplotype analysis (Kitamura et al., 2002; Niu, Zin, Zu, & Liu, 2002; Sham & Curtis, 1995; Stephens, Smith, & Donnelly, 2001), little work has been reported for a comprehensive analysis pipleline for large-scale and high-throughput SNPs screening which fully integrate these functions. HAPSCORE (Zhang, Rowe, Struewing, & Buetow, 2002) is one of the few examples of those pipeline systems; however, it does not include some analysis functions such as automatic linkage disequilibrium (LD) block partitioning and disease association analysis tools. In this report, the author presents a system, LDMiner (Higashi et al., 2003), which represents the pioneer pipeline system that integrates a comprehensive spectrum of functions related to haplotype analysis. This report introduces the details of LDMiner and shows some examples of haplotype analysis with LDMiner, which helps to explain the theory and practice on population-based association study for common diseases.

BACKGROUND

Genetic Variations and Common Diseases

The progress on human genome science is opening doors for the discovery of new diagnostics, preventive strategies, and drug therapies for common complex diseases including diabetes, hypertension, heart disease, cancer, and mental illness. Analysis of human genome primarily focuses on variations in the human DNA sequence, since these differences can affect the potential risk of disease outbreaks or the effectiveness of a drug treatment of the diseases.

A common method for determining the genetic differences between individuals is to find single nucleotide polymorphisms (SNPs). A SNP is defined as a DNA sequence variation referring to an alteration of a single nucleotide (A, T, G, C). SNPs represent the most common genetic variations. In fact, there are millions of SNPs in the human genome (Kruglyak & Nickerson, 2001), and it is estimated that there will be on average one SNP every 1,000 base pairs. SNPs are caused when nucleotides replicate imperfectly or mutate. Although most of these SNPs have no ostentatious impacts on the survival of the species, certain SNPs may confer beneficial effects allowing species to evolve and to adapt to new environments more successfully, while certain others may be detrimental. These SNPs are passed on from generation to generation. After hundreds of years, some SNPs become established in the population.

A number of instances are known for which a particular nucleotide at a SNP locus (i.e., a particular SNP allele) is associated with an individual's propensity to develop a disease. For example, it has been reported that functional SNPs in the lymphotoxin-alpha gene were associated with susceptibility to myocardial infection by means of a large-scale case-control association study using 92,788 SNP markers (Ozaki et al., 2002). There have also been a number of reports that show some SNPs in certain genes can determine whether a drug can treat a disease more effectively in individual with certain genotypes compared to those who do not carry such SNPs. For example, Cummins et al. (2004) reported that there is a strong association in Han Chinese between a genetic marker, the human leukocyte antigen HLA-B*1502, and Stevens-Johnson syndrome induced by carbamazepine, a drug commonly prescribed for the treatment of seizures. These SNPs are often linked to the causative genes, but may not be themselves proved to be causative. These are often called surrogate markers for the disease. These disease-associated SNPs are expected to be validated by a number of investigations including biological experiments and epidemiological studies in the next several years.

Why Study Haplotypes?

A haplotype is a set of closely linked alleles (SNPs) inherited as a unit. For example, let's assume that there is a gene containing three SNPs. We now represent two alleles of each SNP with A and B (A is minor allele). The number of possible combinations of the three SNPs is eight (i.e. A-A-A, A-A-B, A-B-A, A-B-B, B-A-A, B-A-B, B-B-A, B-B-B). However, the total number of common (i.e. frequency > 5%) haplotypes in the population usually converge to a number less than eight, for example, three (A-A-A, A-B-A, B-B-B). Also, the number of common haplotypes varies depending on the chromosome regions. The recent empirical findings in human genetic studies show that genomic sequence is comprised of parts; cold spots (LD blocks) having much less variations than other regions and hot spots having more variations (Cardon & Abecasis, 2003). Some recombination hot spots tend to include a large number of haplotypes, while cold spots (or LD blocks) include only small number of haplotypes.

Haplotypes play important rules for searching the causative sites associated with several common diseases. As Clark (2004) discussed the role of haplotypes in candidate gene studies, there are three primary reasons for considering the haplotype configuration, which are listed as follows.

- **Biological function:** Haplotypes may be defined as functional units of genes; the protein products of the candidate genes may depend on particular combinations of amino acid. For example, ApoE is a protein whose function is influenced by a pair of polymorphic amino acids (Fullerton et al., 2000). The two-site haplotypes best describe the functional differences of carious ApoE protein isoforms rather than those of individual SNPs in the gene.
- **Tracing past evolutionary history:** The variation in populations is inherently structured into haplotypes. The haplotypes cover the informative small segments on ancient ancestral chromosome that may harbor a disease locus. Therefore the information on haplotype structure is useful for localizing and for tracing the past historical events on causative sites. For example, we can trace the evolutionary events by classifying haplotypes into clusters on the basis of the idea of population genetic theory, which gives us a useful insight on the origin of the disease. (Kido et al., 2003; Templeton, 1995)
- **Statistical power advantage:** The haplotype-based analysis, which combines the information of adjacent SNPs into multilocus haplotypes, may increase statistical power of conventional SNP-based analysis to detect disease associated sites. Bader (2001) showed that haplotype-based association tests can have greater power than SNP-based association tests in the case when the disease locus has multiple disease-causing alleles.

Linkage Disequilibrium and Haplotype Analysis

The case-control study, which compares a group of people with disease to a similar group of people without disease, is commonly used in population-based association study design. If we have a strong hypothesis about the special causative alleles, causal alleles can be directly evaluated. For common complex diseases, however, strong hypotheses about the specific causal alleles are generally not available, so the "indirect" association approach has become widespread. In "indirect" association studies, we expect that the actual causative allele is not genotyped but might be located near another marker that is genotyped.

LD is an important concept for indirect association study, which is defined as the lack of independence among different polymorphisms. That is, alleles A and B are said to be in LD if the frequency of pairs of alleles AB is not equal to the product of the frequencies of A and B alleles. In LD analysis, we assume that causative alleles at tightly linked markers will remain in significant LD for extended periods of time from ancestors. Various statistical metrics can be used to summarize the extent of pairwise LD between two linked markers, such as D' and r^2.

One of the important questions for indirect association studies is "how much LD is needed to detect a disease allele using its nearby genetic markers?" Krina and Cardon (2004) discussed the complex interplay among factors that influence allelic association. They outlined the four parameters that affect an odds ratio test of association with a single SNP: (1) the odds ratio of true disease-causing SNP, (2) LD between markers and the causal SNP, (3) the marker allele frequency, and (4) the disease allele frequency. Ideally a disease allele can be detected provided that the extent of LD between the marker and the causative allele is high and the marker allele frequency matches with the allele frequency of the disease allele.

Since haplotype analysis captures regional LD information of SNPs, characterization of LD patterns is important. Characterization of LD patterns across the human genome is at present an area of highly active research. International HapMap project gathers information on LD structure of variation in human populations (The International HapMap Consortium, 2003). This will facilitate genome-wide association analysis and the search for the genetic determinations of complex diseases.

Computational Algorithms for Haplotype Analysis

Defining LD Blocks

The recent finding in human genetics shows that human genome is broken up into regions of weak LD (called hot spots) and strong LD (called cold spots or LD blocks). Strong LD within a genomic region implies that most of the variation of that region can be captured by just a few informative SNPs, called tag SNPs. It is hoped that the regions of strong LD will facilitate the discovery of genes related to common diseases through association studies.

There are a variety of proposed empiric definitions for haplotype blocks. For example, Gabriel et al. (2002) developed a descriptive approach to comparing block boundaries across different populations. They examined adjacent pairs of SNPs in at least two populations, and labeled SNP pairs as concordant or discordant, based on the strength of linkage disequilibrium between two populations. They then calculated the percentage of concordant pairs among all pairs to define the block boundaries. Liu et al. (2004) propose several similarity measures to compare haplotype boundaries across populations. They found that haplotype block boundaries vary among populations and the definition of haplotype blocks is highly dependent on the density of SNPs and the method used to find blocks. The understanding of the similarities and differences of haplotype blocks between different populations will be important for future genome association studies.

Inferring Haplotype Phases

In most cases, haplotypes are not read directly when multiple human SNPs are genotyped, but must be inferred from unphased genotype data. Although a spectrum of molecular haplotyping methods such as single-molecule dilution, long-range allele-specific PCR were developed, these methods are not widely used because of their low-throughput performance and important technical problems remain unresolved. Therefore a number of algorithms for inferring haplotypes from unphased genotype data have been developed.

Niu et al. (2004) provides a good review of recent computational methods of population-based haplotype inference methods. Clark's algorithm is based on the principle of maximum parsimony which resolves the haplotypes by starting with identifying all unambiguous haplotypes (all homozygotes and single-site heterozygotes) and repeatedly adding a new haplotype to the resolved haplotype. (Clark, 1990) Although Clark's algorithm is a relatively straight-forward procedure, the algorithm does not give unique solutions, because the phasing results are dependent on the order of genotypes that need to be phased. Expectation-maximization (EM) algorithm estimates population haplotype probabilities based on maximum likelihood, finding the values of the haplotype probabilities which optimize the probability of the observed data, based on the assumption of Hardy-Weinberg equilibrium (HWE). (Excoffer & Slatkin, 1995) Actually, EM algorithm is based on solid statistical theory and are quite effective as simulation studies demonstrate that its performance is not strongly affected by the departures from HWE. (Niu et al., 2004). Some disadvantages of EM algorithm, however, are that the iteration may lead to locally optimal point when there are many distinct haplotypes, and it cannot handle a large number of loci. Stephens et al. (2001) proposed a coalescence-based Markov-chain Monte Carlo (MCMC) approach: a pseudo-Gibbs sampler (PGS) for reconstructing haplotypes from genotype data. This algorithm is implemented as program named PHASE. Niu et al. introduced a divide-conquer-combine algorithm: partition-ligation (PL), which can handle a large number of loci. This algorithm is implemented as programs named Haplotyper (Bayesian implementation) and PLEM (EM implementation).

Although there is an ongoing debate in regard to whether the best haplotype phasing algorithm exits, different algorithms perform differently in different populations with their own strengths and limitations. Niu et al. demonstrates that different algorithms have different degrees of sensitivity to various extends of population diversities and genotyping error rates using empirical simulations based on real genome data sets.

Statistical Advantages and Methods for Haplotype-Based Disease Association

Haplotype analyses can sometimes provide greater power than single-marker analyses for genetic disease associations, due to the ancestral structure captured in the distribution f haplotypes. Whether haplotype analyses outperform single locus analyses varies depending on the assumptions about the number of disease causing SNPs, the amount of LD among SNPs from the marker and disease causing SNPs. Schaid (2004) reviews the recent works on the statistical advantages of haplotypes. Bader (2001) reports that when the set of measured SNPs includes causative SNPs, single-locus tests are more powerful than haplotype-based tests when the number of causative SNPs is less than the number of haplotypes. In contrast, haplotype analyses can be more powerful than single-locus analyses when the SNPs are in LD with a causative diallelic locus. (Akey & Xiong, 2001) Both single-locus analyses and haplotype analyses lose power when there are multiple alleles at a causative locus, but comparatively speaking haplotype analyses lose less power. In this situation, the power advantage for haplotype analyses is greatest when the marker alleles are not in strong LD with each other, yet in strong LD with the causative alleles.

Schaid provides a comprehensive review on how to evaluate the association of haplotypes with human traits. When haplotypes are directly observed, then statistical methods can be applied to compare the frequencies of haplotypes between cases and controls. If the haplotype- phases are unknown, haplotype frequencies for the cases and controls can be estimated by EM algorithm. One of the common methods for haplotype-based disease association is to use likelihood ratio statistic test (LRT) which tests the equality of haplotype frequencies between cases and controls (Zhao & Sham, 2002). Although LRT is a practical method, one of the disadvantages of LRT is that it may not be adequate when there are many haplotypes, it lacks adjustment for environmental covariates. In order to solve this problem, Schaid introduces regression models for haplotypes in which haplotypes are treated as categorical covariates with other environmental covariates. Schaid also reviews recent developments of statistical regression methods for haplotype analyses.

A HAPLOTYPE SYSTEM ANALYSIS

LDMiner: The Pipeline System for SNP Analysis

We have developed a system named LDMiner that supports disease association studies to detect genes that may cause complex diseases (Higashi et al., 2003). The main functions of LDMiner are as follows:

1. LDMiner is a total analysis system for SNPs, LD blocks haplotypes, and disease association.
2. LDMiner automatically combines SNPs located in the vicinity with respect to each other on a chromosomal region and constructs LD block, and accordingly haplotypes within the block using the EM algorithm (Kitamura et al., 2002).
3. LDMiner automatically assesses the association of haplotypes and diplotypes between cases and controls. By combining SNPs together into LD blocks, we can improve the statistical power for association study.
4. LDMiner visualizes the analyzed results along with a genome viewer. The main viewer displays the genomic structure and is linked to another main viewer showing the in-depth analysis result. These viewers allow the user to easily check and make interpretations of the results.

LD analysis and some advanced association studies can be efficiently performed using LDMiner with handy tools for eliminating the inadequate data and so on. Consequently, the number of SNPs the system can analyze is about 30 to 50 times higher than by the standard manual procedures per unit of time (Higashi et al., 2003).

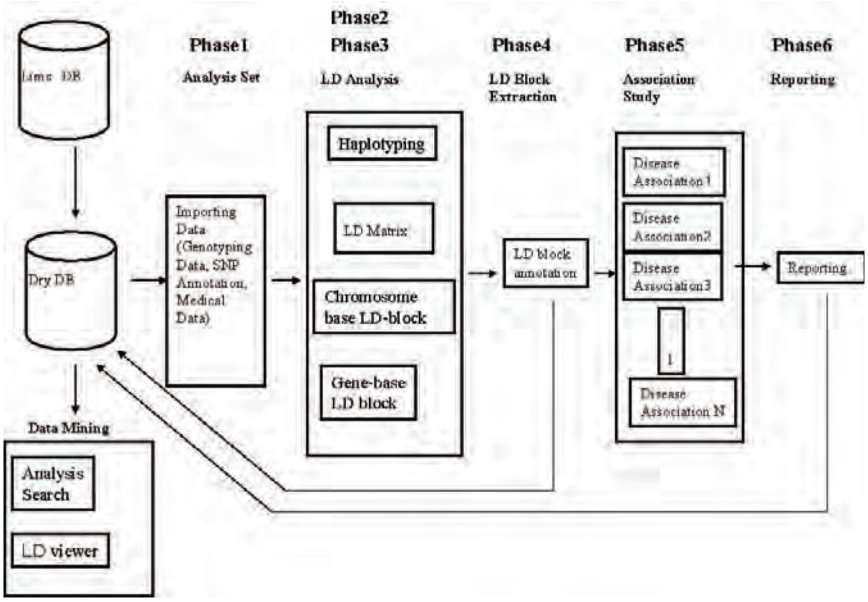
The system design of the LDMiner analysis pipeline is shown in Figure 1. In phase 1, we import genotype and medical data with SNP annotation for the analysis. In phases 2 and 3, we examine the extents of LDs among SNPs within the same genes or the same chromosomal regions as well as to estimate haplotypes and diplotypes using EM algorithm. In phase 4, LD blocks are automatically constructed by a sliding-window approach in searching for strong LD regions as discretized LDblocks. In phase 5, we perform several association tests for either individual SNPs or haplotypes within LD blocks. In phase 6, all association results are automatically annotated with SNPs and LD blocks information in order to easily retrieve and compare. These analysis results are visualized by linking the genome browser which maps LD status into genome structure.

Algorithms

Haplotyping

Given the genotypes of the individuals in a sample, the frequencies of the haplotypes in the population can be estimated. There are primarily three categories of algorithms for the inference of haplotype phases of individual genotype data;

Figure 1. Analysis pipeline with LDMiner

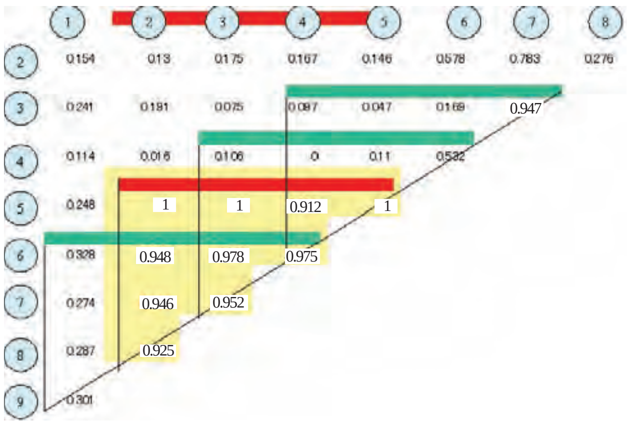


these categories are Clark's algorithm (Clark, 1990), the EM algorithm (Kitamura et al., 2002) and Bayesian algorithm (Stephens et al., 2001). Niu et al. reports the performance comparison of these algorithms and conclude that EM algorithm is shown to be accurate and practical. Furthermore they proposed PL algorithm, which is an extended version of a stochastic sampling algorithm incorporating the idea of "divide-conquer-combine", implemented the program called HAPLOTYPER which is applicable to a large number of SNPs. Although our system can incorporate several haplotyping programs as module, we mainly use EM algorithm implemented as LDSupport programs by Kamatani et al. (2002) and then determine the posterior probability distribution of diplotype configuration (diplotype distribution).

Automatic LD Block Extraction

We define the LD block based on the idea that SNPs within each LD block are tightly linked, which means that most of the pairwise LD measurement of two SNPs within the LD block is high, for example D' ≥ 0.8 Because our major purpose of LD block extraction is to improve the performance of haplotype association analysis, the number of SNPs in the LD block should be limited in order to keep down the degree of freedom of the association study. For example, based on our experience, three to five SNP markers are appropriate for our association study when the majority of the pairwise LD values across all SNPs within the same LD block is strong enough. Otherwise, if we select too many SNPs for the LD block, the power of the multi-locus association test becomes smaller because of the inflation of the number of haplotypes as well as the fact that the posterior probability of the diplotype determination becomes smaller. For these reasons, we propose the window-sliding method for LD block construction where the maximum number of SNPs in the LD block is always lower than predefined window size. One of the examples of

Figure 2. Window-sliding method for LD block extraction



the window-sliding method is shown in Figure 2. In this example, SNP2, 3, 4, 5 were extracted to be within a single LD block because all pairwise LDs among the four SNPs is strong enough (D' > 0.9). Although D' is the quite popular measure of LD, appropriate LD measures and parameters should be selected depending on the situation. LDminer provides a variety of LD measures and parameters for flexibly defining the LD block. We implement this method for both of gene-base and chromosome-base LD block extraction.

Disease Association for Multi-Locus Haplotypes

After building LD blocks, our system automatically performs both single-locus tests and haplotype /displotype association tests. In haplotype / diplotype association tests, we have options to select several multi-locus association tests such as Log Likelihood Test (LRT) (Zhao & Sham, 2002), Monte Carlo tests. (Sham & Curtis, 1995)

LRT is the nonparametric tests for homogeneity in haplotype / diplotype frequencies between cases and controls. For example, the fast EH program (FEHP) developed by Zhao and Sham can be used for this purpose. In their statistics, the maximum log-likelihood was calculated for cases alone, for controls alone, and for cases and controls pooled altogether. Denoting these maximum log-likelihoods as, and, LRT statistic was defined as 2(+ -), which is asymptotically chi-squared with the degree of freedom of, where is the number of haplotypes / diplotypes. Then we obtain asymptotic p-values as well as empirical p-values by the use of permutation procedures.

In Monte Carlo tests, some programs are well known such as CLUMP by Sham and Curtis (1995), which outputs the results of four different association tests from T1 to T4. In order to estimate the odds ratio (OR: a measure of the degree of association which estimates the relative risk of the disease in case-control study) and 95%CI for each haplotype/diplotype, we perform the Fisher's exact test for each haplotype by constructing a 2 × 2 table which consists of the number of haplotype holders /others in the column and the number of case/control in the row.

Viewer for Genome Structure and Tools for Analysis Interpretation

LDMiner also has a supplicated vitalization tool (Higashi et al., 2003). The main viewer displays the genomic structure and is linked to another main viewer showing the in-depth analysis result. As real causative polymorphisms are likely to be located in the vicinity of the analyzed polymorphisms rather than themselves, it is important to examine the surrounding genes and polymorphisms. For this purpose, we also proposed the algorithm on haplotype classification for detecting disease-associated sites (Kido et al., 2003). These viewers and tools allow the user to easily check and make an interpretation of the results.

Figure 3. Selected SNPs and detected LD block for Caucasian, African American and Japanese

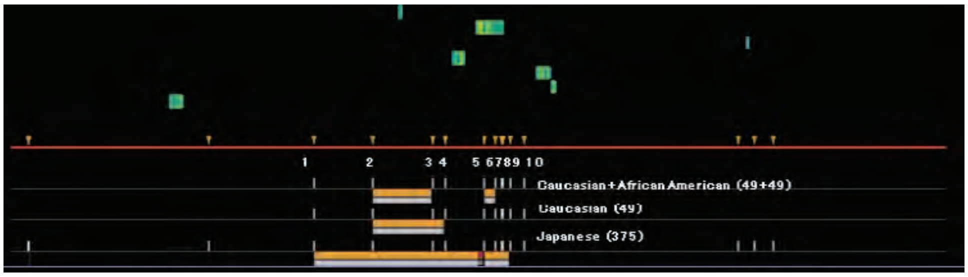


Figure 4. P-value distribution of the single-locus analysis and LD mapping

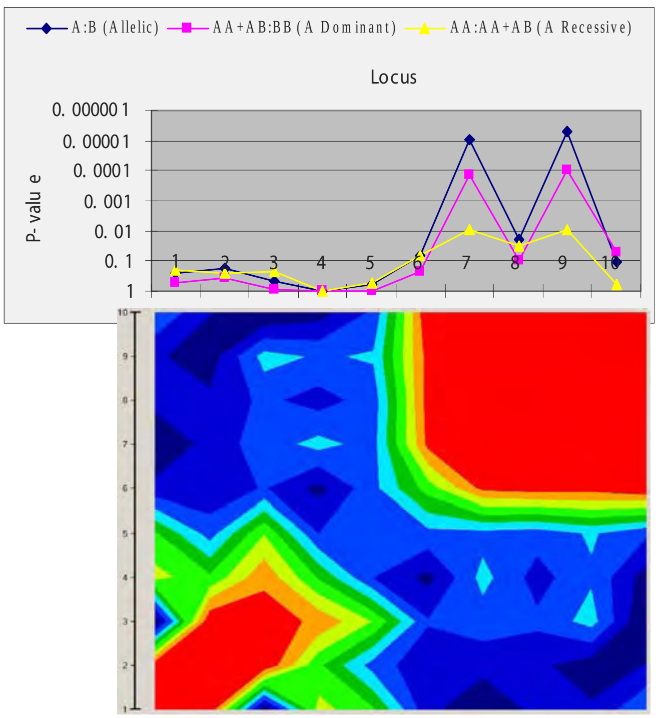
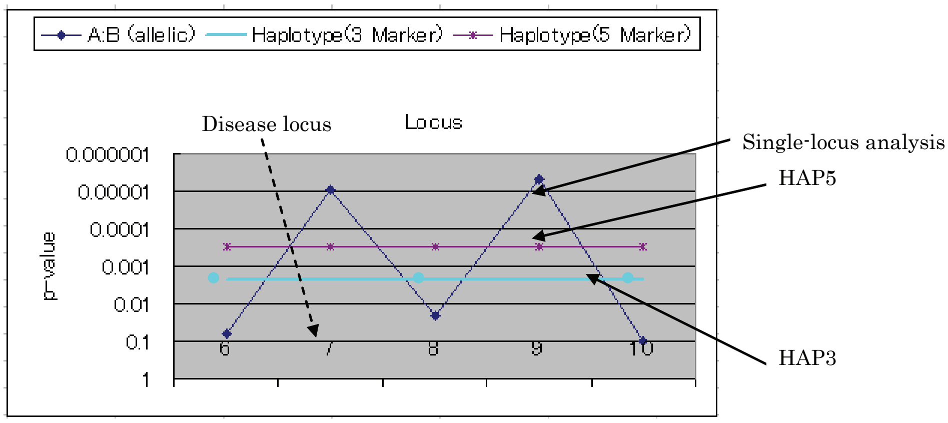


Figure 5. P-value distribution of the haplotype analysis and comparison with those of single-locus analysis



EXAMPLES OF HAPLOTYPE ANALYSES

Experimental Setting: Genotype Data Used

To evaluate our methods, we performed a simulation study based on observed SNPs from genes around NPHS1 in chromosome 19 regions with a Japanese population of 375 independent individuals, a Caucasian population of 49 individuals and an African-American population of 49 individuals. This enables us to base our study on real patterns of LD within those genes. Some genes were genotyped in both exonic and intronic regions. We chose a subset of 10 SNPs mainly from NPHS1 genes denoted by 1 to 10 as shown in Figure 3.

Experimental Setting for the Simulation Study for Association Study

In order to examine whether the multi-locus association tests of LD0blocks have more statistical power than that of single-locus marker and how LD status influences the power of association, we evaluated it with simulated data sets. We designed a disease model for typical common disease described in report by Jannot, Essioux, Reese, & Clerget-Darpoux (2003) and phenotypic trait distributions were simulated under the samples of Japanese 375 (case 148, control 227) individuals, keeping our SNP typing information. In this model, we assumed that SNP 7, which allele frequency is 0.068, is the causative with prevalence rate of 0.08. We then compared the multi-locus haplotype tests with single-locus tests.

Single-Locus Analysis

The results of the single-locus analysis of allelic and genotype association are shown in Figure 4. As we designed, the causative SNP7 shows significant difference between case and control allele frequencies with a p-value is less than 0.00001. SNP9 also has a low p-value in allelic association. This is because LD between SNP7 and SNP9 is strongly high (|D'|=0.978) and the discrepancy between SNP7 and SNP9 frequencies is very small (SNP7 minor allele frequency [MAF]=0.068, and SNP9 MAF = 0.064). LD theory introduced by Zondervan and Cardon (2004) estimates that the expected OR of locus 9 is 2.06 and this fits with our simulation result. However, the SNP6, SNP8, SNP10 are also in the strong LD with locus, their p-values are not significant. This can be explained by the discrepancy between the disease SNP (locus 7) and marker SNP frequency is not small. The MAF of SNP6 is 0.135, that of SNP8 is 0.120 and that of SNP10 is 0.056. This case shows that common variant marker which has high allele frequency can not detect the effect of row variant causative SNP even though they are in the same LD block.

In LD mapping, the red part shows the strong LD (D' is higher than 0.8) and the blue part shows weak LD. Read parts can be extracted as LD block.

Haplotype Analysis

Selecting SNPs for Haplotype Analysis

First we selected 5 SNPs - SNP6, SNP7 (causative), SNP8, SNP9, SNP10 that are in the strong LD block as show in Figure 4. The average of |D'| in this LD block is 0.964. We then select 3 SNPs - SNP6, SNP8 SNP10. This is because, minor allele frequencies of SNP7 and SNP9 are small (less than 0.1) and they are usually considered as inappropriate SNP markers. Here, an interesting question for us is whether haplotype analysis consisting of three SNPs can detect the causative region or not, although all three SNPs are not positive in single-locus analysis.

Results of Haplotype Association: Single-Locus Analysis vs. Haplotype Analysis

The p-value distributions of the haplotype association and comparisons with those of single-locus analysis are shown in Figure 5. The HAP5 consisting of five SNPs has a significant overall p-value by FEHP program; empirical p-value is less than 0.001. However it is not higher than those of single-locus analysis at SNP7 and SNP9. Interestingly, HAP3 consisting of three SNPs has also significant overall p-value, which is less than 0.01. (The empirical p-value by FEHP program with 20000 permutations is 0.0022, and the asymptotic p-value is 0.0007. In Figure 3, we plot the empirical p-value of FEHP.) Furthermore this p-value is higher than those of single-locus analysis at SNP6, SNP8 and SNP10. This is a good example to show the advantage of haplotype analysis; although three SNPs in HAP3 are not positive in single-locus analysis, haplotype analysis can detect the significant difference in the HAP3 block.

Table 1 shows the results of haplotype analysis with three SNP markers (HAP3). Only six haplotypes are estimated by EM algorithm, major haplotype 'BBB' dominates 85.8 % over the population. For the major haplotype 'BBB', the difference of frequencies between case and control is not significant with a p-value obtained by Fisher's exact test of 0.085. The frequency of the minor haplotype 'AAB' - 6.5 %, is significant (p-value is less than 0.0001). It works as a risk haplotype, such that its frequency among cases (11%) is much higher than in controls (3%). The Fisher's p-value of 'AAB' is less than 0.0001 and the OR is 3.75 and its 95% confidence interval is 1.99 – 7.03. Because the posterior probability for determining diplotype is almost 100% (represented by blue bar) or more than 99% (represented by green bar) in most haplotypes, we can validate that haplotype estimation is mostly reasonable.

Diplotype Analysis

We also perform three-locus diplotype association as well as haplotype association study. This is done by determining individual's diplotype after estimating haplotype by EM algorithm (Kitamura et al., 2002). We only choose them from which posterior probability is more than 80%, otherwise we rule out from our analysis. The results are

Table 1. The results of haplotype analysis with 3 SNP markers

Table1:Haplotype Infomation

Haplotype ID	Sequence	Control Case	Frequency in case/control	Posterior probability distribution	Frequency	2x2 Fisher's p-value	Odds ratio	Odds ratio 95% confidence interval
0	Missing	0%(1) 0%(2)			0	0.0	0.0	null
1	Haplotype[BBB]	87%(400) 82%(243)			0.838	0.083636	0.694286	0.458361 - 1.051644
2	Haplotype[AAB]	3%(15) 11%(33)			0.065	2.2E-5	3.745946	1.996388 - 7.028751
3	Haplotype[AAA]	6%(29) 3%(10)			0.052	0.091979	0.522133	0.250532 - 1.088175
4	Haplotype[ABB]	1%(9) 1%(4)			0.017	0.775473	0.689815	0.210468 - 2.260884
5	Haplotype[ABA]	0%(2) 0%(1)			0.004	1.0	0.780069	0.070413 - 8.642005
6	Haplotype[BAB]	0%(1) 0%(1)			0.003	1.0	1.563574	0.097414 - 25.096648

Table 2. The result of diplotype analysis with 3 SNP markers

Table2:Diplotype Infomation

Diplotype ID	Control Case	Frequency in control/case	Posterior probability distribution	Frequency	2x2 Fisher's p-value	Odds ratio	Odds ratio 95% onfidence interval
Missing	0%(0) 0%(1)				null	null	null
Diplotype[1:1]	77%(176) 71%(105)				0.270774	0.756652	0.470371 - 1.217172
Diplotype[1:2]	6%(15) 15%(23)				0.005039	2.655285	1.335409 - 5.279683
Diplotype[1:3]	10%(25) 4%(6)				0.020673	0.348000	0.139146 - 0.870335
Diplotype[1:4]	2%(6) 2%(3)				1.000000	0.776224	0.191075 - 3.153330
Diplotype[1:5]	0%(1) 0%(0)				1.000000	0.517634	0.020944 - 12.793105
Diplotype[1:6]	0%(1) 0%(1)				1.000000	1.565517	0.097154 - 25.226514
Diplotype[2:2]	0%(0) 3%(5)				0.008690	17.763251	0.974761 - 323.702877
Diplotype[3:3]	0%(2) 1%(2)				0.645199	1.569444	0.218627 - 11.266481
Diplotype[4:4]	0%(1) 0%(0)				1.000000	0.517634	0.020944 - 12.793105
Diplotype[4:5]	0%(1) 0%(1)				1.000000	1.565517	0.097154 - 25.226514

shown in Table 2. Minor diplotype [1:2] which represents the combination of haplotype1 ('BBB') and haplotype2 ('AAB') works as risk diplotype (Fisher's p-value is 0.005, OR=2.65). The diplotype also works as risk diplotype in our analysis (Fisher's p-value is 0.008, OR=17), although it is in the rare cases.

SUMMARY

The main purpose of this chapter was to review the roles of haplotype analysis for association study on complex common diseases and provide the computational framework that enables to realize high throughput genome-wide association study. The motivation, background and computational methods for finding causative genes by use of haplotype and LD information have been reviewed. A Haplotype Analysis System, LDMiner, has been introduced with some real examples of association studies. Although there have been significant advances in this area, further developments of computational methods for the genome-wide association studies and the detection of complex gene-gene interactions have been pursued. (Meng, Zaykin, Zu, Wanger, & Ehm, 2003; Ritchie et al., 2001; Toivonen et al., 2000) It is hoped that this report will entice further research and developments in bioinformatics for finding causative genes.

ACKNOWLEDGMENT

We would like to thank Dr. Tianhua Niu at Harvard University for useful discussions and comments.

REFERENCES

Akey J., & Xiong, M. (2001). Haplotype vs single marker linkage disequilibrium tests: What do we gain? *Eur J Hu Genet, 9,* 291-300.

Badano, J. L, & Katsanis, N. (2002). Beyond Mendel: An evolving view of human genetic disease transmission. *Nature Genetics*, *3*, 779-789.

Bader, J. S. (2001). The relative power of SNPs and haplotype as genetic markers for association tests. *Pharmacogenomics*, *2*(1), 11-24

Cardon, L. R, & Abecasis, G.R. (2003). Using haplotype blocks to map human complex trait loci. *Trends Genet., 19,* 135-140.

Clark, A. G. (1990). Inference of haplotypes from PCR-amplified samples of diploid populations. *Mol. Biol. Evol.*, *7*, 111-122.

Clark, A. G. (2004). The role of haplotypes in candidate gene studies. *Genetic Epidemiology, 27*, 321-333.

Cummins, J. M., Rago, C., Kohli, M., Kinzer, K. W., Lengauer, C., & Vogelstein, B. (2004). A marker for Stevens-Hohnson Syndrome. *Nature, 428*,486.

Excoffier, L., & Slatkin, M. (1995). Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.*, *12*, 921-927.

Fan, R., & Knapp, M. (2003). Genome association studies of complex diseases by case-control designs. *Am .J .Hum. Genet.*, *72*,850-868.

Fullerton, S. M., Clark A. G, Weiss, K. M, Nickerson, D. A., Taylor, S. L., Stengard, J. H., et al. (2000). Apolipoprotein E variation at the sequence haplotype level: Implications for the origin and maintenance of a major human polymorphism. *Am. J. Hum. Genet.*, *67*, 881-900.

Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., et al. (2002). The structure of haplotype blocks in the human genome. *Science*, *296*, (pp. 2225-2229).

Higashi, Y., Higuchi, H., Kido, T., Matsumine, H., Baba, M., Morimoto, T., et al. (2003). SNP analysis system for detecting complex disease associated sites. *The Proceeding of IEEE Computational Systems Bioinformatics* (CSB2003), 450-451.

Hung, S. L., Chung, W. H., Liou, L. B., Cgy, C. C., Lin, M., Huang, H. P., et al. (2005). HLA-B*5801 allele as a genetic marker for severe cutaneous adverse reactions caused by allopurinol. *PNAS*, *102*,11, 4134-4139.

International HapMap Project. (n.d.). Retrieved from http://www.hapmap.org/

The International HapMap Consortium. (2003) The International HapMap Project. *Nature*, *18*, 426(6968),789-96.

Jannot, A. S., Essioux, L., Reese, M. G., & Clerget-Darpoux, F. (2003). Improved use of SNP information to detect the role of genes. *Genetic Epidemiology, 25*, 158-167.

Kido, T., Baba, M., Matsumine, H., Higashi, Y., Higuchi, H., & Muramatsu, M. (2003). Haplotype pattern mining & classification for detecting disease associated site. *The proceeding of IEEE Computational Systems Bioinfomratics* (CSB2003), (pp. 452-453).

Kitamura, Y., Moriguchi, M., Kaneko, H., Morisaki, H., Morisaki, T., Toyama, K., & Kamatani, N. (2002). Determination of probability distribution of diplotype configuration (diplotype distribution) for each subject from genotypic data using the EM algorithm. *Annals of Human Genetics*, *66*, 183-193.

Kruglyak, L., & Nickerson, D. A. (2001 March). Variation is the spice of life. *Nature Genetics.*, *27*(3), 234-6.

Liu, N., Sawyer, S. L., Mukherjee, N., Pakstis, A. J., Kidd, J. R., Kidd, K. K., et al. (2004). Haplotype block structures show significant variation among populations. *Genetic Epidemiology*, *27*, 385-400.

Meng, Z., Zaykin, D. V., Zu, C. F., Wanger, M., & Ehm, M. G. (2003). Selection of genetic markers for association analysis, using linkage disequilibrium and haplotypes. *Am. J .Hum. Genet.*, *73*, 115-130.

Niu, T. (2004). Algorithms for Inferring Haplotypes. *Genetic Epidemiology*, *27*, 334-347.

Niu, T., Zin, Z. S., Zu, X., & Liu, J. S. (2002). Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms, *Am. J. Hum. Genet.*, *70*, 157-169.

Ritchie, M. D, Hahn, L. W., Roodi, N., Dupont, W. D., Parl, F. F., & Moore, J. H. (2001). Multifactor dimensionality reduction reveals high-order interactions among estrogen metabolism genes in sporadic breast cancer. Annals of Human Genetics, 69, 138-147.

Schaid, D. J. (2004). Evaluating associations of haplotypes with traits. Genetic Epidemiology, 27, 348-364.

Sham, P. C., & Curtis, D. (1995). Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. Annals of Human Genetics, 59, 97-105.

Stephens, M., Smith, N. J., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. Am. J. Hum . Genet., 68, 978-989.

Toivonen, H. T. Onkamo, P., Vasko, K., Ollikainen, V., Sevon, P., Mannila, H., Herr, M., & Kere, J. (2000). Data mining applied to linkage disequilibrium mapping. Am. J. Hum. Genet., 67, 133-145.

Zhang, J., Rowe, W. L., Struewing, J. P., & Buetow, K. H. (2002). HapScope: A software system for automated and visual analysis of functionally annotated the allelic association. Nature Genetics, 5, 89-100.

Zhao, J. H. (2004). Haplotype block structures show significant variation among populations. Genetic Epidemiology, 27, 385-400.

Zhao, J. H., & Sham, P.C. (2002). Faster haplotype frequency estimation using unrelated subjects. Hum Hered., 53, 36-41.

This work was previously published in Advanced Data Mining Technologies in Bioinformatics, edited by H. Hsu, pp. 214-230, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.20
Computational Fluid Dynamics and Neural Network for Modeling and Simulations of Medical Devices

Yos S. Morsi
Swinburne University, Australia

Subrat Das
Swinburne University, Australia

ABSTRACT

This chapter describes the utilization of computational fluid dynamics (CFD) with neural network (NN) for analysis of medical devices. First, the concept of mathematical modeling and its use for solving engineering problems is presented followed by an introduction to CFD with a brief summary of the numerical techniques currently available. A brief introduction to the standard optimization strategies for NN and the various methodologies in use are also presented. A case study of the design and optimization of scaffolds for tissue engineering heart valve using the combined CFD and NN approach is presented and discussed. This chapter concludes with a discussion of the advantages and disadvantages of the combined NN and CFD techniques and their future potential prospective.

MATHEMATICAL MODELING

Introduction

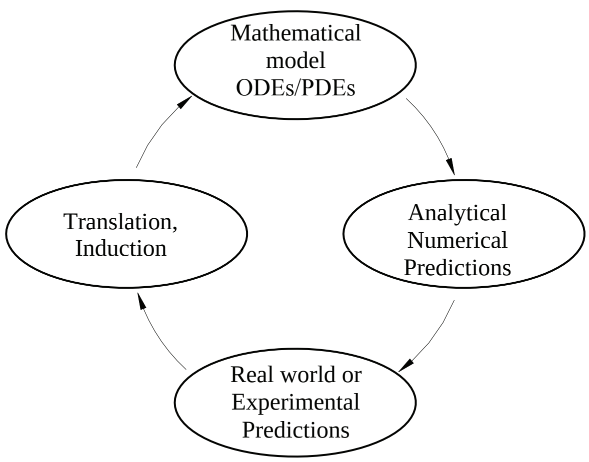
Many engineering systems are of a complex nature and require techniques that relate the relevant variables in the system under consideration. Equations that express physical phenomena between quantities require absolute numerical and dimensional equality. Historically, the use of dimensional analysis of the physics observed experimentally has been very successful in adding to our understanding of the complexity of the problem in hand. Generally speaking, all physical relationships can be expressed in terms of quantities such as mass M, Length L, and time T or other related quantities such as force F, pressure P, stress τ, and so on. The application of such a system may include converting one system of units to another, developing relations or equations, reducing the number of variables required for an experimental program and, in some cases, determining the principles of model design. It should be noted that for a physical system, dimensional analysis can only indicate variables or groups of variables that are functionally related, and it does not give insight of the nature of the correlation and its complexity. One of the most common uses of dimensional analysis is in experimental planning of examining a particular phenomena or system. Moreover, the dimensional analysis does not estimate the actual behavior of the system. This requires the development of a more comprehensive mathematical model which often requires a solution to the governing equations, either ordinary differential equations or partial differential equations (ODEs/PDEs).

The Governing Equations

In recent years, the utilization of mathematical modeling to solve engineering problems has been advanced greatly by the availability of fast and user-friendly software packages. However, these models, even today, still require experimental or physical modeling results to validate and verify the numerical results. However, in general, the process of mathematical modeling may require a continuing four-stage cycle, as illustrated in Figure 1 to have a solution.

The mathematical model is constructed based on fundamental laws of physics, often by making quite a number of assumptions. Some are justified quite rigorously on scientific grounds, while others render the problem tractable or solvable within an economically acceptable time frame. In any case, the mathematical assumptions formulate a mathematical model, and the next two stages of the cycle are intended to test this model and modify it if need be (Figure 1). To test a model, one needs to draw certain conclusions about the real problem at hand. Such terminations are often of two types, those related to previously observed situations (*explanatory in nature*) or those related to new, not previously observed situations (*predictive in nature*). Both types are important for validating a mathematical model, though for purposes of discussion, it is reasonable

Figure 1. The typical loop of mathematical modeling



to refer to both types as predictions. To obtain these outcomes, one first makes mathematical predictions, using mathematical tools, which have been previously developed or are developed for a particular mathematic model. These mathematical predictions are then translated back from the language of the model to the language of the real problem and interpreted as real-world predictions or conclusions. In the final stage, the predictions are checked against real data, either old (in the case of tests of the explanatory power of the model) or new (in the case of tests of its predictive power). On the basis of the new data, which includes the performance of the model's predictions, the model is modified, and the cycle repeats. The cyclical process is continuous and newly generated data must be validated against the explanatory power of the model or against its predictive power.

Mathematical models provide a fundamentally based quantitative relationship between process variables that provide a general insight into the overall behavior of a given system. For a new process, a mathematical model can provide guidance with respect to the general feasibility and the consistency of a new concept with both the physical and theoretical laws. If the process concept is found feasible, the model may identify critical areas that require further validation through possibly experimental investigation.

Traditionally, mathematical model development has relied on individual effort, where a scientist may have worked for a very long period of time to develop customized algorithms or a computer program, often understood by the developer alone. This imposes a major difficulty in extending the theory for further development of the program. However, in the last few decades with availability of software packages, the ready access to inexpensive computers, and the much more widespread application of modeling as a technical tool has assisted greatly in the advancement of mathematical modeling packages.

A mathematical model generally consists of algebraic or differential equations that quantitatively represent a system or process. For example, it may be a relationship that defines the time, pressure, and velocity in a pipe, defines swirling flow in an annulus, or defines the circulation patterns, fluid-structure interactions, composition of two phases in chemical equilibrium, stress distributions, and pressure drops along and across a valve or within a pump.

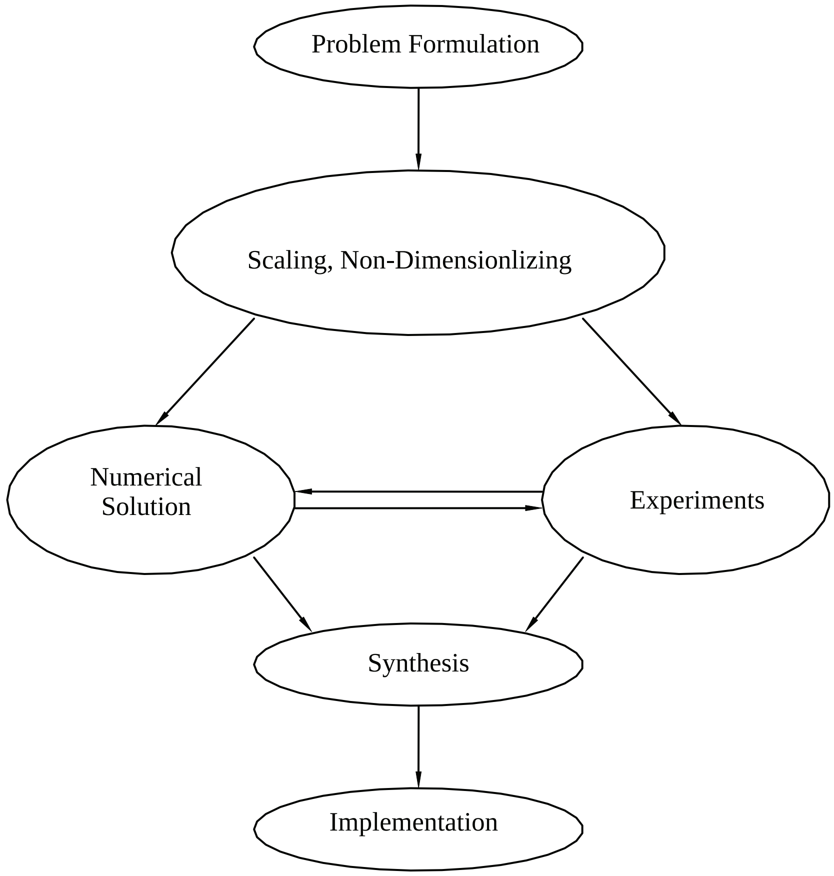
From the above discussion, it follows that a mathematic model may be simple differential equations or a complex set of equations, such as the coupled Navier Stokes equations with turbulence closures which may require certain expertise and advanced computer hardware to solve.

General Methodology for Model Development

A typical flowchart of a model development and implementation is shown in Figure 2. The model development involves the translation of a previously constructed physical picture into concrete mathematical terms, or more precisely, differential equations. Traditionally, problem formulation was usually performed "from scratch" by establishing a control volume and thus developing the appropriate governing equations in a differential form, by letting the dimensions of the control volume shrink to an infinitesimal size. While such a development is instructive and may indeed be appropriate for certain simple systems, at present there are rather easier and speedier ways to proceed:

- Drawing on prior experience with related systems (the rapidly growing body of modeling literature makes this possible);
- Using the "ready-made" building blocks of the key differential relationships;
- Navier-Stokes equations for fluid flow;
- Fourier's equation for heat transfer;
- Fick's equation for diffusion;

Figure 2. General methodology of model development



- Maxwell's equations for deformation processing turbulence models;
- Constitutive equations for deformation processing turbulence models.

It should be stressed that many problems involve some non-standard building blocks, such as meniscus effects, surface waves, hydrodynamic instability, surface tension, and so forth. The modeler will need to take care not to ignore these factors because the effects may have a dominant influence on the problem.

It has been shown that the critical examination of the governing equations represents the next logical step. These equations usually involve several scaled parameters as mentioned previously. Thus, the functional relationships between key process parameters, such as the solidification time proportional to the square of the characteristic dimension, the linear dependence of the melt velocities on current in induction furnaces, the relationship between the Rayleigh number and the characteristic velocity in buoyancy driven flows, Reynolds number effect on wall shear stress, and so forth, do provide a framework for interpreting data, which is unlikely to emerge from the computer output. Furthermore, the analytical solutions or functional relationships provide an excellent means for evaluating the consistency of computed solutions. The computer-generated solutions will make a much greater degree of precision possible to a whole range of process problems, such as the details of the velocity field, the thermal gradients, local values of the heat and mass transfer coefficients, and the details of mixing. Therefore, these two approaches are complementary rather than orthogonal.

At this point, it is important to stress the need to employ the correct boundary conditions because the software packages essentially solve the same types of conservation equations, and the principal factor that distinguishes individual cases is usually associated with the boundary conditions. These numerical solutions need to be checked for internal consistency, that is, by performing overall heat, mass, and momentum balances, testing sensitivity to different starting points and grid sizes, and, most importantly, checking the predictions against analytical asymptotes. The predictions will then have to be compared with experimental measurements to establish the discrepancies between these results by critical examination. If the deviation between these results is large, then it is necessary to check the appropriateness of the assumptions and to update the model.

Outline of the General Mathematical Procedures

The general principle of mathematical procedures, which can hold for any systems, can be outlined as follows:

- Define the quantity conserved, the control volume for which the equations are to be written, and select the independent and dependent variables. If only one independent variable appears, an ordinary differential equation will result.
- Apply the physical laws that yield the input and output expressions for the quantity conserved.
- Substitute these expressions into the appropriate terms in the equation.
- Develop the necessary boundary conditions.
- Solve the resulting differential equation and boundary conditions.

The above procedures can be applied to various engineering problems for design and optimization of various cardiovascular devices such as blood vessels, rotary blood pumps, ventricular assist devices, lung ventilation unit, modeling of liver stent, kidney and various other medical devices. In the following section, a case study is given to illustrate the necessary steps needed for design and optimization of model of artificial heart valves.

The success or failure of mathematical treatment of blood flow problems within cardiovascular devices depends mainly on our ability to solve the governing equations of continuity, momentum, and energy as accurately as possible. These equations contain non-linear terms and a large number of unknown variables, which makes an exact solution extremely difficult to obtain. Therefore, some assumptions of the physical phenomena must be made in order to reduce the number of unknown variables in the equations. In complex turbulent flow, a complete solution is still extremely difficult to obtain and, in attempting to solve the equations of motion, it is necessary to consider the following factors:

- The solution techniques available for the problem and their reliability;
- The methods of approximating the equations;
- The degree of accuracy requiring the validity of assumptions employed.

In general, the exact solutions for complex medical devices are difficult to obtain; moreover, the correctness of the results obtained analytically are somewhat limited due to their simplifying assumptions. With this in mind, numerical analyses have been widely used for several decades. As a result, numerical techniques have been developed to solve different combinations of ordinary/partial differential equations for various engineering and science applications. These methods are briefly summarized below:

Numerical Methods

Finite difference method (FDM), finite-volume method (FVM), and finite-element method (FEM) are widely employed to solve flow and heat transfer problems. Generally, the choice of a particular technique depends on factors such as non-linearity, geometry of boundary, and type of boundary conditions. FDM is a common and successful method used to deal with orthogonal geometries involving simple boundary conditions. In FDM methodology, discretization of the governing equations in 3D problems is simpler and the accuracy can readily be examined by the order of the truncation error in the Taylor's series expansion. However, for problems involving irregular geometry in the solution domain, FVM is more suitable. On the other hand, arbitrary and more complex geometries, involving gradient boundary conditions, FVM can be more cumbersome, and the application of FEM is preferred. Another important use of FEM is as a generalized computer algorithm, making it more advantageous than the other methods (Das & Morsi, 2003). However, development of a computer code that employs FEM is more time consuming compared to an equivalent code that uses FDM or FVM. It should be noted that the topic of "which method is good" is still the subject of debate among researchers. In reality, the choice of a particular method depends on the type of partial differential equations (PDEs) and the computational domain.

Partial Differential Equations (PDEs)

PDEs representing the physical problem may be classified as elliptic, parabolic, and hyperbolic and can be expressed in two-dimensional Cartesian form as

$$A\frac{\partial^2\phi}{\partial x^2}+B\frac{\partial^2\phi}{\partial x\partial y}+C\frac{\partial^2\phi}{\partial y^2}+D\frac{\partial\phi}{\partial x}+E\frac{\partial\phi}{\partial y}+F\phi+G(x,y)=0 \quad (1)$$

Where ϕ is the dependent variable that becomes velocity and temperature in fluid flow and heat transfer problems respectively? The mathematical character of the PDE is defined by the coefficients of A, B, and C. Other coefficients, D and E, can be functions of the spatial coordinates (x, y). The source term is represented by G. All fluid flow and heat transfer problems use a set of PDEs that are the same as the general equation shown above. However, in some situations of complex geometry, typical boundary conditions and the coefficient in the equation (1) can represent a formidable challenge in computational ability.

Computational fluid dynamics (CFD) and computer hardware during the last decade provide researchers in the field with various commercial codes of FVM and FEM such as CFX-TASCflow, FLOW3D, CFX4, CFX5, STAR-CD, Flotran, and so forth. They have helped significantly in various area of engineering to obtain reasonable approximations. However; other researchers have developed their own source codes for specific applications in fluid flow and heat transfer.

Recent improvements in the coupling library MpCCI (previously called COCOLIB) are taking place in cooperation between Computational Dynamics, Intes, GMD, and Daimler Chrysler. This capability will provide a general interface between a flow solver and a structural analysis code for estimating the fluid loads on a defined structure, particularly in multiphysics problems. Many have used this MpCCI interface to couple either their own codes or commercial codes such as fluid dynamics (CFX) with structure analysis (ANSYS, ADINA, ABACUS) and successfully simulated fluid induced large deflection problems. Several CAD development packages, such as Unigraphics, Solid-Edge, AutoCAD, Pro-Engineer, and 3D-CAD, have become an additional advantage for developing the accurate geometries of the computational domain. The development of CPU speed has also helped to simulate more sophisticated turbulence modeling that deals with realistic geometries. Traditionally, the results generated from these codes are validated with qualitative and quantitative information gained from experimental diagnostics techniques such as partical images velocitmatry (PIV) and laser Doppler anemometry (LDA).

In general, despite all the advances in the field, the CFD approach to some typical situations remained computationally expensive in terms of time. On the other hand, it is not suitable for system simulations, particularly for optimization purposes. Moreover, when the solution demands an interactive procedure, numerical solution usually fails to provide the solutions in real time due to its iterative procedure. The major shortcomings of CFD analysis can be listed:

- The system is usually depicted in terms of the set of ordinary differential equations or partial differential equations. The mathematical representation in terms of OPDEs/PDES usually involves approximations that are not relevant to the physical model. Moreover, due to a lack of precise knowledge of the physics of the processes involved or the properties of the materials used, the mathematical model represents an ideal situation rather than a real one.
- There may be several external *disturbances*, such as a change in environmental conditions that affect the response of the system.
- The exact initial conditions to determine the state of the system may not be accurately known.
- The model may be too complicated for exact analytical solutions. Computer-generated numerical solutions may have small errors that are magnified over time.
- The solution may be inherently sensitive to small perturbations in the state of the system, in which case, any error will magnify over time.
- Numerical solutions may be too slow to be of use in real time. This is usually the case if PDEs or a large number of ODEs are involved.

Although the CFD is now increasingly used to model thermal and fluid system performance as a part of the design and engineering process, in nearly all cases, the computational time required for CFD limits its use to providing insight into the physical phenomena investigated after the basic design is chosen rather than as a design and optimization tool. This is the case where NN can be thought of as another alternative tool where a collective set of neurons can be trained with a prior knowledge database obtained from numerical CFD calculations. Thus, it was required to define an intelligent system and its collectiveness.

SYSTEM APPROACH

A *system* may be defined as a small part of the universe that we are interested in and that has some intelligence. At this point, it is difficult to quantify the mode or the degree of intelligence that can be given to a system. However, the existing CFD/experimental data are enormous and can be one of the most important tools to create a system with some knowledge-based intelligence through various evolutionary algorithms (EAs). There are several areas where knowledge-based training approaches are applied to construct intelligent systems through evolutionary algorithms. These include instrument landing systems, automatic pilot, collision-avoidance systems, antilock brakes, smart air bags, intelligent road vehicles, medical diagnostic devices, image processing, pattern recognition, intelligent data analysis, temperature and flow control, process control, intelligent CAD, smart materials, smart manufacturing, Internet search engines, and machine translators. Neural network (NN) is one of approaches through which a knowledge-based system can be constructed.

Neurophysiologists spent many years searching for the engram, that is, the precise location in the brain for specific memories. The brain memory system has a specific structure that consists of many elements (neurons) working in parallel and in connection with one another. These sets of neurons act as a computing system for the brain. Thus, a neural network science based largely on the notion of "brain-style computation" has evolved in which a large number of very simple processing units act simultaneously on a distributed pattern of data. A trained neuron can be thought of as an expert system or a knowledge-based system in the category of information it has been given to analyze. Thus, NN that consists of several neurons has a remarkable ability to derive meaning from complicated or imprecise data that can be used to extract patterns and detect trends that are too complex to be recognized by humans or computer techniques. The application of NN to a variety of engineering applications appears to be a recent development, and a brief discussion on system development is outline below.

Simple System

A system has an input $u(t)$ and an output $y(t)$ where t is the time. A schematic representation simple open system is shown in Figure 3.

One can represent an input-output relationship by $y = \nabla(u)$ where∇ is an operator (addition, substraction, differential, or integral). Thus, if we know the input u(t), then the operations represented by ∇ must be carried out to obtain the output. In some situations, we may have a set of values for u(t) and y(t) and we would like to know what ∇ is. This is a *system identification* problem. If the system identifies the physics precisely, then it is called an expert/intelligent system.

For *closed-loop control*, there is a *feedback* from the output to the input of the system, as shown in Figure 4. The output of the NN model is compared with some standard value (base value). This base value can be from CFD simulation, experimental data, or analytical solution. The difference between the output and the base value can be treated as an error and minimized through a feedback closed loop approach.

Complex systems are made up of a large number of such opened/closed simple systems, each of which may be easy to understand or to solve for. Together, however, they pose a formidable modeling and computational task, and from a neurophysics point of view, each simple system behaves like a neuron.

NEURAL NETWORK IN CFD APPLICATION

Neural networks (NNs) have emerged as an alternative tool (Cao et al., 2004; Parlos et al., 1992) to first principle-based approaches for modeling complex systems in a variety of engineering applications. In a particular area of CFD analysis, the volume flow rates obtained by the two-dimensional CFD analysis were used as target values for learning a neural network. By learning neural network with the target values, the values of learning results were obtained, which almost coincide with the values of CFD (Han, 2003). The authors have used backpropagation rule as a learning method with hyperbolic tangent sigmoid transfer function for connecting the input layer with the first hidden layer. Even in a particular natural convection problem, u, v, p and T, Ra, Pr can be considered as inputs and Nusselt number can be considered (heat transfer rate) as an output. Yuen and Bau (1998) used the same approach for a natural convection problem. Diaz (2000) used

Figure 3. Schematic representation of an open system

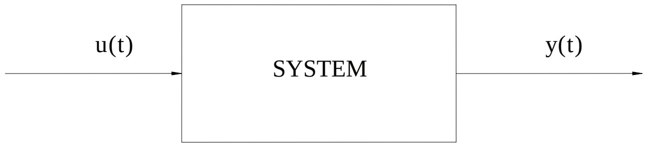
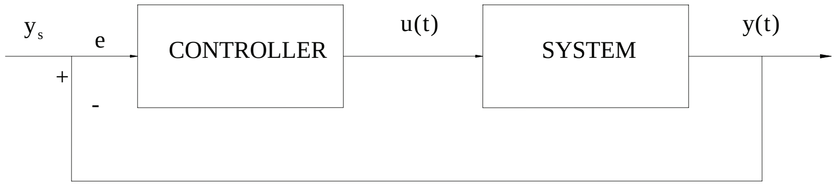


Figure 4. Closed system



NN model for heat exchanger problem. Turbulence model using NN models were also reported (Gillies, 1998; Lee et al., 1997).

NN-based fluid field estimation with dimensional inputs and output has been studied by Xue and Watton (1998). The non-linear transient behavior of fluid power elements was also identified through NN modeling based upon input-output observations only. Xu et al. (1996) applied NN to model a flow through an orifice. Schreck et al. (1994) reported a non-dimensional NN configuration to characterize the unsteady, vortex-dominated flow as it develops over the wing and splitter plate of aircraft using surface pressure measurements and flow visualization. The neural network models were shown to accurately predict both temporal and spatial variations for both the unsteady separated flow fields and the aerodynamic loads. The inputs and output of the NN model were non-dimensionalized according to previous empirical and theoretical results. More recently, Cao et al. (2004) reported a non-dimensional artificial neural network (NDANN) model for accurate predictions of flow force and flow rates under the broad operating conditions of a hydraulic valve. Because of its non-dimensional characteristics, the NDANN fluid field estimator also exhibits accurate input-output scalability, which allows the NDANN model to estimate the fluid force and flow rate even when the operating condition parameter or design geometry parameters are outside the range of the training data. Sablani et. al. (2003) developed an explicit calculation of the friction factor in pipeline flow of Bingham plastic fluids. Their NN approach involved the establishment of an explicit relationship between the Reynolds number, (Re), and the friction factor, (f), under both laminar and turbulent flow conditions. In addition, EAs have also been applied to various optimization problems in thermofluid devices including airfoils, heat exchangers, fluid-structure interaction problems, and missile nozzle inlets for high-speed flow.

The overall modeling approach in NN is a black box approach, since (in principle) the development of NN models does not require prior knowledge of the process. It is sufficient to provide the network with the set of input data and the corresponding outputs. The objective is to train a model with prior knowledge input (PKI) (Chao et al., 2001; Watson et. al., 1999). However, having some idea about such relationship may provide for the fine-tuning of the NN model. In some instances, these ideas have shown exceptional improvements in NN model performance (Porru et al., 2000; Shayya & Sablani, 1998; Sablani, 2001). It is also capable of dealing with uncertainties, noisy data, and non-linear relationships. Brasquet and Le Cloirec (2000) conducted an experimental study and measured air and water pressure drops through a layer of several textile fabrics. Their study focused on the influence of specific parameters of clothes on their dynamic behavior. Using NN, they correlated fluid properties (i.e., viscosity, density, and Reynolds number) and fabric characteristics (i.e., thickness, density, number of openings, and raw material) as input neurons with pressure drop as the output neuron. The NN model they developed predicted pressure drops that closely followed the experimental values. A hybrid neural network modeling approach (Porru et. al., 2000) was proposed for the identification of the dynamic behavior of chemical reactors. They used a multilayer, feedforward neural network to correlate bulk carbon monoxide concentration and temperature with reaction rate. Their NN model was capable of describing the system kinetics over the entire range of the investigated operating conditions whereas a conventional Langmuir–Hinshelwood rate law failed to provide the correct representation. These and other studies reported in the literature underscore the importance of NN in tackling wide range of problems. However, in most of the applications of NN in modeling, a feedforward neural network is used to get a non-linear input/output mapping. A detailed methodology in the development of NN model for cardiovascular application is outlined below.

Case Study Overview: The Use of Artificial Neural Network for the Design and Optimization of Heart Valve

One of the most important applications of this area is prosthetic heart valves, which are commonly used to replace natural heart valves and are also widely used in ventricular assist devices (VAD) in total artificial hearts (TAH). The clinical success of any valve design is based on many factors including the fluid flow phenomena, particularly *in vitro* velocity profiles, shear stresses, regurgitation, and energy losses (Baldwin & Tarbell, 1991; Chandran & Cabell, 1983; Morsi et al., 2001). Thus, the optimization of the valve leaflet or wall-stress development patterns relates various parameters (Qiong et al., 2003). Moreover, if a prosthetic heart valve is to be used, the valve-related problems such as blood cell damage, thrombus formation, calcification, and infection, as well as valve durability, need consideration. In such areas, the fluid-phase is most conveniently described with respect to a Eulerian reference frame, while a Lagrangian formulation is more appropriate for the solid phase. These two formulations are not compatible, and, as a result, the numerical complexity increases manyfold. Several numerical techniques, such as arbitrary Lagrangian Eulerian (ALE), fictitious domain/mortar element (FD//ME), and immersed boundary (IB), have been proposed recently. All these techniques solve the problem (fluid-structure) sequentially. The solution of fluid forces are obtained using conservation of law of mass and momentum equations and then the structural solution follows for each time step. In all the methods mentioned above, the deformation of the mesh poses a formidable computational task, particularly in the case of the complex geometric problems like cardiovascular application.

In the study of hemodynamics and, in particular, vascular disease, one of the most important variables is the shear stress, τ [N/m^2], at the vessel wall. Wall shear stress has considerable clinical relevance because it provides information about both the magnitude of the force that the blood exerts on the vessel wall as well as the force exerted by one fluid layer on another. Shear stress varies with flow conditions (cardiac output, heart rate, etc.) as well as with the local geometry of the vessel (curves, branches, etc.). Excessively high levels of shear stress caused, for example, by atherosclerotic lesions or artificial heart valves, may damage red blood cells (a condition called haemolysis) or the endothelium of the vessel wall. Other abnormal shear stresses, such as very low or strongly oscillatory shear stresses, may also change the biological behavior of some cells or platelets in the blood stream (leading to thrombus formation) or endothelial cells on the vessel wall. Thus, the rheology of the blood, along with the laws of conservation (basic physical laws), is important in the design and development of artificial devices.

A typical solution procedure of fully open valve is shown in Figure 5.

The mesh quality changes as the solution proceeds, as shown in Figure 6(a). Most of the time, the solution diverges due to the poor quality of mesh, and this is the case with cardiovascular problems where it needs a large deflection phenomenon. The flow dynamics are also very complex, shown in Figure 6(b), as the valve deflects from its original position.

A three dimensional mesh generation of complete tri-leaflet valve is shown in Figure 7 to give an idea of its numerical complexities (Morsi & Das, 2004). Nevertheless, few modeling approaches have been proposed using sequential weak coupling. The knowledge gained from the experimental findings also has increased our understanding in the same area.

At this point, it is impossible to carry out any numerical procedure to achieve an optimized design with the variable mentioned above. However, a neural network model can be thought of as an alternative method to deal with such optimization problems with existing numerical/experimental data. A solution can be achieved by building a model of the cardiovascular system of an individual and comparing it with the real time physiological measurements or with existing CFD simulation data. The aim is to exploit the benefits of NNs by developing an improved neural network methodology from the prior knowledge inputs for the qualitative and quantitative visualization of flow development on deformed structures.

The development of an NN model involves two basic steps. These include the generation of (or compilation of available) data required for training, the training of NN networks, and the evaluation and validation of the optimal configuration of the NN model. The procedure used for the development of the NN model is outlined below.

Step 1. Data from CFD/Experiments for Training of Artificial Neural Networks

PIV and LDA are powerful experimental techniques that enable full field velocity vectors of seeding particles suspended in a fluid to be visualized. It is effective for obtaining both qualitative and quantitative information of a flow. Particle tracking velocimetry (PTV) also uses the images of seeding particles for analysis and can give data for wide range of variables. Both these visualization data and CFD data from the numerical techniques can be used (PKI Method) to create an NN model to mimic the actual cardiovascular system for given boundary conditions.

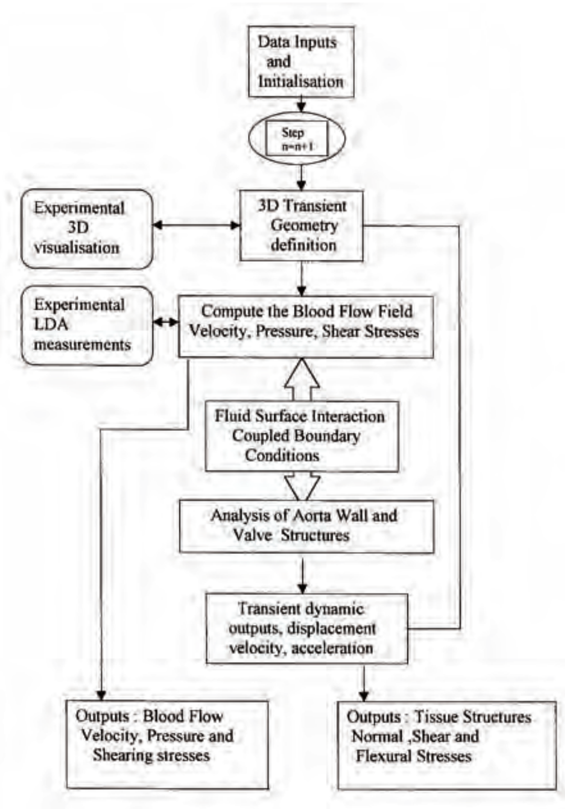
The overall objective of the proposed NN model is to devise and evaluate an explicit procedure for estimating the heart valve leaflet deflection with time. An accurate training of a system of NN model may save computational time substantially for such a complex problem as the fluid-structure interaction of the leaflet of heart valve.

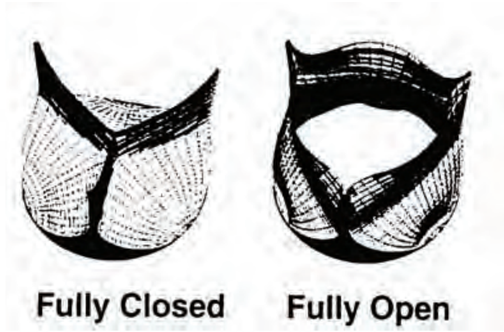
Step 2. Algorithm for Selection and Training of Neural Networks

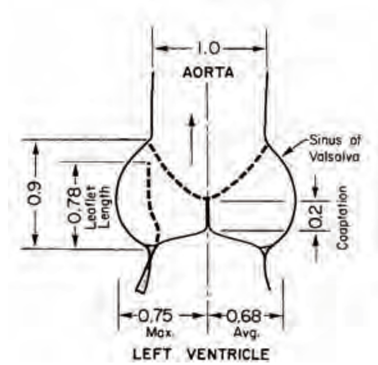
It is necessary to design an NN model to represent the desired input-output mapping for which the following considerations should be taken into account.

- *Type of system output:* stochastic or deterministic;

Figure 5. Typical tri-leaflet configurations for closed and open valve





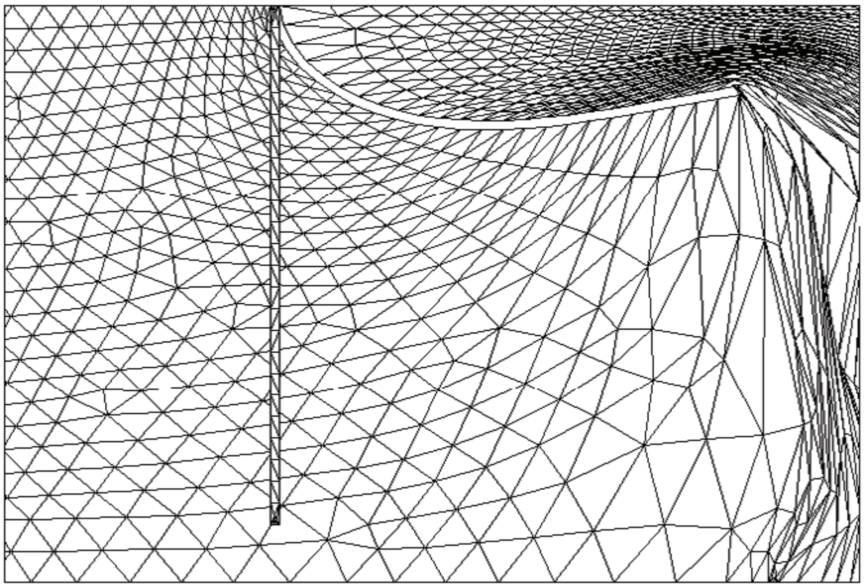


- *Type of neural network connectivity:* feed-forward, recurrent, or something else. This is a critical factor, especially when dealing with dynamical systems or with very large data quantities;
- *Layer transfer functions that depend on system behavior:* non-linear or linear;
- *Training routine:* gradient-based schemes or non-gradient-based schemes (Genetic Algorithms);
- *Scaling of input CFD data and output data:* both must be of constant order of magnitude.

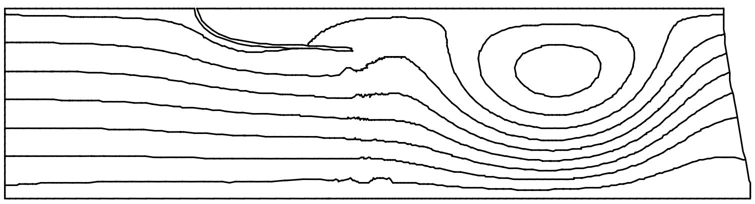
A typical NN structure trained for modeling purposes is shown in Figure 8. A scaling unit is needed to carry the system parameters and responses into a range meaningful for NN structure considered. Once training is over, the weights and bias terms of trained NNs are saved, and NN model is used for design purposes. It has to be pointed out that even though the training phase is long, using NN model takes a very short time, as there are no complex calculations.

The feedforward network structure, as a possible NN model, is shown in Figure 9. There are several ways one can select the input data depending upon objective functions (Outputs). A dataset

Figure 6. (a) Mesh deformation in 2D fluid-structure analysis; (b) streamline plot at peak flow condition



(a)



(b)

of fluid variables (u_f, v_f, P_f,) and structure varibles (u_s and v_s) at all points, along with tube diameter (D), leaflet thickness (d), Reynolds number (Re), and so forth, can be used as input parameters to an NN model with deflection (δ) as an output variable. These points cover the entire domain of solution with Re varied in a large range. It is to be noted that training the NN model with a limited dataset may not capture all the information (modes non-linearities) of the physical phenomena. The input layer consisted of five neurons that corresponded to u_f, v_f, Re, D, and d, while the output layer had one neuron representing the leaflet deflection, δ. However, selecting input data in the case of a dynamic situation of fluid flow is a crucial factor because the flow changes from laminar to turbulent as time increases. Thus, enormous amounts of data are required to train an NN model with all the non-linearities that are associated with a fluid and structural member. However, the use of surface force, that is, instantaneous pressure force (as a polynomial) on the leaflet as input and the deflection as an output may reduce the number of data for a specific optimization. Neural networks are proven to be more sophisticated in analyzing

Figure 7. Computational domain for a typical fluid-structure analysis of tri-leaflet heart valve

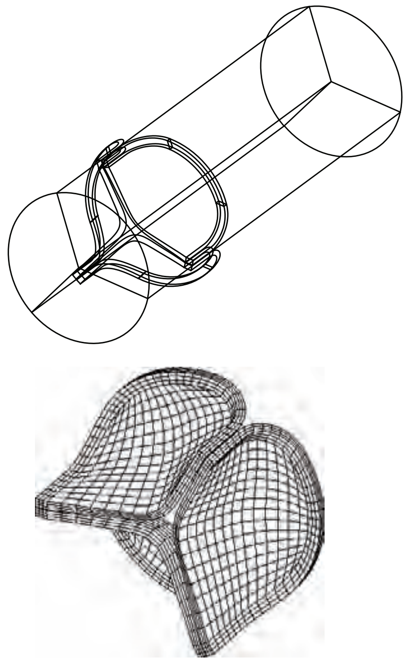
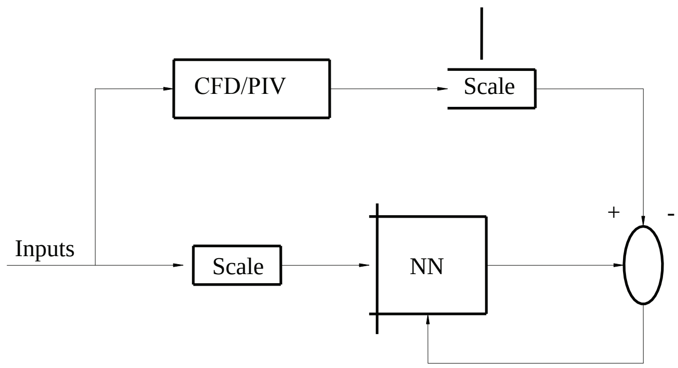


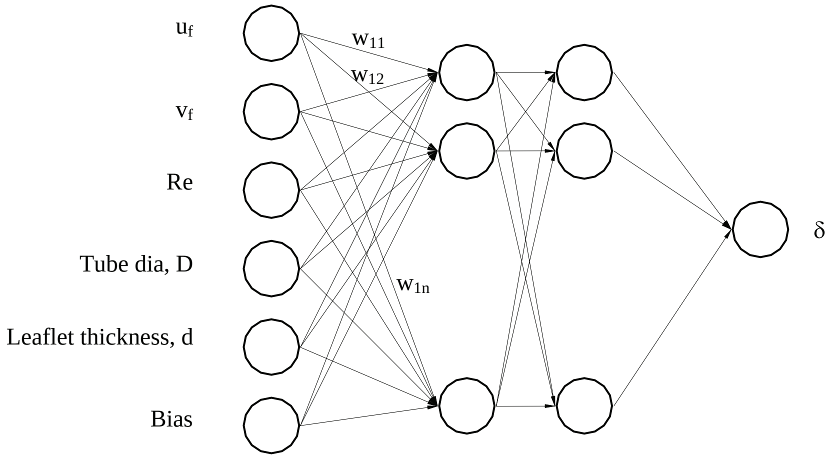
Figure 8. Neural network structure during training phase for model generating



non-linear functions operating on polynomial expansions.

Backpropagation algorithm is a well-known method (Haykin, 1999), as discussed above. It can be utilized for development and training. The building of a backpropagation network involved the specification of the number of hidden layers and the number of neurons in each hidden layer. Hidden layers, which can be more than one, are formed using neurons having continuously differentiable non-linear activation functions, while output neurons have linear activation functions.

Figure 9. A multilayer perceptron structure (MLP)



The number of neurons on the last layer is equal to the number of objective functions. However, in Figure 9, only one objective function is shown. All the neurons are connected to the neurons of the previous and next layers. In addition, several parameters including the learning rule, the transfer function, the learning coefficient ratio, the random number seed, the error minimization algorithm, and the number of learning cycles can be specified to improve the learning strategy. As the problem involves several non-linearities, that is, fluid flow and large material deformation, the accurate choice of these parameters is very difficult. However, a trial and error method can be used and for which several standard procedures are provided by any NNs software. Input data move ahead through the layers according to the connections, and, at the output layer, output of the NN is obtained. Thus, the major step is called training and consists in determining adequate values for these connection. The training is performed by minimizing the evaluation errors of the NN for the known points of the database entries. These outputs, y, are compared to desired values (base value), and weights are changed by means of backpropagation algorithm minimizing error functions depending on error, *e*, between desired and obtained outputs. This stage is crucial and often tedious, since this optimization exercise involves a large number of variables. It is usually performed through gradient-based methods and may suffer from local minimization, resulting in a weak training which provides a poor approximation. Moreover, the results may depend strongly on the number of layers and neurones, which is a real disadvantage for the present application, since the training should be performed automatically.

In general, as the number of hidden layers and the number of neurons within each hidden layer increased, the prediction capabilities of the network also increased. However, the number of connection weights increased significantly as the number of neurons and hidden layers increased, thereby increasing the chances of memorization of the behavior among the dataset (rather than generalization). Hence, the number of hidden layers and the neurons in the hidden layers need to be selected carefully.

The applications of NN models in fluid-structure problems may have to predominantly use the backpropagation network as it depends mostly on existing input and output data. Without the priory knowledge input (CFD data), NN model may be impossible to achieve. Thus, CFD still plays a major role in providing this information for NN models. However, developing the model through these data may save the computational or optimization time significantly.

CONCLUSION

In this chapter, development of an NN model with the existing CFD/experimental results is discussed particularly in the area where a multiphysics phenomena (fluid-structure) along with a complex geometry. Emphasis is given to the backpropagation technique with multilayer perceptron (MLP) because backpropagation techniques are the most suitable for training non-linear data structures. At this stage, the NN models are not widely used in CFD analyses because of the longer time involved in the training procedure with such a large non-linear data range. However, soon we will achieve a convenient mode of training procedure with evolutionary algorithms.

REFERENCES

Baldwin, J. T., & Tarbell, J. M. (1991). Mean velocities and Reynolds stresses within regurgitant jets produced by tilting disc valves. *ASAIO Transactions, 37*, 348-349.

Brasquet, C., & Le Cloirec, P. (2000). Pressure drop through textile fabrics-experimental data modeling using classical models and neural networks. *Chemical Engineering Sciences, 55*, 2767-2778.

Cao, M., Wang, K. W., & DeVries, L. (2004). Steady state hydraulic valve fluid field estimator based on non-dimensional artificial neural network. *Journal of Computing and Information Science in Engineering, 4*, 257-270.

Cao, M., Wang, K. W., Fujii, Y., & Tobler, W.E. (2004). A hybrid neural network approach for the development of friction component dynamic model. A*SME Journal of Dynamic Systems, Management and Control, 126*(1), 144-153.

Chandran K. B., & Cabell, G. N. (1983). Laser anemometry measurements of pulsatile flow past aortic valve prostheses. *Journal of Biomechanics, 16*, 865-873.

Chao, L., Jun, X., & Liangjin, X. (2001). Knowledge-based artificial neural network models for Fin line. *International Journal of Infrared and Millimeter Waves, 22*, 2.

Das, S., & Morsi, Y. (2003). Natural convection in domed porous enclosures: Non-Darcian flow. *Journal of Porous Media, 6*(3), 159-175.

Diaz, G. (2000). *Simulation and control of heat exchangers using artificial neural networks.* PhD thesis, University of Notre Dame, Department of Aerospace and Mechanical Engineering.

Gillies, E. A. (1998). Low-dimensional control of the circular cylinder wake. *Journal of Fluid Mechanics, 371*, 157-178.

Han, S. Y., & Maeng, J. S. (2003). Shape optimization of cut-off in a multi-blade fan/scroll system using neural network. *International Journal of Heat and Mass Transfer, 46*, 2833-2839.

Haykin, S. (1999). *Neural networks: A comprehensive foundation*. New York: Macmillan.

Lee, C., Kim, J., Babcock, D., & Goodman, R. (1997). Application of neural networks to turbulence control for drag reduction. *Physics of Fluids, 9*(6), 1740-1747.

Lin, Q., Morsi, Y. S., Smith, B., & Yang, W. (2004). Numerical simulation and structure verifications of Jellyfish heart valve. *International Journal of Computer Applications in Technology, 21*(1-2), 2-7.

Morsi, Y. S., Zhou, Z., & Hassan, A. (2001). A three dimensional computational investigation of turbulent flow through aortic valve. *Frontiers of Medical and Biomedical Engineering, 11*(1), 1-11.

Parlos, A. G., Atiya, A. F., Chong, K. T., & Tsai, W. K. (1992). Nonlinear identification of process dynamics using neural networks. *Nuclear Technology, 97*(1), 79-96.

Porru, G., Aragonese, C., Baratti, R., & Servida, A. (2000). Monitoring of a CO oxidation reactor through a grey model-based EKF observer. *Chemical Engineering Sciences, 55*, 331-338.

Qiong, L., Morsi, Y.S., Benjamin, S., & William, Y. (2003). Numerical simulation and structure verification of jellyfish heart valve. *International Journal of Computer Applications in Technology, 20*, 1-6.

Sablani, S. S. (2001). A neural network approach for non-iterative calculation of heat transfer coeMcient in fluid-particle systems. *Chemical Engineering and Processing, 40*, 363-369.

Sablani, S. S., Shayya, W. H., & Kacimov, A. (2003). Explicit calculation of the friction factor in pipeline flow of Bingham plastic fluids: A neural network approach. *Chemical Engineering Science, 58*, 99-106.

Schreck, S. J., & Helin, H. E. (1994). Unsteady vortex dynamics and surface pressure topologies on a finite pitching wing. *Journal of Aircraft, 31*(4), 899-907.

Shayya, W. H., & Sablani, S. S. (1998). An artificial neural network for non-iterative calculation of the friction factor in pipeline flow. *Computers and Electronics in Agriculture, 21*(3), 219-228.

Watson, P. M., Gupta, K. C., & Mahajan, R. L. (1999). Applications of knowledge-based artificial neural network modeling to microwave components. *International Journal of RF and Microwave CAE, 9*, 254-260.

Xu, P., Burton, R. T., & Sargent, C. M. (1996). Experimental identification of a flow orifice using a neural network and the conjugate gradient method. *ASME Journal of Dynamic Systems, Measurements and Control, 118*(2), 272-277.

Xue, Y., & Watton, J. (1998). Dynamics modeling of fluid power systems applying a global error descent algorithm to a self-organizing radial basis function network. *Mechatronics, 8*(7), 727-745.

Yuen, P. K., & Bau, H. H. (1998). Controlling chaotic convection using neural nets-theory and experiments. *Neural Networks, 11*(3), 557-569.

This work was previously published in Neural Networks in Healthcare: Potential and Challenges, edited by J. Kamruzzaman, R. Begg, and R. Sarker, pp. 262-283, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.21
Genome-Wide Analysis of Epistasis Using Multifactor Dimensionality Reduction:
Feature Selection and Construction in the Domain of Human Genetics

Jason H. Moore
Dartmouth Medical School, USA

ABSTRACT

Human genetics is an evolving discipline that is being driven by rapid advances in technologies that make it possible to measure enormous quantities of genetic information. An important goal of human genetics is to understand the mapping relationship between interindividual variation in DNA sequences (i.e., the genome) and variability in disease susceptibility (i.e., the phenotype). The focus of the present study is the detection and characterization of nonlinear interactions among DNA sequence variations in human populations using data mining and machine learning methods. We first review the concept difficulty and then review a multifactor dimensionality reduction (MDR) approach that was developed specifically for this domain. We then present some ideas about how to scale the MDR approach to datasets with thousands of attributes (i.e., genome-wide analysis). Finally, we end with some ideas about how nonlinear genetic models might be statistically interpreted to facilitate making biological inferences.

THE PROBLEM DOMAIN: HUMAN GENETICS

Human genetics can be broadly defined as the study of genes and their role in human biology. An important goal of human genetics is to understand the mapping relationship between interindividual variation in DNA sequences (i.e., the genome) and variability in disease susceptibility (i.e., the phenotype). Stated another way, how does one or more changes in an individual's DNA sequence increase or decrease their risk of developing a common disease such as cancer or cardiovascular disease through complex networks biomolecules that are hierarchically organized and highly interactive? Understanding the role of DNA sequences in disease susceptibility is likely to improve diagnosis, prevention and treatment. Success in this important public health endeavor will depend critically on the degree of nonlinearity in the mapping between genotype to phenotype. Nonlinearities can arise from phenomena such as locus heterogeneity (i.e., different DNA sequence variations leading to the same phenotype), phenocopy (i.e., environmentally determined phenotypes), and the dependence of genotypic effects on environmental factors (i.e., gene-environment interactions or plastic reaction norms) and genotypes at other loci (i.e., gene-gene interactions or *epistasis*). It is this latter source of nonlinearity, *epistasis*, that is of interest here. Epistasis has been recognized for many years as deviations from the simple inheritance patterns observed by Mendel (Bateson, 1909) or deviations from additivity in a linear statistical model (Fisher, 1918) and is likely due, in part, to canalization or mechanisms of stabilizing selection that evolve robust (i.e., redundant) gene networks (Gibson & Wagner, 2000; Waddington, 1942, 1957; Proulx & Phillips, 2005).

Epistasis has been defined in multiple different ways (e.g., Brodie, 2000; Hollander, 1955; Philips, 1998). We have reviewed two types of epistasis, biological and statistical (Moore & Williams, 2005). Biological epistasis results from physical interactions between biomolecules (e.g., DNA, RNA, proteins, enzymes, etc.) and occur at the cellular level in an individual. This type of epistasis is what Bateson (1909) had in mind when he coined the term. Statistical epistasis on the other hand occurs at the population level and is realized when there is interindividual variation in DNA sequences. The statistical phenomenon of epistasis is what Fisher (1918) had in mind. The relationship between biological and statistical epistasis is often confusing but will be important to understand if we are to make biological inferences from statistical results (Moore & Williams, 2005).

The focus of the present study is the detection and characterization of statistical epistasis in human populations using data mining and machine learning methods. We first review the concept difficulty and then review a multifactor dimensionality reduction (MDR) approach that was developed specifically for this domain. We then present some ideas about how to scale the MDR approach to datasets with thousands of attributes (i.e., genome-wide analysis). Finally, we end with some ideas about how nonlinear genetic models might be statistically interpreted to facilitate making biological inferences.

CONCEPT DIFFICULTY

Epistasis can be defined as biological or statistical (Moore & Williams, 2005). Biological epistasis occurs at the cellular level when two or more biomolecules physically interact. In contrast, statistical epistasis occurs at the population level and is characterized by deviation from additivity in a linear mathematical model. Consider the following simple example of statistical epistasis in the form of a penetrance function. Penetrance is simply the probability (P) of disease (D) given a particular combination of genotypes (G) that was inherited (i.e., P[D|G]). A single genotype is determined by one allele (i.e., a specific DNA sequence state) inherited from the mother and one allele inherited from the father. For most single nucleotide polymorphisms or SNPs, only two alleles (e.g., encoded by *A* or *a*) exist in the biological population. Therefore, because the

Table 1. Penetrance values for genotypes from two SNPs

	AA (0.25)	*Aa* (0.50)	*aa* (0.25)
BB (0.25)	0	.1	0
Bb (0.50)	.1	0	.1
bb (0.25)	0	.1	0

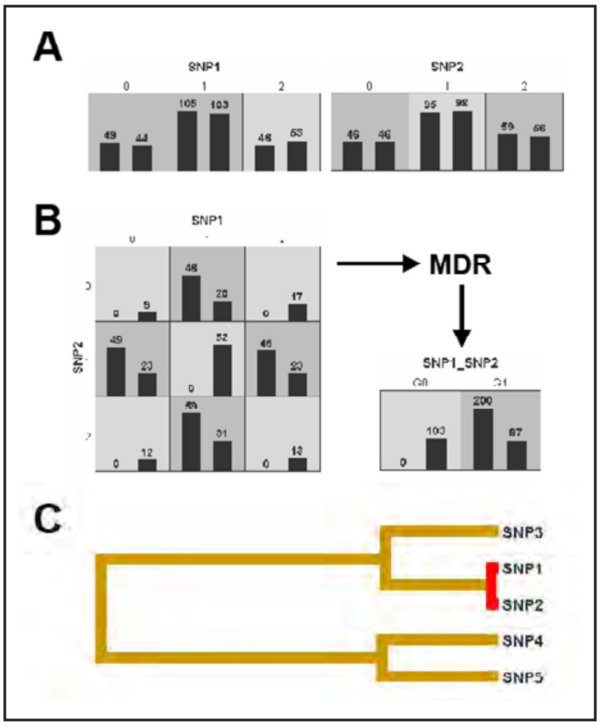
order of the alleles is unimportant, a genotype can have one of three values: *AA*, *Aa* or *aa*. The model illustrated in Table 1 is an extreme example of epistasis. Let's assume that genotypes *AA*, *aa*, *BB*, and *bb* have population frequencies of 0.25 while genotypes *Aa* and *Bb* have frequencies of 0.5 (values in parentheses in Table 1). What makes this model interesting is that disease risk is dependent on the particular *combination* of genotypes inherited. Individuals have a very high risk of disease if they inherit *Aa* or *Bb* but not both (i.e., the exclusive OR function). The penetrance for each individual genotype in this model is 0.5 and is computed by summing the products of the genotype frequencies and penetrance values. Thus, in this model there is no difference in disease risk for each single genotype as specified by the single-genotype penetrance values. This model is labeled M170 by Li and Reich (2000) in their categorization of genetic models involving two SNPs and is an example of a pattern that is not linearly separable. Heritability or the size of the genetic effect is a function of these penetrance values. The model specified in Table 1 has a heritability of 0.053 which represents a small genetic effect size. This model is a special case where all of the heritability is due to epistasis. As Freitas (2001) reviews this general class of problems has high concept difficulty. Moore (2003) suggests that epistasis will be the norm for common human diseases such as cancer, cardiovascular disease, and psychiatric diseases.

MULTIFACTOR DIMENSIONALITY REDUCTION

Multifactor dimensionality reduction (MDR) was developed as a nonparametric (i.e., no parameters are estimated) and genetic model-free (i.e., no genetic model is assumed) data mining strategy for identifying combinations of SNPs that are predictive of a discrete clinical endpoint (Hahn & Moore, 2004; Hahn, Ritchie, & Moore, 2003; Moore, 2004; Moore et al., 2006; Ritchie, Hahn, & Moore, 2003; Ritchie et al., 2001). At the heart of the MDR approach is a feature or attribute construction algorithm that creates a new attribute by pooling genotypes from multiple SNPs. The process of defining a new attribute as a function of two or more other attributes is referred to as constructive induction or attribute construction and was first developed by Michalski (1983). Constructive induction using the MDR kernel is accomplished in the following way. Given a threshold *T*, a multilocus genotype combination is considered high-risk if the ratio of cases (subjects with disease) to controls (healthy subjects) exceeds *T*, else it is considered low-risk. Genotype combinations considered to be high-risk are labeled G_1 while those considered low-risk are labeled G_0. This process constructs a new one-dimensional attribute with levels G_0 and G_1. It is this new single variable that is assessed using any classification method. The MDR method is based on the idea that changing the representation space of the data will make it easier for a classifier such as a decision tree or a naive Bayes learner to detect attribute dependencies. Open-source software in Java and C are freely available from www.epistasis.org/software.html

Consider the simple example presented above and in Table 1. This penetrance function was used to simulate a dataset with 200 cases (diseased subjects) and 200 controls (healthy subjects) for a total of 400 instances. The list of attributes included the two functional interacting SNPs (SNP1 and SNP2) in addition to three randomly

Figure 1. (a) Distribution of cases (left bars) and controls (right bars) across three genotypes (0, 1, 2) for two simulated interacting SNPs; (b) distribution of cases and controls across nine two-locus genotype combinations**; (c) an interaction dendrogram summarizing the information gain associated with constructing pairs of attributes using MDR****



*Note: * The ratio of cases to controls for these two SNPs are nearly identical. The dark shaded cells signify "high-risk" genotypes. ** Considering the two SNPs jointly reveals larger case-control ratios. Also illustrated is the use of the MDR attribute construction function that produces a single attribute (SNP1_SNP2) from the two SNPs. *** The length of the connection between two SNPs is inversely related to the strength of the information gain. Red lines indicate a positive information gain that can be interpreted as synergistic interaction. Brown lines indicate no information gain*

generated SNPs (SNP3 – SNP5). All attributes in these datasets are categorical. The SNPs each have three levels (0, 1, 2) while the class has two levels (0, 1) that code controls and cases. Figure 1a illustrates the distribution of cases (left bars) and controls (right bars) for each of the three genotypes of SNP1 and SNP2. The dark-shaded cells have been labeled "high-risk" using a threshold of T = 1. The light-shaded cells have been labeled "low-risk." Note that when considered individually, the ratio of cases to controls is close to one for each single genotype. Figure 1b illustrates the distribution of cases and controls when the two functional SNPs are considered jointly. Note the larger ratios that are consistent with the genetic model in Table 1. Also illustrated in Figure 1b is the distribution of cases and controls for the new single attribute constructed using MDR. This new single attribute captures much of the information from the interaction and could be assessed using a simple naïve Bayes classifier.

The MDR method has been successfully applied to detecting epistasis or gene-gene interactions for a variety of common human diseases including, for example, sporadic breast cancer (Ritchie et al., 2001), essential hypertension

(Moore & Williams, 2002; Williams et al., 2004), atrial fibrillation (Moore et al., 2006; Tsai et al., 2004), myocardial infarction (Coffey et al., 2004), type II diabetes (Cho et al., 2004), prostate cancer (Xu et al., 2005), bladder cancer (Andrew et al., 2006), schizophrenia (Qin et al., 2005), and familial amyloid polyneuropathy (Soares et al., 2005). The MDR method has also been successfully applied in the context of pharmacogenetics and toxicogenetics (e.g., Wilke, Reif, & Moore, 2005).

Consider the following case study. Andrew et al. (2006) carried out an epidemiologic study to identify genetic and environmental predictors of bladder cancer susceptibility in a large sample of Caucasians (914 instances) from New Hampshire. This study focused specifically on genes that play an important role in the repair of DNA sequences that have been damaged by chemical compounds (e.g., carcinogens). Seven SNPs were measured including two from the *X-ray repair cross-complementing group 1* gene (*XRCC1*), one from the *XRCC3* gene, two from the *xeroderma pigmentosum group D* (*XPD*) gene, one from the *nucleotide excision repair* gene (*XPC*), and one from the *AP endonuclease 1* gene (*APE1*). Each of these genes plays an important role in DNA repair. Smoking is a known risk factor for bladder cancer and was included in the analysis along with gender and age for a total of 10 attributes. Age was discretized to > or ≤ 50 years.

A parametric statistical analysis of each attribute individually revealed a significant independent main effect of smoking as expected. However, none of the measured SNPs were significant predictors of bladder cancer individually. Andrew et al. (2006) used MDR to exhaustively evaluate all possible two-, three-, and four-way interactions among the attributes. For each combination of attributes a single constructed attribute was evaluated using a naïve Bayes classifier. Training and testing accuracy were estimated using 10-fold cross-validation. A best model was selected that maximized the testing accuracy. The best model included two SNPs from the *XPD* gene and smoking. This three-attribute model had a testing accuracy of 0.66. The empirical p-value of this model was less than 0.001 suggesting that a testing accuracy of 0.66 or greater is unlikely under the null hypothesis of no association as assessed using a 1000-fold permutation test. Decomposition of this model using measures of information gain (see Moore et al., 2006; see below) demonstrated that the effects of the two *XPD* SNPs were non-addtive or synergistic suggestive of nonlinear interaction. This analysis also revealed that the effect of smoking was mostly independent of the nonlinear genetic effect. It is important to note that parametric logistic regression was unable to model this three-attribute interaction due lack of convergence. This study illustrates the power of MDR to identify complex relationships between genes, environmental factors such as smoking, and susceptibility to a common disease such as bladder cancer. The MDR approach works well in the context of an exhaustive search but how does it scale to genome-wide analysis of thousands of attributes?

GENOME-WIDE ANALYSIS

Biological and biomedical sciences are undergoing an information explosion and an understanding implosion. That is, our ability to generate data is far outpacing our ability to interpret it. This is especially true in the domain of human genetics where it is now technically and economically feasible to measure thousands of SNPs from across the human genome. It is anticipated that at least one SNP occurs approximately every 100 nucleotides across the $3*10^9$ nucleotide human genome. An important goal in human genetics is to determine which of the many thousands of SNPs are useful for predicting who is at risk for common diseases. This "genome-wide" approach is expected to revolutionize the genetic analysis of common human diseases (Hirschhorn & Daly, 2005; Wang, Barratt, Clayton, & Todd, 2005) and is quickly replacing the traditional "candidate gene" approach that focuses on several genes selected by their known or suspected function.

Moore and Ritchie (2004) have outlined three significant challenges that must be overcome if we are to successfully identify genetic predictors of health and disease using a genome-wide approach. First, powerful data mining and machine learning methods will need to be developed to statistically model the relationship between combinations of DNA sequence variations and disease susceptibility. Traditional methods such as logistic regression have limited power for modeling high-order nonlinear interactions (Moore & Williams, 2002). The MDR approach was discussed above as an alternative to logistic regression. A second challenge is the selection of genetic features or attributes that should be included for analysis. If interactions between genes explain most of the heritability of common diseases, then combinations of DNA sequence variations will need to be evaluated from a list of thousands of candidates. Filter and wrapper methods will play an important role because there are more combinations than can be exhaustively evaluated. A third challenge is the interpretation of gene-gene interaction models. Although a statistical model can be used to identify DNA sequence variations that confer risk for disease, this approach cannot be translated into specific prevention and treatment strategies without interpreting the results in the context of human biology. Making etiological inferences from computational models may be the most important and the most difficult challenge of all (Moore & Williams, 2005).

Combining the concept difficulty described above with the challenge of attribute selection yields what Goldberg (2002) calls a *needle-in-a-haystack* problem. That is, there may be a particular combination of SNPs that together with the right nonlinear function are a significant predictor of disease susceptibility. However, individually they may not look any different than thousands of other SNPs that are not involved in the disease process and are thus noisy. Under these models, the learning algorithm is truly looking for a genetic needle in a genomic haystack. A recent report from the International HapMap Consortium (Altshuler et al., 2005) suggests that approximately 300,000 carefully selected SNPs may be necessary to capture all of the relevant variation across the Caucasian human genome. Assuming this is true (it is probably a lower bound), we would need to scan $4.5 * 10^{10}$ pair wise combinations of SNPs to find a genetic needle. The number of higher order combinations is astronomical. What is the optimal approach to this problem?

There are two general approaches to selecting attributes for predictive models. The filter approach pre-processes the data by algorithmically or statistically assessing the quality of each attribute and then using that information to select a subset for classification. The wrapper approach iteratively selects subsets of attributes for classification using either a deterministic or stochastic algorithm. The key difference between the two approaches is that the classifier plays no role in selecting which attributes to consider in the filter approach. As Freitas (2002) reviews, the advantage of the filter is speed while the wrapper approach has the potential to do a better job classifying. We discuss each of these general approaches in turn for the specific problem of detecting epistasis or gene-gene interactions on a genome-wide scale.

A FILTER STRATEGY FOR GENOME-WIDE ANALYSIS

There are many different statistical and computational methods for determining the quality of attributes. A standard strategy in human genetics is to assess the quality of each SNP using a chi-square test of independence followed by a correction of the significance level that takes into account an increased false-positive (i.e., type I error) rate due to multiple tests. This is a very efficient filtering method but it ignores the dependencies or interactions between genes. Kira and Rendell (1992) developed an algorithm called Relief that is capable of detecting attribute dependencies. Relief estimates the quality of attributes through a type of nearest neighbor algorithm that selects neighbors (instances) from the same class and from the different class based on the vector of values across attributes. Weights (W) or quality estimates for each attribute (A) are estimated based on whether the nearest neighbor (nearest hit, H) of a randomly selected instance (R) from the same class and the nearest neighbor from the other class (nearest miss, M) have the same or different values. This process of adjusting weights is repeated for m instances. The algorithm produces weights for each attribute ranging from -1 (worst) to +1 (best). The Relief pseudocode is outlined below:

```
set all weights W[A] = 0
for i = 1 to m do begin
    randomly select an instance R_i
    find nearest hit H and nearest miss M
    for A = 1 to a do
        W[A] = W[A] − diff(A, R_i, H)/m
        + diff(A, R_i, M)/m
    end
```

The function diff(A, I_1, I_2) calculates the difference between the values of the attribute A for two instances I_1 and I_2. For nominal attributes such as SNPs it is defined as:

diff(A, I_1, I_2) = 0 if genotype(A, I_1) = genotype(A, I_2),
1 otherwise

The time complexity of Relief is $O(m{*}n{*}a)$ where *m* is the number of instances randomly sampled from a dataset with *n* total instances and a attributes. Kononenko (1994) improved upon Relief by choosing *n* nearest neighbors instead of just one. This new ReliefF algorithm has been shown to be more robust to noisy attributes (Kononenko, 1994; Robnik-Šikonja & Kononenko, 2001, 2003) and is widely used in data mining applications.

ReliefF is able to capture attribute interactions because it selects nearest neighbors using the entire vector of values across all attributes. However, this advantage is also a disadvantage because the presence of many noisy attributes can reduce the signal the algorithm is trying to capture. Moore and White (2007a) proposed a "tuned" ReliefF algorithm (TuRF) that systematically removes attributes that have low quality estimates so that the ReliefF values if the remaining attributes can be reestimated. The pseudocode for TuRF is outlined below:

```
let a be the number of attributes
for i = 1 to n do begin
    estimate ReliefF
    sort attributes
    remove worst n/a attributes
end
return last ReliefF estimate for each attribute
```

The motivation behind this algorithm is that the ReliefF estimates of the true functional attributes will improve as the noisy attributes are removed from the dataset.

Moore and White (2007a) carried out a simulation study to evaluate the power of ReliefF, TuRF, and a naïve chi-square test of independence for selecting functional attributes in a filtered subset. Five genetic models in the form of penetrance functions (e.g., Table 1) were generated. Each model consisted of two SNPs that define a nonlinear relationship with disease susceptibility. The heritability of each model was 0.1 which reflects a moderate to small genetic effect size. Each of the five models was used to generate 100 replicate datasets with sample sizes of 200, 400, 800, 1600, 3200 and 6400. This range of sample sizes represents a spectrum that is consistent with small to medium size genetic studies. Each dataset consisted of an equal number of case (disease) and control (no disease) subjects. Each pair of functional SNPs was combined within a genome-wide set of 998 randomly generated SNPs for a total of 1000 attributes. A total of 600 datasets were generated and analyzed.

ReliefF, TuRF and the univariate chi-square test of independence were applied to each of the datasets. The 1000 SNPs were sorted according to their quality using each method and the top 50, 100, 150, 200, 250, 300, 350, 400, 450 and 500 SNPs out of 1000 were selected. From each subset we counted the number of times the two functional SNPs were selected out of each set of 100 replicates. This proportion is an estimate of the power or how likely we are to find the true SNPs if they exist in the dataset. The number of times each method found the correct two SNPs was statistically compared. A difference in counts (i.e., power) was considered statistically significant at a type I error rate of 0.05. Moore and White (2007a) found that the power of ReliefF to pick (filter) the correct two functional attributes was consistently better ($P \leq 0.05$) than a naïve chi-square test of independence across subset sizes and models when the sample size was 800 or larger. These results suggest that ReliefF is capable of identifying interacting SNPs with a moderate genetic effect size (heritability=0.1) in moderate sample sizes. Next, Moore and White (2007a) compared the power of TuRF to the power of ReliefF. They found that the TuRF algorithm was consistently better ($P \leq 0.05$) than ReliefF across small SNP subset sizes (50, 100, and 150) and across all five models when the sample size was 1600 or larger. These results suggest that algorithms based on ReliefF show promise for filtering interacting attributes in this domain. The disadvantage of the filter approach is that important attributes might be discarded prior to analysis. Stochastic search or wrapper methods provide a flexible alternative.

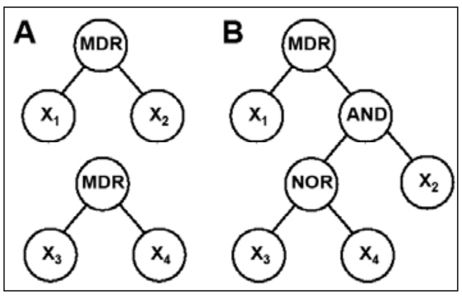
A WRAPPER STRATEGY FOR GENOME-WIDE ANALYSIS

Stochastic search or wrapper methods may be more powerful than filter approaches because no attributes are discarded in the process. As a result, every attribute retains some probability of being selected for evaluation by the classifier. There are many different stochastic wrapper algorithms that can be applied to this problem. Moore and White (2007b) have explored the use of genetic programming (GP). Genetic programming (GP) is an automated computational discovery tool that is inspired by Darwinian evolution and natural selection (Banzhaf, Nordin, Keller, & Francone, 1998; Koza 1992, 1994; Koza, Bennett, Andre, & Keane, 1999; Koza et al., 2003; Langdon, 1998; Langdon & Poli, 2002). The goal of GP is evolve computer programs to solve problems. This is accomplished by first generating random computer programs that are composed of the basic building blocks needed to solve or approximate a solution to the problem. Each randomly generated program is evaluated and the good programs are selected and recombined to form new computer programs. This process of selection based on fitness and recombination to generate variability is repeated until a best program or set of programs is identified. Genetic programming and its many variations have been applied successfully in a wide range of different problem domains including data mining and knowledge discovery (e.g., Freitas, 2002), electrical engineering (e.g., Koza et al., 2003), and bioinformatics (e.g., Fogel & Corne, 2003).

Moore and White (2007b) developed and evaluated a simple GP wrapper for attribute selection in the context of an MDR analysis. Figure 2a illustrates an example GP binary expression tree. Here, the root node consists of the MDR attribute construction function while the two leaves on the tree consist of attributes. Figure 2b illustrates a more complex tree structure that could be implemented by providing additional functions and allowing the binary expression trees to grow beyond one level. Moore and White (2007b) focused exclusively on the simple one-level GP trees as a baseline to assess the potential for this stochastic wrapper approach.

The goal of this study was to develop a stochastic wrapper method that is able to select attributes that interact in the absence of independent main effects. At face value, there is no reason to expect that a GP or any other wrapper method would perform better than a random attribute selector because there are no "building blocks" for this problem when accuracy is used as the fitness measure. That is, the fitness of any given classifier would look no better than any other with just one of the correct SNPs in the MDR model. Preliminary studies by White, Gilbert, Reif, and Moore (2005) support this idea. For GP or any other wrapper to work there needs to be recognizable building blocks. Moore and White (2007b) specifically evaluated whether including pre-processed attribute quality estimates using TuRF (see above) in a multiobjective fitness function improved attribute selection over a random

Figure 2. (a) Example of a simple GP binary expression with two attributes and an MDR function as the root node; (b) example of what a more complex GP tree might look like



search or just using accuracy as the fitness. Using a wide variety of simulated data, Moore and White (in press) demonstrated that including TuRF scores in addition to accuracy in the fitness function significantly improved the power of GP to pick the correct two functional SNPs out of 1000 total attributes. A subsequent study showed that using TuRF scores to select trees for recombination and reproduction performed significantly better than using TuRF in a multiobjective fitness function (Moore & White, 2006).

This study presents preliminary evidence suggesting that GP might be useful for the genome-wide genetic analysis of common human diseases that have a complex genetic architecture. The results raise numerous questions. How well does GP do when faced with finding three, four or more SNPs that interact in a nonlinear manner to predict disease susceptibility? How does extending the function set to additional attribute construction functions impact performance? How does extending the attribute set impact performance? Is using GP better than filter approaches? To what extent can GP theory help formulate an optimal GP approach to this problem? Does GP outperform other evolutionary or non-evolutionary search methods? Does the computational expense of a stochastic wrapper like GO outweigh the potential for increased power? The studies by Moore and White (2006, 2007b) provide a starting point to begin addressing some of these questions.

STATISTICAL AND BIOLOGICAL INTERPRETATION

Multifactor dimensionality reduction is powerful attribute construction approach for detecting epistasis or nonlinear gene-gene interactions in epidemiologic studies of common human diseases. The models that MDR produces are by nature multidimensional and thus diffuct to interpret. For example, an interaction model with four SNPs, each with three genotypes, summarizes

81 different genotype (i.e., level) combinations (i.e., 3^4). How do each of these level combinations relate back to biological processes in a cell? Why are some combinations associated with high-risk for disease and some associated with low-risk for disease? Moore et al. (2006) have proposed using information theoretic approaches with graph-based models to provide both a statistical and a visual interpretation of a multidimensional MDR model. Statistical interpretation should facilitate biological interpretation because it provides a deeper understanding of the relationship between the attributes and the class variable. We describe next the concept of interaction information and how it can be used to facilitate statistical interpretation.

Jakulin and Bratko (2003) have provided a metric for determining the gain in information about a class variable (e.g., case-control status) from merging two attributes into one (i.e., attribute construction) over that provided by the attributes independently. This measure of *information gain* allows us to gauge the benefit of considering two (or more) attributes as one unit. While the concept of information gain is not new (McGill, 1954), its application to the study of attribute interactions has been the focus of several recent studies (Jakulin & Bratko, 2003; Jakulin et al., 2003). Consider two attributes, A and B, and a class label C. Let H(X) be the Shannon entropy (see Pierce, 1980) of X. The information gain (IG) of A, B, and C can be written as (1) and defined in terms of Shannon entropy (2 and 3).

$$IG(ABC) = I(A;B|C) - I(A;B) \quad (1)$$

$$I(A;B|C) = H(A|C) + H(B|C) - H(A,B|C) \quad (2)$$

$$I(A;B) = H(A) + H(B) - H(A,B) \quad (3)$$

The first term in (1), I(A;B|C), measures the *interaction* of A and B. The second term, I(A;B), measures the *dependency* or correlation between A and B. If this difference is positive, then there is evidence for an attribute interaction that cannot be linearly decomposed. If the difference is negative, then the information between A and B is redundant. If the difference is zero, then there is evidence of conditional independence or a mixture of synergy and redundancy. These measures of interaction information can be used to construct interaction graphs (i.e., network diagrams) and an interaction dendrograms using the entropy estimates from Step 1 with the algorithms described first by Jakulin and Bratko (2003) and more recently in the context of genetic analysis by Moore et al. (2006). Interaction graphs are comprised of a node for each attribute with pairwise connections between them. The percentage of entropy removed (i.e., information gain) by each attribute is visualized for each node. The percentage of entropy removed for each pairwise MDR product of attributes is visualized for each connection. Thus, the independent main effects of each polymorphism can be quickly compared to the interaction effect. Additive and nonadditive interactions can be quickly assessed and used to interpret the MDR model which consists of distributions of cases and controls for each genotype combination. Positive entropy values indicate synergistic interaction while negative entropy values indicate redundancy.

Interaction dendrograms are also a useful way to visualize interaction (Jakulin & Bratko 2003; Moore et al., 2006). Here, hierarchical clustering is used to build a dendrogram that places strongly interacting attributes close together at the leaves of the tree. Jakulin and Bratko (2003) define the following dissimilarity measure, D (5), that is used by a hierarchical clustering algorithm to build a dendrogram. The value of 1000 is used as an upper bound to scale the dendrograms.

$$D(A,B) = \begin{cases} |I(A;B;C)|^{-1} & \text{if } |I(A;B;C)|^{-1} < 1000 \\ 1000 & \text{otherwise} \end{cases} \quad (5)$$

Using this measure, a dissimilarity matrix can be estimated and used with hierarchical cluster analysis to build an interaction dendrogram. This facilitates rapid identification and interpretation of pairs of interactions. The algorithms for the entropy-based measures of information gain are implemented in the open-source MDR software package available from www.epistasis.org. Output in the form of interaction dendrograms is provided. Figure 1c illustrates an interaction dendrogram for the simple simulated dataset described above. Note the strong synergistic relationship between SNP1 and SNP2. All other SNPs are independent which is consistent with the simulation model.

SUMMARY

We have reviewed a powerful attribute construction method called multifactor dimensionality reduction or MDR that can be used in a classification framework to detect nonlinear attribute interactions in genetic studies of common human diseases. We have also reviewed a filter method using ReliefF and a stochastic wrapper method using genetic programming (GP) for the analysis of gene-gene interaction or epistasis on a genome-wide scale with thousands of attributes. Finally, we reviewed information theoretic methods to facilitate the statistical and subsequent biological interpretation of high-order gene-gene interaction models. These data mining and knowledge discovery methods and others will play an increasingly important role in human genetics as the field moves away from the candidate-gene approach that focuses on a few targeted genes to the genome-wide approach that measures DNA sequence variations from across the genome.

ACKNOWLEDGMENT

This work was supported by National Institutes of Health (USA) grants LM009012, AI59694, HD047447, RR018787, and HL65234. We thank Mr. Bill White for his invaluable contributions to the methods discussed here.

REFERENCES

Altshuler, D., Brooks L. D, Chakravarti, A., Collins, F. S., Daly, M. J., & Donnelly, P. (2005). International HapMap consortium: A haplotype map of the human genome. *Nature, 437*, 1299-1320.

Andrew, A. S., Nelson, H. H., Kelsey, K. T., Moore, J. H., Meng, A. C., Casella, D. P., et al. (2006). Concordance of multiple analytical approaches demonstrates a complex relationship between DNA repair gene SNPs, smoking, and bladder cancer susceptibility. *Carcinogenesis, 27*, 1030-1037.

Banzhaf, W., Nordin, P., Keller, R. E., & Francone, F. D. (1998). *Genetic programming: An introduction: On the automatic evolution of computer programs and its applications*. San Francisco: Morgan Kaufmann Publishers.

Bateson, W. (1909). *Mendel's principles of heredity*. Cambridge, UK: Cambridge University Press.

Brodie III, E. D. (2000). Why evolutionary genetics does not always add up. In J. Wolf, B. Brodie III, M. Wade (Eds.), *Epistasis and the evolutionary process* (pp. 3-19). New York: Oxford University Press.

Cho, Y. M., Ritchie, M. D., Moore, J. H., Park, J. Y., Lee, K. U., Shin, H. D., et al. (2004). Multifactor-dimensionality reduction shows a two-locus interaction associated with Type 2 diabetes mellitus. *Diabetologia, 47*, 549-554.

Coffey, C. S., Hebert, P. R., Ritchie, M. D., Krumholz, H. M., Morgan, T. M., Gaziano, J. M., et al. (2004). An application of conditional logistic regression and multifactor dimensionality reduction for detecting gene-gene interactions on risk of myocardial infarction: The importance of model validation. *BMC Bioinformatics, 4*, 49.

Fisher, R. A. (1918). The correlations between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh, 52*, 399-433.

Fogel, G. B., & Corne, D. W. (2003). *Evolutionary computation in bioinformatics.* San Francisco: Morgan Kaufmann Publishers.

Freitas, A. (2001). Understanding the crucial role of attribute interactions. *Artificial Intelligence Review, 16*, 177-199.

Freitas, A. (2002). *Data mining and knowledge discovery with evolutionary algorithms.* New York: Springer.

Gibson, G., & Wagner, G. (2000). Canalization in evolutionary genetics: A stabilizing theory? *BioEssays, 22*, 372-380.

Goldberg, D. E. (2002). *The design of innovation*. Boston: Kluwer.

Hahn, L. W., & Moore, J. H. (2004). Ideal discrimination of discrete clinical endpoints using multilocus genotypes. *In Silico Biology, 4*, 183-194.

Hahn, L. W., Ritchie, M. D., & Moore, J. H. (2003). Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics*, *19*, 376-382.

Hirschhorn, J. N., & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics, 6*, 95-108.

Hollander, W. F. (1955). Epistasis and hypostasis. *Journal of Heredity, 46*, 222-225.

Jakulin, A., & Bratko, I. (2003). Analyzing attribute interactions. *Lecture Notes in Artificial Intelligence*, *2838*, 229-240.

Kira, K., & Rendell, L. A. (1992). A practical approach to feature selection. In D. H. Sleeman, & P. Edwards (Eds.), In *Proceedings of the Ninth International Workshop on Machine Learning*, San Francisco (pp. 249-256)

Kononenko, I. (1994). Estimating attributes: Analysis and extension of relief. In *Proceedings of the European Conference on Machine Learning* (pp. 171-182). New York: Springer.

Koza, J. R. (1992). *Genetic programming: On the programming of computers by means of natural selection*. Cambrindge, MA: The MIT Press.

Koza, J. R. (1994). *Genetic programming II: Automatic discovery of reusable programs*. Cambrindge, MA: The MIT Press.

Koza, J. R., Bennett III, F. H., Andre, D., & Keane, M. A. (1999). *Genetic programming III: Darwinian invention and problem solving.* San Francisco: Morgan Kaufmann Publishers.

Koza, J. R., Keane, M. A., Streeter, M. J., Mydlowec, W., Yu, J., & Lanza, G. (2003). *Genetic programming IV: Routine human-competitive machine intelligence*. New York: Springer.

Langdon, W. B. (1998). *Genetic programming and data structures: Genetic programming + data structures = automatic programming!* Boston: Kluwer.

Langdon, W. B., & Poli, R. (2002). *Foundations of genetic programming.* New York: Springer.

Li, W., & Reich, J. (2000). A complete enumeration and classification of two-locus disease models. *Human Heredity, 50*, 334-349.

McGill, W. J. (1954). Multivariate information transmission. *Psychometrica, 19*, 97-116.

Michalski, R. S. (1983). A theory and methodology of inductive learning. *Artificial Intelligence, 20*, 111-161.

Moore, J. H. (2003). The ubiquitous nature of epistasis in determining susceptibility to common human diseases. *Human Heredity, 56*, 73-82.

Moore, J. H. (2004). Computational analysis of gene-gene interactions in common human diseases using multifactor dimensionality reduction. *Expert Review of Molecular Diagnostics, 4*, 795-803.

Moore, J. H., Gilbert, J. C., Tsai, C.-T., Chiang, F. T., Holden, W., Barney, N., et al. (2006). A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. *Journal of Theoretical Biology, 241*, 252-261.

Moore, J. H., & Ritchie, M. D. (2004). The challenges of whole-genome approaches to common diseases. *Journal of the American Medical Association, 291*, 1642-1643.

Moore, J. H., & White, B. C. (2007a). *Tuning relief for genome-wide genetic analysis* (LNCS). New York: Springer.

Moore, J. H., & White, B. C. (2007b). Genome-wide genetic analysis using genetic programming. The critical need for expert knowledge. *Genetic programming theory and practice IV.* New York: Springer.

Moore, J. H., & White, B. C. (2006). Exploiting expert knowledge in genetic programming for genome-wide genetic analysis. *Lecture Notes in Computer Science*. New York: Springer.

Moore, J. H., & Williams, S. W. (2002). New strategies for identifying gene-gene interactions in hypertension. *Annals of Medicine, 34,* 88-95.

Moore, J. H., & Williams, S. W. (2005). Traversing the conceptual divide between biological and statistical epistasis: Systems biology and a more modern synthesis. *BioEssays, 27,* 637-646.

Phillips, P. C. (1998). The language of gene interaction. *Genetics, 149,* 1167-1171.

Proulx, S. R., & Phillips, P. C. (2005). The opportunity for canalization and the evolution of genetic networks. *American Naturalist, 165*, 147-162.

Qin, S., Zhao, X., Pan, Y., Liu, J., Feng, G., Fu, J., et al. (2005). An association study of the N-methyl-D-aspartate receptor NR1 subunit gene (GRIN1) and NR2B subunit gene (GRIN2B) in schizophrenia with universal DNA microarray. *European Journal of Human Genetics, 13,* 807-814.

Ritchie, M. D., Hahn, L. W., & Moore, J. H. (2003). Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, phenocopy, and genetic heterogeneity. *Genetic Epidemiology, 24,* 150-157.

Ritchie, M. D., Hahn, L. W., Roodi, N., Bailey, L. R., Dupont, W. D., Parl, F. F., et al. (2001). Multifactor dimensionality reduction reveals high-order interactions among estrogen metabolism genes in sporadic breast cancer. *American Journal of Human Genetics, 69*, 138-147.

Robnik-Sikonja, M., & Kononenko, I. (2001). Comprehensible interpretation of Relief's Estimates. In *Proceedings of the Eighteenth International Conference on Machine Learning*, San Francisco (pp. 433-440).

Robnik-Siknja, M., & Kononenko, I. (2003). Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning, 53,* 23-69.

Soares, M. L., Coelho, T., Sousa, A., Batalov, S., Conceicao, I., Sales-Luis, M. L., et al. (2005). Susceptibility and modifier genes in Portuguese transthyretin V30M amyloid polyneuropathy: Complexity in a single-gene disease. *Human Molecular Genetics, 14,* 543-553.

Tsai, C. T., Lai, L. P., Lin, J. L., Chiang, F. T., Hwang, J. J., Ritchie, M. D., et al. (2004). Reninangiotensin system gene polymorphisms and atrial fibrillation. *Circulation*, *109*, 1640-1646.

Waddington, C. H. (1942). Canalization of development and the inheritance of acquired characters. *Nature, 150*, 563-565.

Waddington, C. H. (1957). *The strategy of the genes.* New York: MacMillan.

Wang, W. Y., Barratt, B. J., Clayton, D. G., & Todd, J. A. (2005). Genome-wide association studies: Theoretical and practical concerns. *Nature Reviews Genetics, 6*, 109-118.

White, B. C., Gilbert, J. C., Reif, D. M., & Moore, J. H. (2005). A statistical comparison of grammatical evolution strategies in the domain of human genetics. In *Proceedings of the IEEE Congress on Evolutionary Computing* (pp. 676-682). New York. IEEE Press.

Wilke, R. A., Reif, D. M., & Moore, J. H. (2005). Combinatorial pharmacogenetics. *Nature Reviews Drug Discovery, 4*, 911-918.

Williams, S. M., Ritchie, M. D., Phillips III, J. A., Dawson, E., Prince, M., Dzhura, E., et al. (2004). Multilocus analysis of hypertension: A hierarchical approach. *Human Heredity, 57*, 28-38.

Xu, J., Lowery, J., Wiklund, F., Sun, J., Lindmark, F., Hsu, F.-C., et al. (2005). The interaction of four inflammatory genes significantly predicts prostate cancer risk. *Cancer Epidemiology Biomarkers and Prevention*, *14*, 2563-2568.

This work was previously published in Knowledge Discovery and Data Mining: Challenges and Realities, edited by X. Zhu and I. Davidson, pp. 17-30, copyright 2007 by Information Science Publishing (an imprint of IGI Global).

Chapter 7.22
Gene Expression Programming and the Evolution of Computer Programs

Cândida Ferreira
Gepsoft, UK

ABSTRACT

In this chapter an artificial problem solver inspired in natural genotype/phenotype systems — gene expression programming — is presented. As an introduction, the fundamental differences between gene expression programming and its predecessors, genetic algorithms and genetic programming, are briefly summarized so that the evolutionary advantages of gene expression programming are better understood. The work proceeds with a detailed description of the architecture of the main players of this new algorithm (chromosomes and expression trees), focusing mainly on the interactions between them and how the simple yet revolutionary structure of the chromosomes allows the efficient, unconstrained exploration of the search space. And finally, the chapter closes with an advanced application in which gene expression programming is used to evolve computer programs for diagnosing breast cancer.

EVOLUTIONARY ALGORITHMS IN PROBLEM SOLVING

The way nature solves problems and creates complexity has inspired scientists to create artificial systems that learn by themselves how to solve a particular problem. The first attempts were done in the 1950s by Friedberg (1958; Friedberg et al., 1959), but ever since highly sophisticated systems have been developed that apply Darwin's ideas of natural evolution to the artificial world of computers and modeling. Of particular interest to this work are the Genetic Algorithms (GAs) and the Genetic Programming (GP) technique, as they are the predecessors of Gene Expression Programming (GEP), the most recent development in evolutionary computation and the theme of this chapter. A brief introduction to these three techniques is given below.

Genetic Algorithms

Genetic algorithms were invented by John Holland in the 1960s and they also apply biological evolution theory to computer systems (Holland, 1975). Like all evolutionary computer systems, GAs are an oversimplification of biological evolution. In this case, solutions to a problem are usually encoded in strings of 0s and 1s (chromosomes), and populations of such strings (individuals or candidate solutions) are used in order to evolve a good solution to a particular problem. From generation to generation candidate solutions are reproduced with modification and selected according to fitness. Modification in the original genetic algorithm was introduced by the genetic operators of mutation, crossover, and inversion.

It is worth pointing out that GAs' individuals consist of naked chromosomes or, in other words, GAs' individuals are simple replicators. And like all simple replicators, the chromosomes of genetic algorithms function simultaneously as genotype and phenotype: they are both the object of selection and the guardians of the genetic information that must be replicated and passed on with modification to the next generation. Consequently, the whole structure of the replicator determines the functionality and, therefore, the fitness of the individual. For instance, in such systems it would not be possible to use only a particular region of the replicator as a solution to a problem: The whole replicator is always the solution: nothing more, nothing less.

Figure 1. Tree crossover in genetic programming (arrows indicate the crossover points)

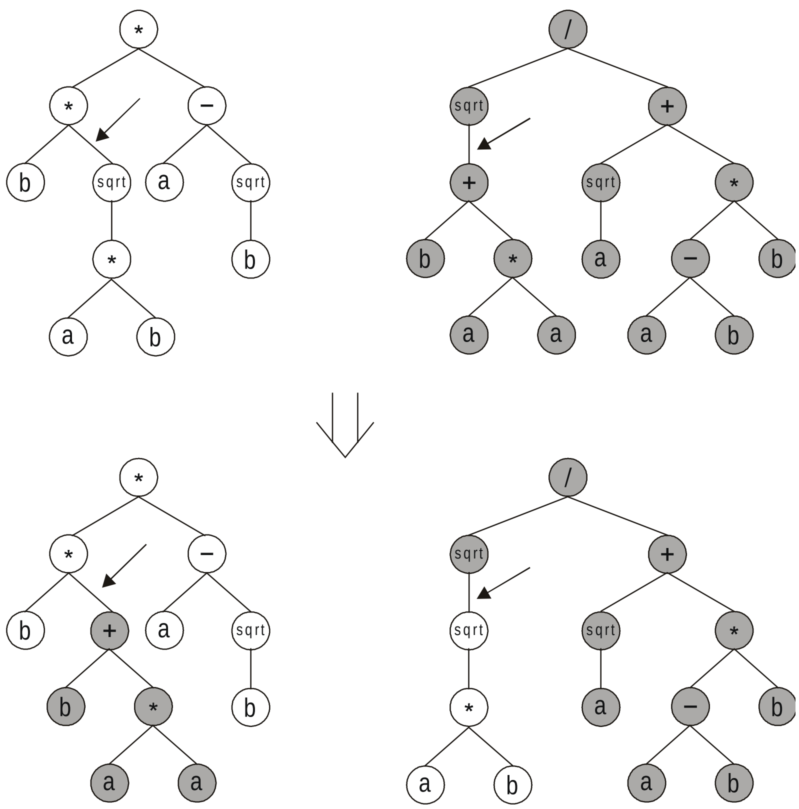


Figure 2. Tree mutation in genetic programming (The arrow indicates the mutation point. The new branch randomly generated by the mutation operator in the daughter tree is shown in gray.)

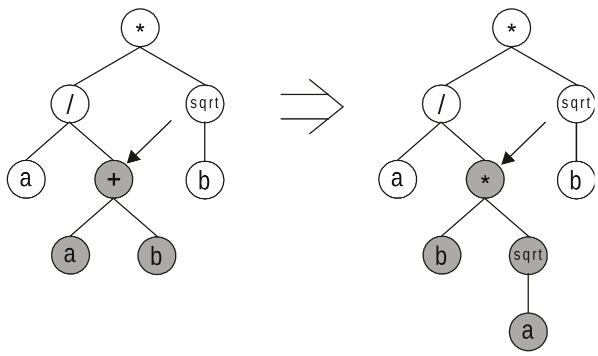


Figure 3. Permutation in genetic programming (The arrow indicates the permutation point. Note that the arguments of the permuted function traded places in the daughter tree.)

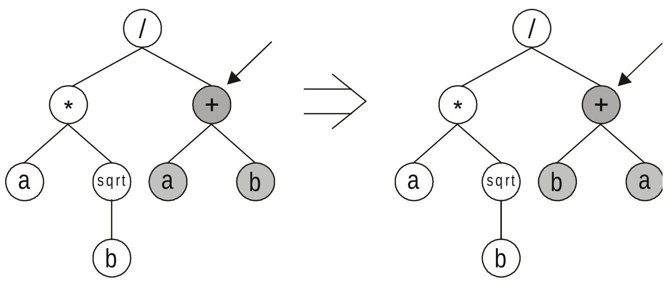
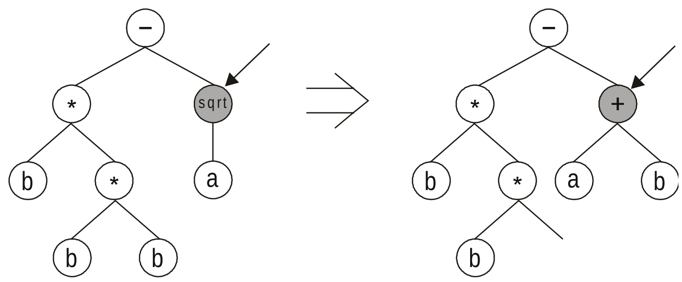


Figure 4. Illustration of a hypothetical event of point mutation in genetic programming (The arrow indicates the mutation point. Note that the daughter tree is an invalid structure.)



Genetic Programming

Genetic programming, invented by Cramer in 1985 (Cramer, 1985) and further developed by Koza (1992), solves the problem of fixed length solutions through the use of nonlinear structures (parse trees) with different sizes and shapes. The alphabet used to create these structures is also more varied, creating a richer, more versatile system of representation. Notwithstanding, the created individuals also lack a simple, autonomous genome. Like the linear chromosomes of genetic algorithms, the nonlinear structures of GP are also cursed with the dual role of genotype/phenotype.

The parse trees of genetic programming resemble protein molecules in their use of a richer alphabet and in their complex and unique hierarchical representation. Indeed, parse trees are capable of exhibiting a great variety of functionalities. The problem with these complex replicators is that their reproduction with modification is highly constrained in evolutionary terms because the modifications must take place on the parse tree itself and, consequently, only a limited range of modification is possible. Indeed, special kinds of genetic operators were developed that operate at the tree level, modifying or exchanging particular branches between trees.

Although at first sight this might appear advantageous, it greatly limits this technique (we all know the limits of grafting and pruning in nature). Consider for instance crossover, the most used and often the only search operator used in genetic programming. In this case, selected branches are exchanged between two parent trees to create offspring (Figure 1). The idea behind its implementation was to exchange smaller, mathematically concise blocks in order to evolve more complex, hierarchical solutions composed of smaller building blocks.

The mutation operator in GP is also very different from natural point mutation. This operator selects a node in the parse tree and replaces the branch underneath by a new randomly generated branch (Figure 2). Notice that the overall shape of the tree is not greatly changed by this kind of mutation, especially if lower nodes are preferentially chosen as mutation targets.

Permutation is the third operator used in genetic programming and the most conservative of the three. During permutation, the arguments of a randomly chosen function are randomly permuted (Figure 3). In this case the overall shape of the tree remains unchanged.

In summary, in genetic programming the operators resemble more of a conscious mathematician than the blind way of nature. But in adaptive systems the blind way of nature is much more efficient and systems such as GP are highly constrained. For instance, the implementation of other operators in genetic programming such as the simple yet high-performing point mutation (Ferreira, 2002c) is unproductive, as most mutations result in syntactically incorrect structures (Figure 4). Obviously, the implementation of other operators such as transposition or inversion raises similar difficulties and the search space in GP remains vastly unexplored.

Although Koza described these three operators as the basic GP operators, crossover is practically the only genetic operator used in most genetic programming applications (Koza, 1992). Consequently, no new material is introduced in the genetic pool of GP populations. Not surprisingly, huge populations of parse trees must be used with the aim of creating all the necessary building blocks with the inception of the initial population in order to guarantee the discovery of a good solution only by moving the initial building blocks around.

Finally, due to the dual function of the parse trees (genotype and phenotype), genetic programming is incapable of a simple, rudimentary expression: in all cases, the entire parse tree is the solution.

Gene Expression Programming

Gene expression programming was invented by the author in 1999 (Ferreira, 2001), and incorporates both the simple, linear chromosomes of fixed length similar to the ones used in genetic algorithms and the ramified structures of different sizes and shapes similar to the parse trees of genetic programming. This is equivalent to say that in gene expression programming the genotype and phenotype are finally separated and the system can now benefit from all the advantages this brings about.

Thus, the phenotype of GEP consists of the same kind of ramified structure used in genetic programming. But the ramified structures created by GEP (expression trees) are the expression of a totally autonomous genome. Therefore, with gene expression programming, the second evolutionary threshold — the phenotype threshold — is crossed (Dawkins, 1995). This means that, during reproduction, only the genome (slightly modified) is passed on to the next generation and we no longer need to replicate and mutate rather cumbersome structures: All the modifications take place in a simple linear structure which only later will grow into an expression tree.

The fundamental steps of gene expression programming are schematically represented in Figure 5. The process begins with the random generation of the chromosomes of a certain number of individuals (the initial population). Then these chromosomes are expressed and the fitness of each individual is evaluated against a set of fitness cases (also called selection environment). The individuals are then selected according to their fitness (their performance in that particular environment) to reproduce with modification, leaving progeny with new traits. These new individuals are, in their turn, subjected to the same developmental process: expression of the genomes, confrontation of the selection environment, selection, and reproduction with modification. The process is repeated for a certain number of generations or until a good solution has been found.

The pivotal insight of gene expression programming consisted of the invention of chromosomes capable of representing any parse tree. For that purpose a new language — *Karva* language — was created in order to read and express the information encoded in the chromosomes. The details of this new language are given in the next section.

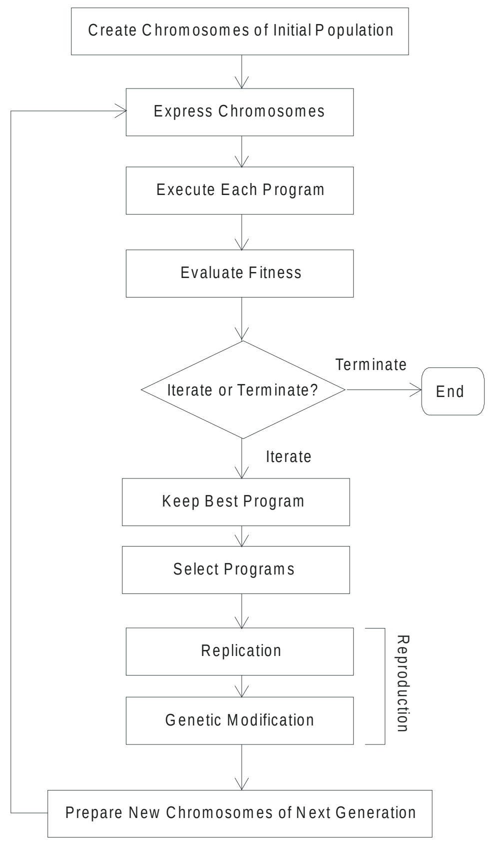
Furthermore, the structure of the chromosomes was designed to allow the creation of multiple genes, each coding for a smaller program or sub-expression tree. It is worth emphasizing that gene expression programming is the only genetic algorithm with multiple genes. Indeed, the creation of more complex individuals composed of multiple genes is extremely simplified in truly functional genotype/phenotype systems. In fact, after their inception, these systems seem to catapult themselves into higher levels of complexity such as multicellular systems, where different cells put together different consortiums of genes (Ferreira, 2002a).

The basis for all this novelty resides on the revolutionary structure of GEP genes. The simple but plastic structure of these genes not only allows the encoding of any conceivable program but also allows their efficient evolution. Due to this versatile structural organization, a very powerful set of genetic operators can be easily implemented and used to search very efficiently the solution space. As in nature, the search operators of gene expression programming always produce valid structures and therefore are remarkably suited to creating genetic diversity.

THE ARCHITECTURE OF GEP INDIVIDUALS

We know already that the main players in gene expression programming are the chromosomes and the expression trees (ETs), the latter being

Figure 5. Flowchart of gene expression programming



the expression of the genetic information encoded in the former. As in nature, the process of information decoding is called translation. And this translation implies obviously a kind of code and a set of rules. The genetic code is very simple: a one-to-one relationship between the symbols of the chromosome and the nodes they represent in the trees. The rules are also very simple: they determine the spatial organization of nodes in the expression trees and the type of interaction between sub-ETs. Therefore, there are two languages in GEP: the language of the genes and the language of expression trees and, thanks to the simple rules that determine the structure of ETs and their interactions, we will see that it is possible to infer immediately the phenotype given the sequence of a gene, and vice versa. This means that we can choose to have a very complex program represented by its compact genome without losing meaning. This unequivocal bilingual notation is called *Karva* language. Its details are explained in the remainder of this section.

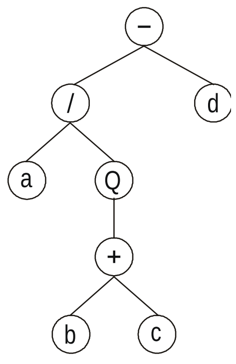
Open Reading Frames and Genes

The structural organization of GEP genes is better understood in terms of open reading frames (ORFs). In biology, an ORF or coding sequence of a gene begins with the start codon, continues with the amino acid codons, and ends at a termination codon. However, a gene is more than the respective ORF, with sequences upstream of the start codon and sequences downstream of the stop codon. Although in GEP the start site is always the first position of a gene, the termination point does not always coincide with the last position of a gene. Consequently, it is common for GEP genes to have non-coding regions downstream of the termination point. (For now we will not consider these non-coding regions, as they do not interfere with expression.)

Consider, for example, the algebraic expression:

$$\frac{a}{\sqrt{b+c}}-d \qquad \textbf{(1)}$$

It can also be represented as a diagram or ET:



where "Q" represents the square root function.

This kind of diagram representation is in fact the phenotype of GEP chromosomes. And the genotype can be easily inferred from the phenotype as follows:

```
01234567
-/daQ+bc
```

(2)

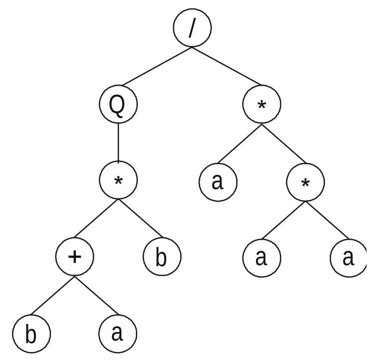
which is the straightforward reading of the ET from left to right and from top to bottom (exactly as we read a page of text). The expression (2) is an open reading frame, starting at "-" (position 0) and terminating at "c" (position 7). These ORFs were named *K-expressions* from Karva language.

Consider another open reading frame, the following K-expression:

```
012345678901
/Q**a*+baaba
```

(3)

Its expression as an ET is also very simple and straightforward. In order to express the ORF correctly, we must follow the rules governing the spatial distribution of functions and terminals. First, the start of a gene corresponds to the root of the expression tree, which is placed in the topmost line. Second, in the next line, below each function, are placed as many branch nodes as there are arguments to that function. Third, from left to right, the nodes are filled consecutively with the next elements of the K-expression. Fourth, the process is repeated until a line containing only terminals is formed. In this case, the following expression tree is formed:



which mathematically corresponds to

$$\frac{\sqrt{(b+a)b}}{a^3}.$$

Looking at the structure of GEP ORFs only, it is difficult or even impossible to see the advantages of such a representation, except perhaps for its simplicity and elegance. However, when open reading frames are analyzed in the context of a gene, the advantages of this representation become obvious. As previously stated, GEP chromosomes have fixed length, and they are composed of one or more genes of equal length. Consequently, the length of a gene is also fixed. Thus, in gene expression programming, what varies is not the length of genes, which is constant, but the length of the ORFs. Indeed, the length of an open reading frame may be equal to or less than the length of the gene. In the first case, the termination point coincides with the end of the gene, and in the latter, the termination point is somewhere upstream of the end of the gene.

What is the function of these non-coding regions of GEP genes? We will see that they are the essence of gene expression programming and evolvability, for they allow the modification of the genome using several genetic operators without restrictions, always producing syntactically correct programs. Thus, in gene expression programming, the fundamental property of genotype/phenotype systems — syntactic closure — is intrinsic, allowing the totally unconstrained restructuring of the genotype and, consequently, an efficient evolution.

In the next section we are going to analyze the structural organization of GEP genes in order to understand how they invariably code for syntactically correct programs and why they allow an unconstrained application of virtually any genetic operator.

Structural Organization of Genes

The genes of gene expression programming are composed of a head and a tail. The head contains symbols that represent both functions and terminals, whereas the tail contains only terminals. For each problem, the length of the head h is chosen, whereas the length of the tail t is a function of h and the number of arguments n of the function with more arguments (also called maximum arity) and is evaluated by the equation:

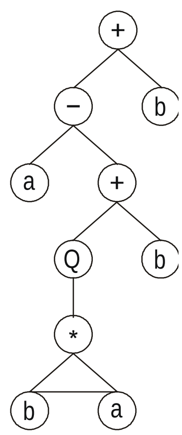
$$t = h\,(n-1) + 1 \qquad \textbf{(4)}$$

Consider a gene for which the set of functions F = {Q, *, /, -, +} and the set of terminals T = {a, b}. In this case $n = 2$; if we chose an $h = 11$, then $t = 11\,(2 - 1) + 1 = 12$; thus, the length of the gene g is 11 + 12 = 23. One such gene is shown below (the tail is shown in bold):

```
0123456789012
+-ba+Qb*ba/abaaaabbaabb
```

(5)

It codes for the following expression tree:



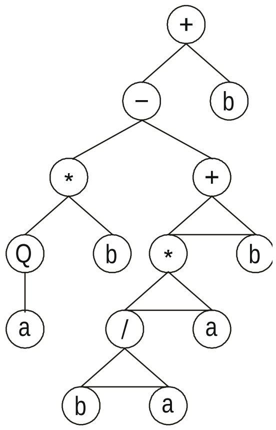
or the equivalent mathematical expression (*a*-). In this case, the open reading frame ends at position 9, whereas the gene ends at position 22.

Suppose now a mutation occurred at position 3, changing the “a” into “*”. Then the following gene is obtained:

```
01234567890123456789012
+-b*+Qb*ba/abaaaabbaabb
```

(6)

And its expression gives:

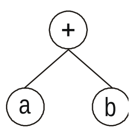


which mathematically corresponds to $b(\sqrt{a}\text{-}1)$. In this case, the termination point shifts four positions to the right (position 13), enlarging and changing significantly the daughter tree.

Obviously the opposite also might happen, and the daughter tree might shrink. For example, consider again gene (5), and suppose a mutation occurred at position 1, changing the "-" into "a":

```
01234567890123456789012
+aba+Qb*ba/abaaaabbaabb
```
(7)

Its expression results in the following ET:



In this case, the ORF ends at position 2, shortening the original ET in seven nodes.

So, despite their fixed length, each gene has the potential to code for expression trees of different sizes and shapes, where the simplest is composed of only one node (when the first element of a gene is a terminal) and the largest is composed of as many nodes as the length of the gene (when all the elements of the head are functions with maximum arity).

It is evident from the examples above that any modification made in the genome, no matter how profound, always results in a structurally correct program. Obviously, the structural organization of genes must be preserved, always maintaining the boundaries between head and tail. We will be able to fully appreciate the plasticity of GEP chromosomes in the section "Genetic Operators and Evolution," where the mechanisms and effects of different genetic operators will be thoroughly analyzed.

Multigenic Chromosomes

The chromosomes of gene expression programming are usually composed of more than one gene of equal length. For each problem or run, the number of genes, as well as the length of the head, are chosen *a priori*. Each gene codes for a sub-ET and the sub-ETs interact with one another, forming a more complex multi-subunit expression tree.

Consider, for example, the following chromosome with length 39, composed of three genes, each with length 13 (the tails are shown in bold):

```
0123456789012012345678901201234567
89012
*Qb+*/bbbabab-a+Qbabbababa/ba-
/*bbaaaaa
```
(8)

It has three open reading frames, and each ORF codes for a sub-ET (Figure 6). The start of each ORF is always given by position 0; the end of each ORF, though, is only evident upon construction of the corresponding sub-ET. As shown in Figure 6, the first open reading frame ends at position 9; the second ORF ends at position 5; and the last ORF ends at position 2. Thus, GEP chromosomes contain several ORFs of different sizes, each ORF coding for a structurally and functionally unique sub-ET. Depending on the problem at hand, these sub-ETs may be selected individually depending on their respective outputs, or they may form a more complex, multi-subunit expression tree and be selected as a whole. In these multi-subunit structures, individual sub-ETs interact with one another by a particular kind of posttranslational interaction or linking. For instance, algebraic sub-ETs can be linked by addition or multiplication whereas Boolean sub-ETs can be linked by OR, AND or IF.

The linking of three sub-ETs by addition is illustrated in Figure 6 c. Note that the final ET could be linearly encoded as the following K-expression:

```
012345678901234567890
++/*-baQba++Qb*/abbba
```

(9)

However, the use of multigenic chromosomes is more appropriate to evolve solutions to complex problems, for they permit the modular construction of complex, hierarchical structures, where each gene codes for a smaller and simpler building block. These smaller building blocks are separated from each other, and thus can evolve independently. Not surprisingly, these multigenic systems are much more efficient than unigenic ones (Ferreira 2001, 2002a).

GENETIC OPERATORS AND EVOLUTION

Genetic operators are the core of all evolutionary algorithms, and two of them are common to all evolutionary systems: selection and replication. Indeed, all artificial systems use a scheme to select individuals more or less according to fitness. Some schemes are totally deterministic, whereas others include a touch of unpredictability. Gene expression programming uses one of the latter, namely, a fitness proportionate roulette-wheel scheme (see, e.g., Goldberg, 1989) coupled with the cloning of the best individual (simple elitism), as it mimics nature very faithfully and produces very good results.

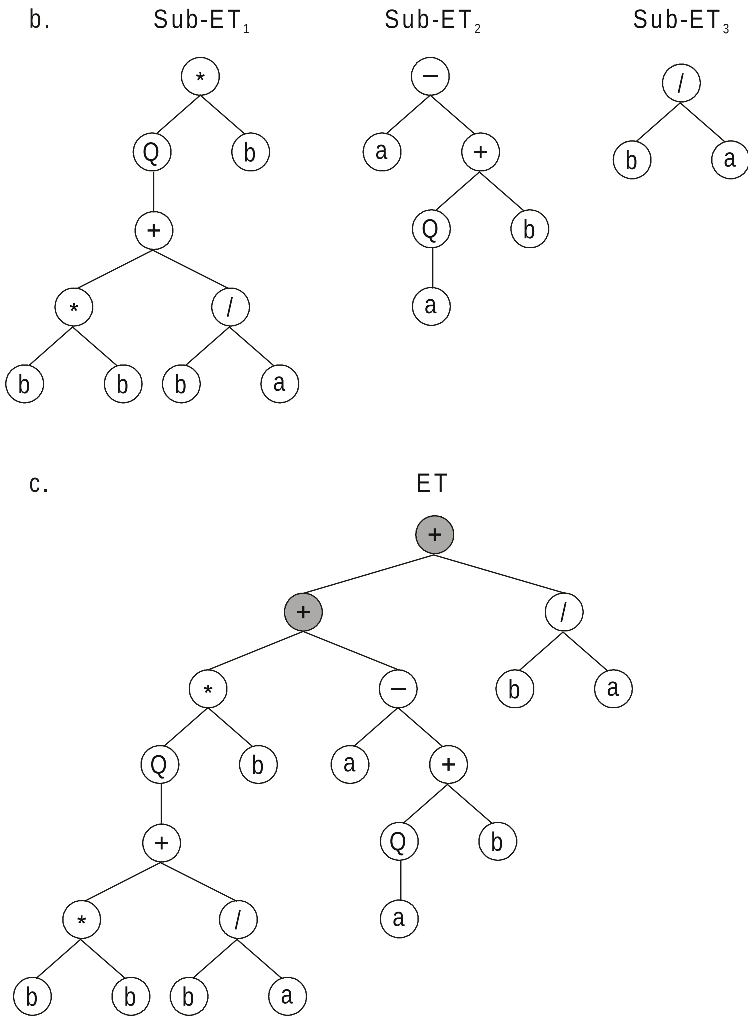
Thus, according to fitness and the luck of the draw, individuals are selected to be replicated. Although crucial, replication is the most uninteresting operator. During replication, chromosomes are dully copied and passed on to the next generation. The fitter the individual, the higher the probability of passing on its genes to the next generation. So, during replication, the genomes of the selected individuals are copied every time the roulette picks them up. And the roulette is spun as many times as there are individuals in the population so that the same population size is maintained from generation to generation.

Although the center of the storm, by themselves, selection and replication do nothing in terms of adaptation. In fact, by themselves they can only cause genetic drift, making populations less and less diverse with time until all the individuals are exactly the same. So, the corner stone of all evolutionary systems is genetic modification. And different algorithms create this modification differently. For instance, genetic algorithms normally use mutation and recombination; genetic programming uses almost exclusively tree recombination; and gene expression programming uses mutation, inversion, transposition, and recombination.

With the exception of GP, which is severely constrained in terms of tools of genetic modification, in both GAs and GEP it is possible to implement easily a vast set of search operators because the search operators act on simple linear chromosomes. In fact, a varied set of search operators was implemented in gene expression programming in order to shed some light on the dynamics of evolutionary systems, but what is important is to provide for the necessary degree of genetic diversification in order to allow an efficient evolution. Nevertheless, mutation (by far the most efficient operator) by itself is capable

Figure 6: Expression of GEP genes as sub-ETs — (a) A three-genic chromosome with the tails in bold; Position zero marks the start of each gene; (b) Sub-ETs codified by each gene, which correspond respectively to $b\sqrt{b^2+b/a}$, *a-(*$\sqrt{a}$*+b), and b/a; (c) The result of posttranslational linking with addition, which obviously corresponds to* $b\sqrt{b^2+b/a} + a - \sqrt{a} - b + b/a$; *the linking functions are shown in gray*

a. `012345678901201234567890120123456789012`
`*Qb+*/`**`bbbabab`**`-a+Qba`**`bbababa`**`/ba-/*`**`bbaaaaa`**



of wonders. However, the interplay of mutation with other operators not only allows an efficient evolution but also allows the duplication of genes and their subsequent differentiation; the creation of small repetitive sequences; and so forth, making things really interesting.

In the remainder of this section we will see how the search operators work and how their implementation in gene expression programming is a child's play due to the simple fact that the genome is completely autonomous and consequently is not tied up in the structural complexities of the computer programs encoded within.

Mutation

In gene expression programming, mutations can occur anywhere in the chromosome. However, the structural organization of chromosomes must remain intact; that is, in the heads any symbol can change into another (function or terminal), whereas in the tails terminals can only change into terminals. This way, the structural organization of chromosomes is preserved, and all the new individuals produced by mutation are structurally correct programs.

Consider the following three-genic chromosome:

```
01234567890120 1 2 3 4 56789012
0123456789 0 1 2
/ b Q a * * b b b b a b a - - + * Q -
abbbabaQ*a+**baabbba
```

Suppose a mutation changed the "*" at position 5 in gene 1 to "a"; the "-" at position 1 in gene 2 to "Q"; and the "a" at position 2 in gene 3 to "*". In this case the following chromosome is obtained:

```
01234 5 67890120 1 2 3 4567890
12012 3 4 56789012
/ b Q a * a b b b b a b a - Q + * Q -
abbbabaQ**+**baabbba
```

Note that if a function is mutated into a terminal or vice versa, or a function of one argument is mutated into a function of two arguments or vice versa, the expression tree is usually modified drastically. Note also that the mutation on gene 1 is an example of a neutral mutation, as it occurred in the non-coding region of the gene. It is worth emphasizing that the non-coding regions of GEP chromosomes are ideal places for the accumulation of neutral mutations, which are known to play an important role in evolution (Ferreira, 2002b; Kimura, 1983). In summary, in gene expression programming there are no constraints both in the kind of mutation and the number of mutations in a chromosome as, in all cases, the newly created individuals are syntactically correct programs.

Inversion

We know already that the modifications bound to make a big impact occur usually in the heads of genes. Therefore, the inversion operator was restricted to these regions. Here any sequence might be randomly selected and inverted.

In gene expression programming, the inversion operator randomly chooses the chromosome, the gene to be modified, and the start and termination points of the sequence to be inverted. It is worth pointing out that this is the first time the inversion operator is described in gene expression programming.

Consider, for instance, the following three-genic chromosome:

```
01 2 3 4 5 6 789012012 3 4
5 67890120123456789012
/+aQ*aaabaaab/aa/baaababab-
Q++aQababaab
```

Suppose that the sequence "aQ*" in gene 1 (positions 2-4) was picked up to be inverted. Then the following chromosome is formed:

```
012 3 4 5 6 7 8 90120123456789
012012 3 4 56789012
/+*Qaaaabaaab/aa/baaababab-
Q++aQababaab
```

It is worth pointing out that, since the inversion operator was restricted to the heads of genes, there is no danger of a function ending up in the tails and, consequently, all the new individuals created by inversion are syntactically correct programs.

Transposition and Insertion Sequence Elements

The transposable elements (also called transposons) of gene expression programming are fragments of the genome that can be activated and then jump to another place in the chromosome. In GEP there are three kinds of transposable elements: (1) short fragments with a function or terminal in the first position that transpose to the head of genes except the root (insertion sequence elements or IS elements); (2) short fragments with a function in the first position that transpose to the start position of genes (root IS elements or RIS elements); (3) and entire genes that transpose to the beginning of chromosomes.

IS Transposition

Any sequence in the genome might become an IS element and, therefore, these elements are randomly selected throughout the chromosome. A copy of the transposon is made and inserted at any position in the head of a gene, except the first position. The transposition operator randomly chooses the chromosome, the start and termination points of the IS element, and the target site. It is worth pointing out that the implementation of this operator as described here slightly differs from the original implementation (Ferreira, 2001), where the length of the IS elements was chosen *a priori*.

Consider the following chromosome composed of three genes, each with a head size of six:

```
01234567890120123456789012012345678
9012
+*+-Q/baaaabbQ+aa*abaaaaba*+-a/-
aabbbba
```

Suppose that the sequence "a/-" in gene 3 (positions 3-5) was picked up as an IS element to be then inserted between positions 1-2 in gene 2, obtaining:

```
01234567890120123456789012012345678
9012
+*+-Q/baaaabbQ+a/-abaaaaba*+-a/-
aabbbba
```

Note that, in this case, a perfect copy of the transposon appears at the site of insertion. Note also that a sequence with as many symbols as the IS element is deleted at the end of the head (in this case, the sequence "a*a" was deleted). Thus, despite this insertion, the structural organization of chromosomes is maintained and, therefore, all the new individuals created by IS transposition are syntactically correct programs.

Root Transposition

All root IS elements start with a function, and therefore must be chosen among the sequences of the heads. For that, a point is randomly chosen in the head and the gene is scanned downstream until a function is found. This function becomes the start position of the RIS element. If no functions are found, the operator does nothing.

The RIS transposition operator randomly chooses the chromosome, the gene to be modified, and the start and termination points of the RIS element. It is worth noticing that this operator is slightly different from the original RIS transposition (Ferreira, 2001), as the length of the transposon is randomly chosen by this simpler RIS transposition.

Consider the following chromosome composed of three genes, each with a head size of six:

```
01234567890120123456789012012345678 9012
01234567890120123456789012012345678 9012
/+b*Q/babababaQ*aQ*QaaababaQa*/
+abbbaaab
```

Note that during transposition, the whole head shifts to accommodate the RIS element, losing, at the same time, the last symbols of the head (as many as there are in the transposon). In this case, the sequence "+b*" was deleted and the transposon became only partially duplicated. As with IS transposition, the tail of the gene subjected to RIS transposition and all nearby genes remain unchanged. Note, again, that all the programs newly created by this operator are syntactically correct, as it also preserves the structural organization of the chromosome.

Gene Transposition

In gene transposition an entire gene works as a transposon and transposes itself to the beginning of the chromosome. In contrast to the other forms of transposition, in gene transposition, the transposon (the gene) is deleted at the place of origin.

The gene transposition operator randomly chooses the chromosome to be modified and then randomly chooses one of its genes (except the first, obviously) to transpose. Consider the following chromosome composed of three genes:

```
01234567890120123456789012012345678 9012
-ab+a-babaaaaQ+bab/babbbba*-*Q*-
abbabab
```

Suppose gene 3 was chosen to undergo gene transposition. In this case the following chromosome is obtained:

```
012345678901201234567890120123456
7 8 9 0 1 2
*-*Q*-abbabab-ab+a-babaaaaQ+bab/
babbbba
```

Apparently, gene transposition is only capable of shuffling genes and, for sub-ETs linked by commutative functions, this contributes nothing to adaptation in the short run. Note, however, that when the sub-ETs are linked by a non-commutative function, the order of the genes matters and, in this case, gene transposition becomes a macromutator. However, gene transposition becomes particularly interesting when it is used in conjunction with recombination, for it allows not only the duplication of genes but also a more generalized shuffling of genes or smaller building blocks.

Recombination

In gene expression programming there are three kinds of recombination: one-point recombination, two-point recombination, and gene recombination. In all types of recombination, two chromosomes are randomly chosen and paired to exchange some material between them, creating two new daughter chromosomes.

One-Point Recombination

In one-point recombination the parent chromosomes are paired and split up at exactly the same point. The material downstream of the recombination point is afterwards exchanged between the two chromosomes.

Consider the following parent chromosomes, each composed of three genes:

```
0123456 7 8 9 0 1 201234 5 6
7 8 90120123456 7 8 9 012
*b-Qb /
a b a a bQ* * Q + * bbaaabb*Q-
- Q Q a ab b bbb
```

```
-/bQa+aabbbba/Q*b/aababaaa-/a/a/abaaabb
```

Suppose bond 4 in gene 2 (between positions 3 and 4) was randomly chosen as the crossover point. Then, the paired chromosomes are both cut at this bond, and exchange between them the material downstream of the crossover point, forming the offspring below:

```
012345678901201234567890120 1 2 3 4 56789012
*b-Qb/aaabaabQ**Q/aababaaa-/a/a/abaaabb
-/bQa+aabbbba/Q*b+*bbaaabb*Q--QQa-abbbbb
```

It is worth emphasizing that GEP chromosomes can cross over any point in the genome, continually disrupting old building blocks and continually forming new ones. Furthermore, due to both the multigenic nature of GEP chromosomes and the existence of non-coding regions in most genes, entire genes and intact open reading frames can be swapped between parent chromosomes. Thus, the disruptive tendencies of one-point recombination (splitting of building blocks) coexist side by side with its more conservative tendencies (swapping of genes and ORFs), making one-point recombination (and of course two-point recombination too) a very well balanced genetic operator. Furthermore, like all the other recombinational operators, when one-point recombination is used together with gene transposition, it is also capable of duplicating genes.

Two-Point Recombination

In two-point recombination two parent chromosomes are paired and two points are randomly chosen as crossover points. The material between the recombination points is afterwards exchanged between the parent chromosomes, forming two new daughter chromosomes.

Consider the following pair of recombining chromosomes:

```
01234 5 6 7 8 9 0 1 20 1 2 3 45678 9 0 1 2 012345 6 7 8 9 012
bb * a a b a a a aQ*a/ b + b b b a a a b Q + / a a + b abaabb
+-Qa/Qaabbaba+Q-+/+abbbaaa//+-/+bababab
```

Suppose bond 7 in gene 1 (between positions 6 and 7) and bond 4 in gene 3 (between positions 3 and 4) were chosen as crossover points. Then, the following daughter chromosomes are created:

```
012345 6 7 8 9012012 3 4 5 67 890120123456789012
/bQbb*aabbaba+Q-+/+abbbaaa//+-a+babaabb
+-Qa/QaabaaaaQ*a/b+bbbaaabQ+/a/+bababab
```

It is worth emphasizing that two-point recombination is more disruptive than one-point recombination in the sense that it recombines the genetic material more thoroughly, constantly destroying old building blocks and creating new ones. But like one-point recombination, two-point recombination has also a conservative side and it is good at swapping entire genes and open reading frames. And, as observed for one-point recombination, two-point recombination can also give rise to duplicated genes if it were used together with gene transposition.

Notwithstanding, if the goal is to evolve good solutions, one-point or two-point recombination should never be used as the only source of genetic variation, as they tend to homogenize populations (Ferreira, 2002c). However, together with mutation, inversion and transposition, these operators are an excellent source of genetic variation and are more than sufficient to evolve good solutions to virtually all problems.

Gene Recombination

In the third kind of GEP recombination, entire genes are exchanged between two parent chromosomes, forming two daughter chromosomes containing genes from both parents. The exchanged genes are randomly chosen and occupy exactly the same position in the parent chromosomes. Consider the following parent chromosomes:

```
012345 6 7 8 9 0 1 20123456
7 8 9 0120 1 2 3 456789012
/+*--b b a a a b a b*+b-
a a b Q **+*b a b a bbab
Q//b-baababaa/ab/QQbaababaQ*+a++bbaaaaa
```

Suppose gene 2 was chosen to be exchanged. In this case the following offspring is formed:

```
0123456 7 8 9 012012345678901201
23456789012
/ + * - - b b a a a b a b / a b /
QQbaababaQ**+*bababbab
Q//b-baababaa*+b--aabaaaabQ*+a++bbaaaaa
```

Note that, with this kind of recombination, similar genes can be exchanged but, most of the times, the exchanged genes are very different and new material is introduced in the population.

It is worth emphasizing that this operator is unable to create new genes: the individuals created by gene recombination are different arrangements of existing genes. Obviously, if gene recombination were used as the unique source of genetic variation, more complex problems could only be solved using very large initial populations in order to provide for the necessary diversity of genes. However, GEP evolvability is based not only in the shuffling of genes (achieved by gene recombination and gene transposition), but also in the constant creation of new genetic material, which is carried out essentially by mutation, inversion and transposition (both IS and RIS transposition) and, to a lesser extent, by recombination (both one-point and two-point recombination).

EVOLVING COMPUTER PROGRAMS FOR DIAGNOSING BREAST CANCER

In this section we are going to use gene expression programming to design a computer program that can be used to diagnose breast cancer. The data sets used both for training and testing were obtained from PROBEN1 — a set of neural network benchmark problems and benchmarking rules (Prechelt, 1994). Both the technical report and the data sets are available through anonymous FTP from Neural Bench archive at Carnegie Mellon University (machine *ftp.cs.cmu.edu,* directory */afs/cs/project/connect/bench/contrib/prechelt*) and from the machine *ftp.ira.uka.de* in directory */pub/neuron*. The file name in both cases is *proben1.tar.gz*.

In this diagnosis task the goal is to classify a tumor as either benign (0) or malignant (1) based on nine different cell analyses (input attributes or terminals).

The model presented here was obtained using the *cancer1* data set of PROBEN1, where the binary 1-of-*m* encoding in which each bit represents one of *m*-possible output classes was replaced by a 1-bit encoding ("0" for benign and "1" for malignant). The first 350 samples were used for training and the last 174 were used to test the performance of the model in real use. This means that absolutely no information from the test set samples or the test set performance are available during the adaptive process. Thus, both the classification accuracy and classification error on the test set can be used to evaluate the generalization performance of the evolved models.

For this problem a very simple function set was chosen, composed only of the four arithmetic operators, that is, F = {+, -, *, /}, where each function was weighted five times; the set of terminals consisted of the nine attributes used in this problem and were represented by T = {d0, ..., d8} which correspond, respectively, to clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli, and mitoses.

In classification problems where the output is often binary, it is important to set criteria to convert real-valued numbers into 0 or 1. This is the 0/1 rounding threshold R that converts the output of a chromosome into "1" if the output is equal to or greater than R, or into "0" otherwise. For this problem we are going to use $R = 0.1$.

The fitness function used to evaluate the performance of each candidate model is very simple and is based on the number of samples correctly classified. Thus, the fitness f_i of an individual program i corresponds to the number of hits and is evaluated by the formula:

$$\text{if } n > C, \text{ then } f_i = n; \text{ else } f_i = 0 \quad \textbf{(10)}$$

where n is the number of sample cases correctly evaluated, and C is the number of samples in the class with more members (predominant class).

As it is customary in genetic programming (Koza, 1992) and gene expression programming (Ferreira, 2001), the parameters used per run are summarized in a table (Table 1). Note that, in this case, a small population of 50 individuals and chromosomes composed of three genes with an $h = 8$ and sub-ETs linked by addition were used. The program below was discovered after 423 generations (genes are shown separately and a dot is used to separate each element):

```
*.+.d0.*.*.*.+.*.d7.d0.d8.d3.d5.d4.d3.d5.d6
*.+.*.d0.*.*.*.-.d5.d4.d1.d5.d1.d0.d2.d1.d1
*.d5.*.+.d6.*.+.d5.d2.d2.d1.d1.d7.d0.d7.d4.d0
```
(11a)

It has a fitness of 340 evaluated against the training set of 350 fitness cases and maximum fitness on the test set of 174 examples. This means that this model is very good indeed, with a classification accuracy of 100% and a classification error of 0% in the test set. In the training set a classification accuracy of 97.14% and a classification error of 2.86% were obtained.

Note that for the expression of chromosome (11a) to be complete the sub-ETs must be linked by addition and the 0/1 rounding threshold must be taken into account. With the software APS 3.0 by *Gepsoft*, the model (11a) above can be automatically converted into a fully expressed computer program or function, such as the C++ function below:

```
int apsModel(double d[])
{
    const double ROUNDING_THRESHOLD = 0.1;
    double dblTemp = 0;
    dblTemp = ((((d[0]*d[8])*(d[3]+d[5]))+((d[4]*d[3])*d[7]))*d[0]);
    dblTemp += ((d[0]+((d[0]-d[2])*d[5]))*((d[4]*d[1])*(d[5]*d[1])));
    dblTemp += (d[5]*(((d[5]*d[2])+(d[2]+d[1]))*d[6]));
    return (dblTemp >= ROUNDING_THRESHOLD ? 1:0);
}
```
(11b)

Similarly, all models evolved by gene expression programming can be immediately converted into virtually any programming language through the use of grammars, including the universal representation of parse trees (Figure 7). These trees can then be used to grasp immediately the mathematical intricacies of the evolved models and therefore are ideal for extracting knowledge from data. As can be clearly seen in Figure 7, all the cell analyses seem to be relevant to an accurate diagnosis of breast cancer. This is, indeed, one of the great advantages of gene expression programming: the possibility of extracting knowledge almost instantaneously, as the models evolved

Table 1. Settings used in the breast cancer problem

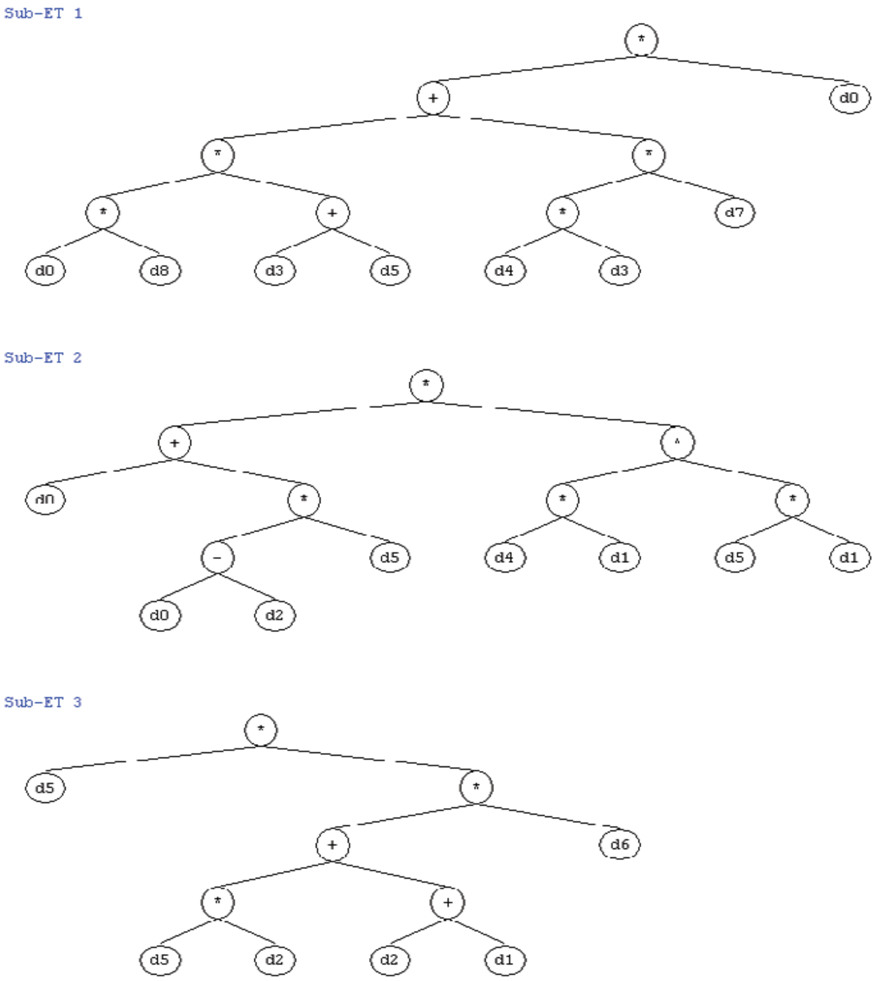
Number of generations	500
Population size	50
Number of training samples	350
Number of testing samples	174
Function set	(+-*/)5
Terminal set d	0 - d9
Rounding threshold	0 .1
Head length	8
Number of genes	3
Linking function	+
Chromosome length	51
Mutation rate	0.044
Inversion rate	0 .1
IS transposition rate	0.1
RIS transposition rate	0.1
One-point recombination rate	0.3
Two-point recombination rate	0.3
Gene recombination rate	0.1
Gene transposition rate	0.1

by GEP can be represented in any conceivable language, including the universal diagram representation of expression trees.

CONCLUSION

In this chapter the details of implementation of gene expression programming were thoroughly explained, giving other researchers the possibility of implementing it themselves. Furthermore, this new algorithm was summarily compared to genetic algorithms and genetic programming in order to bring into focus the fundamental differences between the three techniques and, consequently, enable readers to appreciate the advantages a full-fledged genotype/phenotype system brings into evolutionary computation. In addition, the classification task solved in this work clearly demonstrates the modeling prowess of this new technique: The compact computer programs evolved by gene expression programming in its native Karva code can be immediately used to generate highly sophisticated computer programs in virtually any programming language through the use of grammars as is already done in commercially available software.

Figure 7: The sub-ETs of the model (11b) evolved by gene expression programming to diagnose breast cancer. (Expression trees drawn by Gepsoft APS 3.0.)



REFERENCES

Cramer, N.L. (1985). A representation for the adaptive generation of simple sequential programs. In J.J. Grefenstette (Ed.), *Proceedings of the First International Conference on Genetic Algorithms and their Applications.* Erlbaum.

Dawkins, R. (1995). *River out of Eden.* Weidenfeld and Nicolson.

Ferreira, C. (2001). Gene expression programming: A new adaptive algorithm for solving problems. *Complex Systems*, *13*(2), 87-129.

Ferreira, C. (2002a). *Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence*. Portugal: Angra do Heroísmo.

Ferreira, C. (2002b). Genetic representation and genetic neutrality in gene expression programming. *Advances in Complex Systems, 5*(4), 389-408.

Ferreira, C. (2002c). Mutation, transposition, and recombination: An analysis of the evolutionary dynamics. In H.J. Caulfield et al. (Eds.), *Proceedings of the 6th Joint Conference on Information Sciences, 4th International Workshop on Frontiers in Evolutionary Algorithms* (pp. 614-617). Research Triangle Park, NC.

Friedberg, R.M. (1958). A learning machine: Part I. *IBM Journal, 2*(1), 2-13.

Friedberg, R.M., Dunham, B., & North, J.H. (1959). A learning machine: Part II. *IBM Journal, 3*(7), 282-287.

Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison-Wesley.

Holland, J.H. (1975). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. University of Michigan Press.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge, UK: Cambridge University Press.

Koza, J.R. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: MIT Press.

Prechelt, L. (1994). PROBEN1 – A set of neural network benchmark problems and benchmarking rules. *Technical Report* 21/94, Univ. Karlsruhe, Germany.

This work was previously published in Recent Developments in Biologically Inspired Computing, edited by L.N. de Castro and F.J. Von Zuben, pp. 82-103, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.23
Checking Female Foeticide in the Information Age

Chetan Sharma
Datamation Foundation Charitable Trust, India

Divya Jain
Datamation Foundation Charitable Trust, India

INTRODUCTION

In India, the practice of sex-selective abortion or female foeticide (in which an unborn baby is aborted or killed before birth simply because it is not a boy) is only the latest manifestation of a long history of gender bias, evident in the historically low and declining population ratio of women to men. Moreover, the medical fraternity in India has been quick to see entrepreneurial opportunities in catering to insatiable demands for a male child. Until recently, the technology was prohibitively expensive. The three chief pre-natal diagnostic tests being used to determine the sex of a foetus (sexing) are amniocentesis, chronic villi biopsy (CVB) and ultrasonography. Amniocentesis is meant to be used in high-risk pregnancies, in women older than 35 years. CVB is meant to diagnose inherited diseases like thalassaemia, cystic fibrosis and muscular dystrophy. Ultrasonography is the most commonly used technique. It is non-invasive and can identify up to 50% of abnormalities related to the central nervous system of the foetus. But sexing has become its preferred application. A ban on the government departments at the center and in the states, making use of pre-natal sex determination for the purpose of abortion a penal offence, led to the commercialization of the technology; private clinics providing sex determination tests through amniocentesis multiplied rapidly and widely. These tests are made available in areas that do not even have potable water, with marginal farmers willing to take loans at 25% interest to have the test. Advertisements appear blatantly encouraging people to abort their female foetuses to save the future cost of dowry. The portable ultrasound machine has allowed doctors to go from house to house in towns and villages. In a democracy, it is difficult to restrict rights to business and livelihood if the usual parameters are fulfilled. An argument by Rathee (2001) brings to light the fact that the recent technological developments in medical practice combined with a vigorous pursuit of growth of the private health sector have led to the mushrooming of a variety of sex-selective services. This has happened not only in urban areas, but deep within rural countryside, also—areas where the other dimensions of healthcare and development are yet to penetrate. Indeed, the indications are that given these lethal combinations, the phenomenon of sex-selective abortions is growing nationwide. Furthermore, these discriminatory services are being provided and projected in the name of "democratic choice" as a measure of "upliftment" of women, since they are being saved from dowry deaths, burning and other forms of torture and violence they would have undergone once they were born. This pure greed for money is also equated by a large section of doctors to "people's demand."

BACKGROUND

In an age when females have left no stones unturned in almost every field, there are people who still accord a lower status to women. "What's wrong with a girl child? In this era of gender equality, why this sexual discrimination?" In some of our Indian societies, while a childless woman is perceived as incomplete, one who has given birth to daughters is partially complete. Only the one who has produced a son enjoys a status of sorts. The problem is intimately related to the institution

Figure 1.

of dowry. "If it's a girl child, we will have to spend first on her education and then on her marriage and dowry ... It doesn't stop there. We will also have to meet some of her expenses after marriage. How I wish I get a son!" The bias against females is also related to the fact that sons are looked at as a type of insurance. Even our religions have been prejudiced against women. According to Manu (2000 BC-2000 AD), a woman has to be reborn as a man to attain moksha (redemption). A man cannot attain moksha unless he has a son to light his pyre. Also, it says a woman who gives birth to only daughters may be left in the 11^{th} year of marriage. This obviously shows the gender bias in our male-dominated, patriarchal society.

FEMALE FOETICIDE AND LAW

There has been an inability to discuss the issue of foeticide without the larger debate on abortion, which is legally allowed and has been seen as a triumph of the women's rights movement in the country. India has allowed abortion on broad medical and social grounds since the Medical Termination of Pregnancy (MTP) Act was passed in 1971, further trivializing the issue of foeticide. The Pre-Conception and Pre-Natal Diagnostic Techniques (Prohibition of Sex Selection) Act and Rules 1994 (as amended up to 2002) (the PCPNDT Act), a result of determined action by NGOs against grossly unethical medical terminations of healthy pregnancies, mandates that sex selection by any person, by any means, before or after conception, is prohibited. But while the Act seeks to regulate and prevent misuse of pre-natal diagnostic techniques, it rightly cannot deny them, either. The PCPNDT Act allows pre-natal diagnosis only for chromosomal abnormalities, genetic metabolic disorders and congenital abnormalities. The law, however, permits ultrasound clinics, clinics for medical termination of pregnancies and assisted reproductive facilities as a routine matter and legitimate business. Then there is the legally binding Code of Medical Ethics, constituted by the Indian Parliament in the Medical Council Act (1956) that many doctors conveniently ignore. Doctors are legally bound to report medical malpractice. The PCPNDT Act mandates that any person conducting ultrasonography or any other pre-natal diagnostic technique must maintain proper records. The Act requires use of a written form, duly signed by the expectant mother, as to why she has sought diagnosis. Violations are punishable by imprisonment and a fine. The law also permits abortions for failure of contraception. It is a huge challenge for the government to detect violations of the PNDT Act, since it is a crime of collusion and by consensus.

The Ground Reality

Female foeticede is now more widespread in the country than ever before. The practice that was restricted to a few states a few years ago has now spread all over the country. Not only the poor, but even the middle classes and the rich in India are biased against the girl child and women. The provisional figures of Census[1] 2001, which have frightened even the Health Department, show that the Child Sex Ratio (CSR, the number of girls in the age group of 0 to 6 years per 1,000 boys of the same age group) has declined sharply from 945 females per 1000 males in 1991 to 927 females. Furthermore, the CSR is actually worse than the national average in Himachal Pradesh, Punjab, Chandigarh, Uttaranchal, Haryana, Delhi, Rajasthan, UP, Gujarat, Maharashtra and Daman and Diu. The Overall Sex Ratio (OSR, which looks at the ratio of all females to males, not just those in the 0-6 age group) is even worse in places like Chandigarh (773:1000), Delhi (821:1000) and Punjab (874:1000). Health experts say these are some of the lowest sex ratios in the world. Here is a paradox. During the final decade of the last millennium, the CSR recorded its highest fall, dropping 18 points from 945 to 927. However, during the same decade, the OSR recorded its highest gain, rising 6 points from 927 to 933. In every census since 1951, the two ratios have gone more or less hand in hand, dropping steadily except in the late 1970s (see Table 1). The sudden fall in the number of girls in the youngest age group is believed to be proof of the increased incidence of sex-selective abortions or female foeticide. Most of these abortions are the result of the misuse of sex determination technologies, such as ultrasound scanning and amniocentesis.

What is a woman's role in all this? Whether in choosing her spouse, contraception methods and the first pregnancy, place of delivery and so forth, does her choice or decision really matter at all? Or is it really possible for a woman to decide about having an abortion, if she has to survive as a daughter-in-law in the family? Let's assume her husband supports the birth of a female child; even then, she will prefer to have a male child to get respect from her in-laws family, in order to save her marriage.

- Mrs. A from Jalandhar district was married in Chandigarh. Both the husband and wife are highly educated (post-graduates). The husband earns a salary of Rs.4,000 per month. At present they live in a joint family consisting of the father-in-law, mother-

Table 1. CSR and OSR (Source: Census of India)

	1951	**1961**	**1971**	**1981**	**1991**	**2001**
CSR	983	976 (-7)	964 (-12)	962 (-2)	945 (-17)	927 (-18)
OSR	946	941 (-5)	930 (-11)	934 (+4)	(927-7)	933 (+6)

> in-law, brother-in-law and his wife. The brother-in-law has a 13-year-old only son. All the members are highly qualified—one of the brothers-in-law and his wife are doctors and live separately. Mrs. A has one daughter. She conceived a second time after keeping a conscious spacing of three years. In the fifth month of pregnancy, she went for amniocentesis and found the foetus to be a female. She did not reveal this to her in-laws, but secretly, with her husband's consent, she underwent an abortion. The researcher was able to contact her before the event, and found Mrs. A to be depressed and pale with fear and the guilt of having conceived a daughter. Sobbing, she said that she was ashamed of having a female foetus. While her husband did not mind having a daughter again, she herself wanted to have only two children, and to earn the respect of the family, she presumed that it is important to have a son. What a dilemma!

If a woman doesn't have any say in this matter, don't you think that this forced female foeticide is an act of violence against women? So where is the cure of this disease? With law? Doctors in India believe 2 million foetuses are killed every year through abortion, simply because they are female. Although it is an illegal practice, with the increasing availability of sex determination tests, it is impossible to keep track of such cases. There is little doubt that in India the PCPNDT Act of 1994 has not been very effective. The facts revealed by the census speak for themselves. We need, rather, to attempt more broad-reaching strategies that will address the economic and cultural roots of the problem. What is required is that the available legislation for prevention of sex determination needs strict implementation, alongside the launching of programs aimed at altering attitudes, including those prevalent in the medical profession. In rural areas, as the number of marriageable women declines, men would tend to marry younger women, leading to a rise in fertility rates and thus a high rate of population growth. The abduction of girls is an associated phenomenon. It was recently reported in *The Hindustan Times* that young girls from Assam and West Bengal are kidnapped and sold into marriage in Haryana. It is only by a combination of monitoring, campaigning and effective legal implementation that the deep-seated attitudes and practices against women and girls can be eroded.

Girls and women not only face inequity and inequality, they are even denied the right to be born if their families do not wish so. In fact, many families do not wish their women folk to deliver baby daughters. Confronted with this situation, it is high time to take preventive measures against female foeticide. Both local leadership and government agencies should plan a concerted long-term program aimed at the educational and socio-economic advancement of the community—for example, by social welfare and poverty alleviation programs—to improve the economic status of women. To wean people away from the traditional practices, voluntary organizations should come forward to promote social education and awareness among the people through cultural programs, public debate and so forth. There is a need for sustained campaigning and active monitoring of the act. State governments should realize the importance and priority of the law and not merely treat it with their usual complacency. Structures for implementation of the 1994 law need to be created at the district level. Volunteers have to be actively mobilized to monitor registration and functioning of sex determination clinics at different districts. Cases have to be filed against the violators and social consciousness has to be raised against the crime. Members of society and religious leaders have a positive role in creating a morally reformed society. The long-term task is to foster a culture of goodness and human dignity that inoculates individuals and institutions against the infection of this despicable human practice.

INTERVENTIONIST ROLE OF INFORMATION AND COMMUNICATION TECHNOLOGIES

Progress of science and technology is mandatory for the progress of a nation, but what matters most is its manifestation and beneficial application. Today, information technology has changed the communication paradigm, making it no longer difficult to reach a large number of people more or less at the same time; and that, too, enables them to respond, interact as well as obtain a copy of the information within a low-cost. information and communication technologies (ICTs), apart from making people aware of this heinous crime, can also play a highly interventionist role by pro-actively pursuing cases against erring doctors, booking them under the law of the land as well as helping people in general change their opinion about a "girl child." An important tool helping the government of India accomplish their cherished goal of altogether stopping female foeticide, set up by Datamation Foundation Charitable Trust, is a Web site solely dedicated to Female Foeticide (www.indiafemalefoeticide.org). This major ICT-based campaign and advocacy program is to help prevent occurrences of sex-selective tests and selective abortions of female foetuses in India. The purpose is to mobilize the population that apparently wants female foeticide, a medical system happy to supply the necessary technology, and a section of feminists arguing that female foeticide is about reproductive rights and choices.

FUTURE TRENDS

Online Attempts to Check Female Foeticide

The India Female Foeticide portal includes some distinctive features. The "Femicide" section provides some rich background information on the prevalence and practice of Female Foeticide, including reasons for the sex-selective tests. The site not only covers the regulatory aspects—PNDT Act of 1994 and the Medical Council of India's code of conduct to crack down sex selection – but also includes a complaint-lodging process in the best tradition of e-governance. The complaint-lodging process is a very critical feature, as it protects the identity of the complainant and yet provides an effective vehicle for booking the doctor, maternity home, ultrasound clinic or radiology clinic. The complaints are retrieved into a database format at Datamation, where they are handed over to the central PNDT Cell (Ministry of Health and Family Welfare) and other respective state and district authorities for re-addressal. To date, the Foundation has received about 580 complaints, and has passed these to respective authorities. The interpretation of the law in form of demographic data is also put on the Web site. Another important platform for the Web site is the "Pledge Support" page, which highlights two features—Pledge Support and Information. Through the "pledge support" feature, one can enter information regarding the type of volunteer service the person or the organization is ready to offer, and the "information" option allows one to enter information about any ultra-sound clinics, doctors, radiologists and so forth to enable database building.

Serving as a knowledge repository on the issue is the "What's New" page. The attempt here is awareness generation and capacity building both among the community and the stakeholders. Through e-mails, to date the Foundation has succeeded in spreading awareness among 8 million people worldwide on this issue. The "Links/Resources "page contains the contact addresses of different NGOs and agencies working for the prevention of female foeticide, and gives data for state-wise, sex ratio and female foeticide and sex selection. To enable awareness generation through sound and visual media, radio and video links have also been added to the Web site. An important feature started under this initiative is the e-Newsletter, through which the Foundation disseminates news, articles, reports and other studies on a weekly basis to make the various segments (students, NGOs, medical community, educationalists, government officials, researchers) of the society aware about issues of gender inequality, female foeticide and sex-selective abortions. "Youth Voices," a newly started section on the portal, is an effort to motivate and encourage youth participation in eradicating this menace and spreading awareness among other groups. To enable communication on this issue among various segments of society, we have started a discussion forum that allows free-flow and exchange of ideas among people.

CONCLUSION

Fully understanding that an evil such as this cannot be addressed in isolation, we are also closely examining related social malaise, such as dowry, women's underemployment and exploitation in society, education standards of the girl child as well as high school dropouts among girls, early marriages and the arranged marriage system. It is our endeavor to develop sustainable development models for each of the above-listed social malaise in India so that these have an impact on improving the ratio of females in the Indian society.

REFERENCES

Balakrishnan, R. (1994). The social context of sex selection and the politics of abortion in India. *Harvard School of Public Health*, 267-286.

Dani, A. (2004). Death in the womb sex selection law fails to check foeticide.

Grewal, I., & Kishore, J. (2004). *Female foeticide in India*.

Gurung, M. (2004). The two-child norm only leads to female foeticide. *InfoChange News & Feature*.

Kaur, M. *Harvard School of Public Health*.

Kaur, M. (1993). Sociological perspective of female foeticide. *The Journal of Family Welfare, 39*(1), 40-43.

Kaur, P. (2002). Abortion vs. selective conceiving or heading toward imbalanced society? *South Asian Women's Forum*.

Nair, P. (2004). Sex ratio: The hidden horrors. *India Together*.

Rathee, M. (2001). Eradicate scourge of female foeticide, *People's Democracy*, *XXV*(39).

Sarna, K. (2003). Female foeticide on the rise in India. *Nursing Journal of India*.

SPK. (2004). Why this gender bias towards womankind?

Sridhar, L. (2004). Female foeticide: The collusion of the medical establishment. *InfoChange News & Features*.

Stein, R. (2004). A boy for you, a girl for me: Technology allows choice. *Washington Post*, A01.

KEY TERMS

Child Sex Ratio (CSR): The number of girls in the age group of 0 to 6 years per 1,000 boys of the same age group.

Dowry: Money or property brought by a woman to her husband at marriage.

Female Foeticide: Practice in which an unborn baby is aborted or killed before birth simply because it is not a boy.

Manu: The primordial father of the human race and sovereign of the earth who first instituted religious ceremonies and devised a code of laws.

Moksha: In Hinduism, liberation and bliss (nirvana) from the cycle of rebirth (samsara); also called *mukti* or *moksha*.

Overall Sex Ratio: The ratio of all females to males, not just those in the 0-6 age group.

ENDNOTE

[1] When we talk of the Indian Census, it is important to understand that our sex ratio is defined as the number of females per 1,000 men, unlike the West, where sex ratio is the complete opposite and defined as the number of males per 1,000 females.

This work was previously published in the Encyclopedia of Gender and Information Technology, edited by E. Trauth, pp. 90-95, copyright 2006 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 7.24
Integrated Process and Data Management for Healthcare Applications

Stefan Jablonski
University of Bayreuth, Germany

Rainer Lay
ProDatO Integration Technology GmbH, Germany

Christian Meiler
ProDatO Integration Technology GmbH, Germany

Matthias Faerber
University of Bayreuth, Germany

Bernhard Volz
University of Bayreuth, Germany

Sebastian Dornstauder
University of Bayreuth, Germany

Manuel Götz
University of Bayreuth, Germany

Sascha Müller
University of Erlangen-Nuremberg, Germany

ABSTRACT

Healthcare applications are demanding with respect to their IT needs. Due to their complexity they need strong IT support which must provide enough freedom to their users in the medical domain because of the high degree of flexibility required by them. We have chosen a process based technology, called data logistics, which can cope with these almost conflicting requirements. Data logistics is based on clinical processes. However, when data logistics processes are executed, this happens transparently to the users and provides them a huge degree of freedom in determining the sequence of process execution. In order to cope with these challenging features, the process execution environment for data logistics processes is generated according to our model driven development approach called process driven architecture (PDA). Several clinical studies illustrate

the feasibility, effectiveness, and efficiency of the data logistics approach. This article both exposes the rationale of the data logistics approach and provides a deep inside into the implementation concepts.

INTRODUCTION

General Factors

The medical application domain has been a great challenge for information technology solutions for decades (Anderson, 1997; Berg & Toussaint, 2003; Lenz, Elstner, Siegele, & Kuhn, 2002) and is sometimes even considered as the "killer application" in the field of process support (Dadam, Reichert, & Kuhn, 2000). This estimation is based on the high complexity of the medical domain (Anderson, 1997; Lenz et al., 2002), which results from multiple factors. According to Anderson, J. G. (1997), Berg, (1999), and Kuhn, Giuse, Bakker, Ball and Gel (2001), these factors can be divided into two groups: technical and socio organizational.

Technical Factors

There are several technical factors that lead to high complexity. First, medical processes are by nature not static and thus require high flexibility during execution (Dadam et al., 2000). Depending on the current health condition of a patient, various treatments must be made possible (Dadam et al., 2000). Consequently, the challenge during process execution is "not to restrict the physician or the nurse" (Dadam et al., 2000), but to silently assist and support decisions of the medical staff. Second, the underlying process models need to be quickly adaptable to new requirements, as medical processes change constantly over time (Lenz & Kuhn, 2004). Typical reasons for that are new medical research findings and organizational or juridical amendments. An organically grown heterogeneous process landscape is the third factor; this heterogeneity leads to isolated integration islands. Processes differ in their characteristics like complexity, duration, number of actors, and so forth. Dadam et al. (2000) require either sophisticated or radical integration solutions.

Socio Organizational Factors

The unique combination of several socio-organizational factors is typical for the medical application domain. First, clinical routine is characterized by high workload and continuous high demands on all process participants. This is not surprising, as the medical staff is "responsible for many patients and they have to provide an optimal treatment process for each of them" (Dadam et al., 2000). Second, the process participants differ in their knowledge about the clinical applications. So, "cooperation between organizational units as well as the medical personnel [becomes] a vital task" (Dadam et al., 2000) for the clinical application domain. Third, the complexity of processes is very high, which results from the numerous and sophisticated dependencies within and between processes. This complexity is hard to manage for all process members. Last but not least, IT is often considered critically by the medical staff. According to Anderson (1997), "physicians have been unwilling to change their traditional practice of using the paper medical record." Anderson (1997) also notes that physicians accept information systems that allow for an improved patient treatment process but resist against information system that try to automate medical decisions. By involving process participants in information system development at an early stage, it can be assured that the realization of a medical process complies with their way of organizing work. This early involvement can increase IT acceptance (Anderson, 1997; Berg, 1999; Handler, Feied, Coonan, Vozenilek, Gillam, Peacock et al., 2004; Lenz, Buessecker, Herlofsen, Hinrichs, Zeiler, & Kuhn, 2005). We will now replenish and concretize the above requirements by experiences deduced from our research projects.

Analyzing the Clinical Application Domain

The results of this article have been acquired in two projects founded by the German Research Foundation:

1. Our sub-project C.5 of the Collaborative Research Center ("Sonderforschungsbereich") 539 "Glaucoma Including Pseudo-exfoliation-Syndroms" (SFB 539. (2007) deals with intelligent linking and adaptation of IT systems in order to improve the quality of the treatment of glaucoma patients. The research work is characterized by both clinical routine and clinical research. Consequently, an adequate information system has not only to support patient treatment but also the clinical research work which is split into two scientific goals. The first goal is the support of long running research studies based on quality assured data. The second goal is the development of an integrated semimobile workplace for early diagnosis of the glaucoma disease.
2. Our second project, "Fundamentals of Process Oriented Information Systems for Integrated Healthcare Networks," has been driven by the demand for process-oriented information systems (IS) in healthcare networks. IS have to support the treatment of patients according to clinical pathways (Beyer, Kuhn, Meiler, Jablonski & Lenz, 2004). Independent from their close relationship to IT, clinical pathways provide a concept to optimize the treatment of a group of patients with a similar diagnosis. Pathways are initiated, designed, and consented by a healthcare provider, for example, a hospital, and describe the treatment process considering both evidence-based medical guidelines and the organization of the institution. Clinical pathways need to be supported by medical IS.

In an overall perspective, medical IS have to meet the general requirements discussed in the last section but also have to bring optimal patient treatment and efficient support of clinical research work into a perfect balance. So the daily medical work is covered by three main areas (see Figure 1):

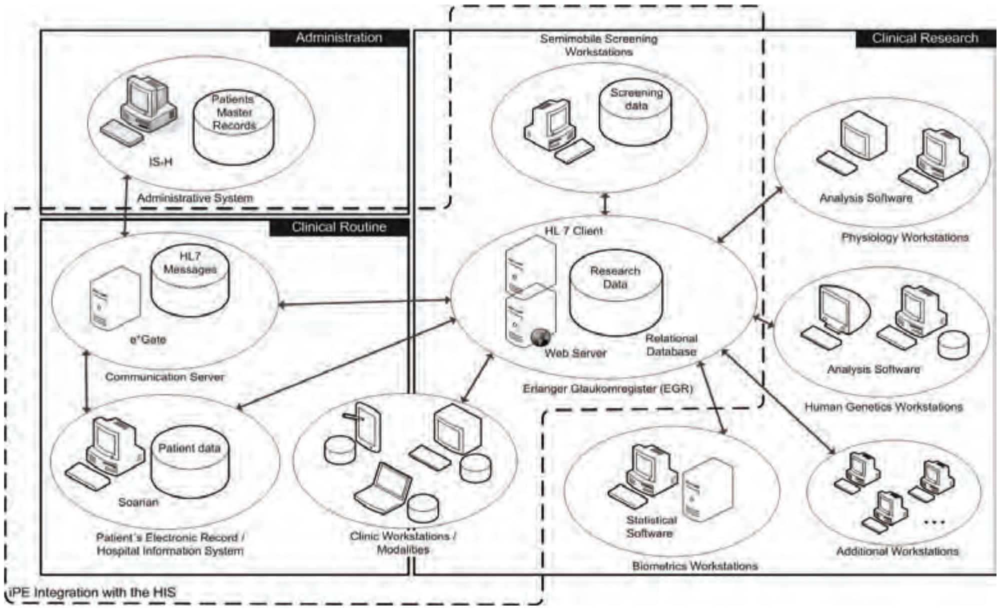
1. **Patient treatment:** Patients must be treated according to newest medical knowledge standardized by medical guidelines and clinical pathways in order to achieve optimal results.
2. **Clinical research:** Medical knowledge must be enhanced by clinical research mostly based on clinical studies.
3. **Administrative tasks:** Medical work is accompanied and supported by many administrative tasks to ensure correct factoring based on the concept of diagnosis related groups (DRGs).

Figure 1 shows a typical scenario based on these three areas as deduced from our research projects. The details shall not be discussed here, as the complexity is obvious.

Formalized Problem Analysis

Having identified concrete problem areas, we will formalize the main IT-related challenges and requirements from a technical point of view. A general problem statement derived from the observations of the two previous subsections boils down to two major issues. On the one hand, the complexity of healthcare applications has to be coped with. We foster a process based solution in this article. On the other hand, the complex healthcare applications are heavily depending on data that have to be shared between them.

Figure 1. Overview of our project setting: Clinical routine, clinical research and administration



Again, we foster a process based solution which is especially tailored to data integration problems. In the following we analyze these problem areas by first starting with the investigation of the data integration problem.

In general there are many different scenarios for data integration. The ideal setting is that all applications access a central data store, preferably a database system. However, not all applications—especially not healthcare applications—can follow this (simplified and naive) model. Figure 2 depicts a typical scenario of a healthcare application, presented in an abstract form. There are a number of so-called modalities (MOD_x), that is, technical equipment like a digital camera for making pictures of the papilla (e.g., KOWA fundus camera) or a laser scanning device (e.g., Heidelberg Retina Tomograph 2 [HRT2]). Typically, data created at a certain modality (e.g., images) are of different quality and structure (format); both are legacy and often not standardized. Also, data are first stored locally at the modalities themselves; these local data stores are here called $DB\text{-}MOD_x$.

Besides modalities, various information systems (IS_y) are deployed in clinical applications, such as a hospital information system (HIS). Also these information systems maintain their proprietary data management facilities; we call them $DB\text{-}IS_y$. Typically, data created at modalities have to be integrated into these information systems. However, data exchange among information systems is also necessary, and data also have to be transferred from information systems to modalities (e.g., patient identifications). Thus, from an abstract perspective, data have to be exchanged between and among modalities and information systems, respectively. Hence, we generalize modalities and information systems to (the term) "healthcare applications" (APPs) since we are going to develop a general data integration mechanism and can therefore abstract from concrete systems.

Data exchange between APPs does not always occur synchronously and at once, that is, when data are entered into the system (e.g., at a particular modality). For example, before such data are passed on to some consuming systems (APPs), the examination results at that modality must first be reviewed. Only if the results are assessed as valid can they be transferred to other APPs. The knowledge what data are when passed on to what consuming APPs can best be drawn from so-called clinical processes.

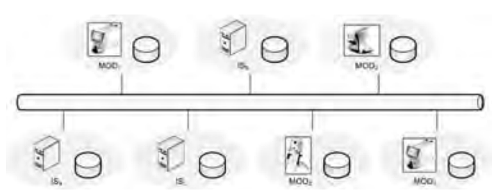
A very promising solution approach for the data integration issue in healthcare application is supported by process technology. Albeit, the usage of conventional process management technology is not practicable. For instance, conventional workflow management systems (as an example of conventional process technology) bear conceptual gaps in the field of adaptability and flexibility which makes their usage in this application scenario inadequate (Jablonski, Heinl, Horn, Neeb, Stein & Teschke, 1999). In order to prepare this discussion and to present our principle solution idea, some highly relevant constraints (Cx) from the healthcare application area have to be identified:

1. **(C1) Data transformation:** The data artifacts created by the different APPs have different formats, use different terminologies, and refer to different ontologies, but they must be synchronized with each other.
2. **(C2) Data identification:** Data produced at the different APPs use different identification schemes.
3. **(C3) Execution flexibility:** Data exchange between the APPs is very cumbersome. This is mostly due to the fact that the clinical personnel are very flexible in using the modalities during examinations. So it can hardly be predicted when an APP will be used; the physicians decide very spontaneously when taking the patients' actual condition into account. Consequently an IT system must not dictate the order of steps within an examination.
4. **(C4) Evidence based:** All examinations have to obey certain clinical and/or medical requirements. These are normally derived from clinical pathways and medical guidelines (Hellmann, 2003).
5. **(C5) Modeling flexibility:** Our application scenario is continuously changing since it is representing a research process. Due to newly discovered facets, the research process is constantly adjusted.
6. **(C6) Ease of use:** Data exchange between the APPs must occur transparently to the clinical personnel, as far as possible. As the personnel already have to handle multiple computer applications, it does not want to be burdened by just another IT system (Berg, 1999; Berg & Toussaint, 2003). Data management is regarded as an "auxiliary" task that should best be done by IT systems autonomously.

An analysis of the above requirements shows that they belong to three categories:

1. **Technical:** To cope with different formats, terminologies, ontologies (C1)), and identification schemes (C2) requires a technical solution. Among other things, mappings between data contribute to this matter.
2. **Application specific:** Clinical application scenarios tend to be spontaneous (C3), but restricted (C4) and dynamic (C5). This can be traced back to the enormous complexity of a patient treatment process.
3. **Usage:** Transparency of data management is a major need since it is regarded as an

Figure 2. Clinical (health care) application scenario



auxiliary, technical task, not as a medically motivated requirement (C6).

Our contribution is a process based data logistics mechanism that—transparently to the clinical users—controls the data exchange between the APPs. It is a process based concept since:

1. processes are regarded as being a suitable means to cope with complex application scenarios (C4),
2. processes can easily be modified and adjusted to changing requirements (C5), and
3. processes deliver a sophisticated structural framework to integrate diverse (technical) tasks of data exchange, that is, terminology, format, and ontology adjustment, and synchronization of identification schemes (C1, C2).

The above list reveals that still constraints C3 (flexible execution) and C6 (transparency of execution) must be dealt with. These constraints will not be met by processes directly, but by our data logistics concept.

In the next section, process management in healthcare application is discussed; especially the data logistics approach is presented. Our multi-staged approach to efficiently generate flexible process execution environments for healthcare processes—the process driven architecture (PDA)—is described in the following section. The implementation of our process management system (iPE) is detailed in the section following process driven architecture.

PROCESS MANAGEMENT IN HEALTHCARE APPLICATIONS

The Data Logistics Approach

Our data logistics method is based on an extended process management concept to cover all requirements identified in Section 1.3 (C1 to C6). The first step of our development method (Section 3) is to model clinical applications as processes (Beyer et al., 2004; Korycan, T. (1995). In order to sustain consistency and adequacy, these process definitions are based on medical guidelines and clinical pathways.

It is well recognized that process modeling is an excellent means to cope with complex and huge applications (Hammer & Champy, 1993; Ireson, 1997).

Especially the interactions between multiple participants and systems can be modeled efficiently such that constraint C4 is met perfectly. It is also well known from literature (Beyer et al., 2004) that process models are quickly and easily adoptable to changes. The graphical representation of processes facilitates change operations effectively and efficiently (constraint C5) even for domain experts such as the medical personnel.

Since we choose a perspective oriented process model, the technical requirements C1 and C2 can also be met without special efforts. In the perspective oriented process model, data constitute a separate perspective (Ireson, 1997).

Terminology, format, and ontology transformations, as well as synchronization of identifications are specific tasks associated with this data perspective.

Still, constraints C3 and C6 have to be addressed. They cannot be directly covered by normal process management. In order to cope with them, process management has to be extended by certain concepts, which we refer to as data logistics. Data logistics assumes the following:

- When the medical staff performs an examination, it is not guarded directly by the clinical processes defined in the first step of our method. Instead, it can use an application (APP) as spontaneous as if the data logistics have not been installed (C3), that is, the APPs could be used spontaneously. However, there must be some mechanism that controls the medically correct order of examinations. We will discuss this in Section 2.3.

- When an APP is used, data typically will be produced and stored in the local data stores ($DB\text{-}MOD_x$ and $DB\text{-}IS_y$, respectively). According to the data logistics specifications—which are derived from the defined clinical processes—these data are transformed and transported to specific sinks. This happens totally transparent to the users (C6).

The content of the second step of our data logistics method is to derive the data logistics processes from clinical processes. We call them data logistics workflows (DL workflows) in order to distinguish them from the "normal" clinical processes defined in the first step of our method. DL workflows run permanently and take care that data exchange between APPs is performed at the right time between the right APPs. One of the main features of DL workflows is their adjacency to clinical processes, that is, also to medical guidelines and clinical pathways. So, the very technical character of data logistics is directly connected to the clinical application field.

Having introduced the data logistics method, three main issues have to be investigated in the next sections:

- How to specify clinical processes?
- How to derive DL workflows from clinical processes?
- How to execute and implement DL workflows?

Processes

The following discussion about process management mostly concentrates on workflow management. Workflow management is regarded to aim at the execution of processes, which is our major interest. We will not provide a comprehensive introduction into workflow management in this section; instead we refer to Ireson, (1997) for a broad discussion of this topic.

According to the perspective oriented workflow model; a workflow consists of several perspectives which are weaved together. The functional perspective builds the skeleton of a workflow; it describes the tasks a workflow has to accomplish. The organizational perspective identifies the agents (e.g., physicians) responsible to perform a task and the operational aspect specifies applications supporting the execution of a workflow step. The control flow perspective defines the execution order between workflows. Finally the data aspect defines in/out parameters of workflows and how data flow between workflows.

In Figure 3 (left side) a very simple workflow model is depicted. Workflow WF1 is executed before workflow WF2 and WF3 respectively. The solid lines describe the data flow between workflows (WF). For instance, WF1 produces a data item OUT1 that becomes the input IN2 of WF2. In this figure we just use the parameters OUT1 and IN2 for readability; each input and output of a workflow can comprise a whole set of in- and out-parameters (e.g., documents). Additionally a control flow is defined between WF1 and WF2

Figure 3. A sample clinical workflow

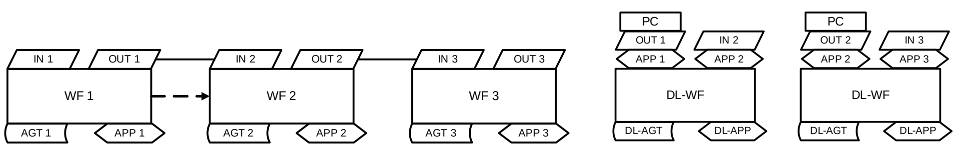
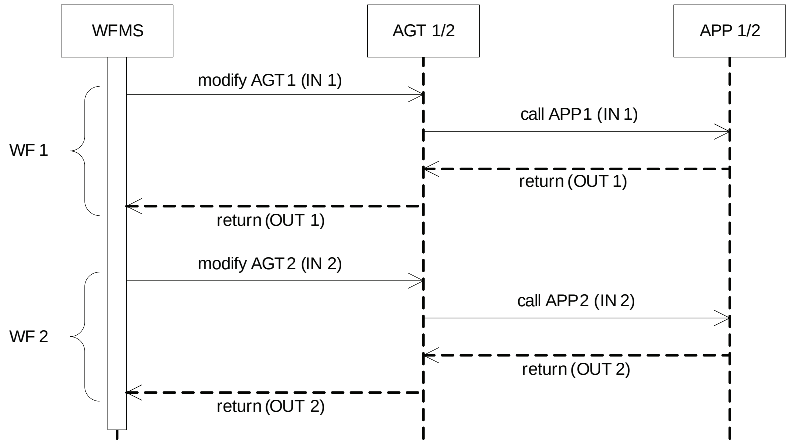


Figure 4. Processing of workflows in a conventional workflow management system (WfMS)



that, for instance, specifies that WF2 may not be executed before a certain time has passed. AGT1 denotes the agents that are responsible to perform WF1 whereas APP1 denotes the applications called by WF1 (e.g., an ophthalmic system). For simplicity we assume in the following that there is just one single agent and application specified for each workflow.

We will now provide arguments why the direct enactment of clinical processes by a workflow management system would not be acceptable for our clinical application scenario. Figure 4 shows as a simplified (unified modeling language [UML]) sequence diagram for workflow processing. The workflow management system (WfMS) selects a workflow that is ready for execution; in our case this is WF1 (see Figure 3). The WfMS notifies the declared agent (AGT) (here: AGT1) to perform the workflow step. The agent calls an application (here: APP1) that has to be executed. Required input data are always provided on the calls.

Figure 4 shows the substantiate duty of the WfMS which is to proactively trigger and initiate work to do. That means the following: If an agent has not been triggered by the WfMS it will not be able to perform its task. This behavior is contrary to what we have experienced in clinical applications and what we have postulated in Section 1.3 (constraint C3). There, we were assuming that the agents (e.g., physicians) are spontaneously selecting ophthalmic modalities for examinations. That means they cannot wait for a WfMS to offer such a modality for use, but they want freely—just piloted by medical requirements—to be able to select a modality. This is an absolute contradiction to the behavior a WfMS induces. As a marginal note, many trials to use WfMS technology in clinical environments were failing due to this discrepancy (Koppel, Metlay, Cohen, Abaluck, Localio, Kimmel et al., 2005).

In order to return to one of the primary goals we have to revisit our application scenario. Originally, data have to be exchanged between applications. This task is not directly fulfilled by the workflow depicted in Figure 3. Workflow management focuses the coordination of tasks, here WF1 and WF2. Indirectly, those tasks are also coordinated and connected through the association of these tasks with agents and application systems.

For the upcoming discussion of data logistics we need a semiformal model for workflow management. Therefore, we define the perspectives of a workflow in the following:

Control flow (CF) is a mapping CF⊆WF×WF which connects individual workflows of the set of all workflows (WF) with each other.

Data flow (DF) is a mapping DF⊆(WF, d_x) ×(WF, d_y). ((WF_x, d_x), (WF_y, d_y)) is element of DF if d_x is an output of WF_x and d_y is an input of WF_y (and both are type compatible, that is, data item d_x can be mapped onto data item d_y). Please note that there might be multiple DF or CF connections between two workflows.

A workflow WF_x∈WF is a structure ($DESC_x$, AGT_x, APP_x) where $DESC_x$ is its description, AGT_x is the agent responsible to execute the workflow, and APP_x is the application called within WF_x. Consequently, WF⊆DESC×AGT×APPP where DESC is the set of all activity descriptions, AGT is the set of all agents, and APPP is the set of all available applications. Without loss of generality, we assume that there are normally just one agent and one application associated with a workflow.

With this simple workflow model at hand we are now able to model clinical workflows. Especially as we use the customizable perspective oriented process (workflow) model of our process modeling tool iPM (ProDatO GmbH. (2006), we are able to model clinical processes appropriate. Meiler (2005) illustrates how the iPM workflow model can be effectively customized to describe clinical processes within a process driven application development methodology like the one we introduce in Section 3. Having accomplished this, the first task formulated at the end of Section 2.1 is fulfilled.

The simple workflow model also facilitates a formal derivation of DL workflows from clinical processes. This was the second open task formulated at the end of Section 2.1. The basic idea is that data exchange requirements between modalities (expressed in a DL workflow) can be derived from clinical processes as the example at the beginning of this section reveals: by combining the workflows WF1 and WF2 as shown in Figure 3 (left side) a data flow between APP1 and APP2 is specified.

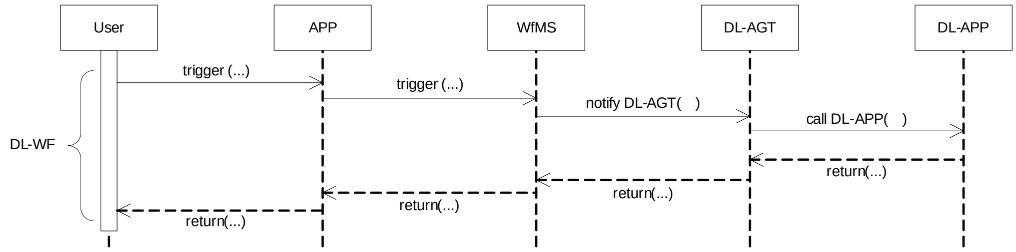
Data Logistics

In this section we present the data logistics approach. We start with an example that illustrates the basic idea (Section 2.3.1). Section 2.3.2 demonstrates the formal transformation of clinical processes into DL workflows.

An Illustrating Example

Data logistics aims at the coordination of data exchange between APPs. What does that mean with respect to the examples of both Figure 3 and Figure 4? When APP1 is executed (under the supervision of WF1) then the output data OUT1 has to flow from APP1 to APP2 (here it becomes the input data IN2). This is equal to the fact that after the use of application APP1, created data are

Figure 5. Processing of data logistics



distributed to APP2. Without loss of generality, our example data flow is just between two applications. It is no conceptual problem to extend this to multiple APPs.

Data logistics focuses on data exchange between APPs. Thus, we now change the perspective. Instead of observing the clinical processes (workflows, respectively) (Figure 3 and Figure 4), we look onto the DL workflows derived from them. A DL workflow needs a couple of special components:

- **DL-APP:** This is a special purpose application that is capable of shipping data from data sources to sinks, that is, to perform a data transport between APPs. In order to function it needs the following parameters: the participating APPs (data source and data sink), the corresponding output and input data, and the process context.
- **DL-AGT:** This is a special purpose agent who is associated with an APP, that is, each APP maintains a specific DL-AGT. The DL-AGT is triggered when the corresponding APP is used. The DL-AGT then initializes the local DL-APP, that is, it initializes data logistics and thus performs the data transport.
- **DL-WF:** This is a DL workflow that is always associated with a DL-APP and a DL-AGT.

The above components are sufficient to describe DL workflows. Figure 3 (right side) exposes the two DL workflows which are derived from the clinical process of Figure 3 (left side). The general rule is to create one DL workflow for each data connection of a clinical process (here: data flow between APP1 and APP2, and between APP2 and APP3). Section 2.3.2 will explain this issue in detail.

In Figure 5 the execution of a DL workflow is illustrated. The following sequence of operations takes place:

- A user is calling an APP, that is, the corresponding wrapper.
- The wrapper of this APP recognizes this and informs the WfMS about the usage.
- The WfMS notifies the DL-AGT of this APP to be prepared for initializing data exchange; this normally happens after the user ceases to use the APP.
- The DL-AGT calls the DL-APP with the corresponding parameters (see above) to initiate data exchange.

There are a couple of observations that must be discussed here in detail:

- There is only one DL workflow to be performed, namely DL-WF. Depending on the provided parameters, it is controlling data exchange between corresponding APPs. Similar, there is only one specific agent (DL-AGT) and one specific application (DL-APP) required. We will later go into details of this discussion.
- Since there is just a “normal” workflow, a “normal” agent, and a “normal” application involved, a conventional WfMS is suitable as execution platform.
- Although Figure 5 is simplifying the processing it nicely demonstrates that now the initializing actor is the user of an APP. This well corresponds to requirement C3. Besides, the data exchange takes place transparently such that constraint C6 is also fulfilled.

Aside from implementation details which are explained later, the demonstrated DL workflow is fulfilling all the requirements identified in Section 1.3. It is derived from clinical processes (C4, C5), it allows spontaneous user interaction (C3), and it organizes data exchange transparently (C6). The remaining two technical requirements C1 and C2 will be explained in Section 2.3.2.

Formal Derivation of DL Workflows

One of the most important advantages of our data logistics approach is that it has both a robust content related origin (the clinical processes) and a unique formal foundation.

The transformation from clinical processes to DL workflows focuses on data flows between process steps of a clinical process. Thus, for any two workflows WF_i and WF_j connected by a data flow (from WF_i to WF_j) the following DL workflow DL-WF can be derived: DL-WF = ($DESC_{DL-WF}$, DL-AGT, DL-APP)

$DESC_{DL-WF}$ is the description of the data logistic workflow, DL-AGT is the general data logistics agent, and DL-APP is the general data logistics application. These three components of the DL workflow will be the same for all clinical processes. Nevertheless the input and output parameters of a DL-WF are more interesting than the above aspects. For WF_i and WF_j with input data IN_i, IN_j, output data OUT_i, OUT_j, and applications APP_i, APP_j, the input data for the DL-WF are `PC`, OUT_i, APP_i. The output data for the DL-WF are `PC`, IN_j, APP_j.

The parameter PC indicates the process context (PC) in which a data logistics workflow is performed. Through this context information the two workflows WF_i and WF_j are determined. This context information is necessary, since the APPs can be reused in various clinical processes for different purposes. So the data logistics system must recognize not only the usage of a certain APP, but also the context (i.e., clinical process) this APP is used in. This determines where to transmit the data created at the APP.

Executing DL Workflows

Unfortunately, simply transporting data from APP1 to APP2 is by far not enough in a real world application scenario like we encountered. In fact many different data structures are used by the participating applications, modalities, and information systems. A closer investigation of the relevant data reveals that the data in general differ in three respects: format, (applied) ontology, and terminology. As a consequence the DL workflows not only have to consider data transport but also data transformation; the latter deals with the differences between formats, ontologies, and terminologies of different APPs.

There are two principal approaches to cope with the challenge of a set of heterogeneous data structures. The simple, but laborious strategy would be to implement a specialized transformer for every pair of applications that exchange data. This approach is in principle feasible, as the use of the process context (PC [Section 2.3.2]) allows us to mostly automatically select what transformer to use during the execution of the DL workflow. Yet, in an application scenario that involves more than a hand full of data flows, the number of required transformers rises according to the following formula, where *n* is the number of participating data structures:

$$\frac{n*(n-1)}{2}$$

It is obvious, that this approach is cumbersome and error prone in a real world setting. This problem is well known in the medical domain and several standards like health level 7 (HL7) (Health Level Seven Inc., 2006) and the clinical document architecture (CDA) (Dolin, Alschuler, Beebe, Biron, Boyer, Essin et al., 2001) try to solve it. Thus we decided to use the same strategy and offered a generic data structure that serves as application independent data container. This significantly reduces the number of needed transformers, as every application needs only a single transformer to convert its data structure to the common data structure and vice-versa. In our projects we mainly use the Erlanger Glaukomregister (Figure 1) and its generic relational database schema as common data structure. In Section 4.3 we outline an example where we further standardized our glaucoma research data structure by using CDA to be able to link our research database to the HIS "Soarian" and to an electronic patient record.

After this comprehensive discussion of the basic concepts of our data logistics approach, the next section gives a brief introduction to our holistic development approach for process oriented applications.

PROCESS DRIVEN ARCHITECTURE (PDA)

We developed a comprehensive methodology to efficiently support the semiautomatic generation of process oriented applications, called process driven architecture (PDA). It is used to provide a modeling and execution environment for DL workflows.

Foundations of the Process Driven Architecture

Our approach is inspired by several concepts derived from the area of model driven software development and especially the Object Management Group's (OMG) model driven architecture (MDA) (Kleppe, Warmer, & Bast, 2003). The MDA is a highly structured general purpose approach to facilitate the semiautomatic generation of software, based on the OMG's unified modeling language (UML) (2007). It allows for the specification of software based on customized UML models. The models are built upon the terminology of the application domain, generally referred to as the domain specific language (DSL). The MDA promotes an iterative development cycle based on three, increasingly system specific model levels: the computation independent model (CIM), the platform independent model (PIM), and the platform specific model (PSM).

The most abstract, but at the same time most domain specific model level is the CIM, which may be modeled using any kind of domain specific description (e.g., narrative use cases). This fuzzy method to describe a software system on a very high semantic abstraction level might be regarded as advantage, as it offers much flexibility, otherwise it might be regarded as disadvantage, as it avoids efficient tool support to specify and transform the CIM to the next model level. In fact, using the MDA often results in a semantic gap between the CIM and the PIM, as it in general involves multiple human interpretations of the imprecisely described CIM.

The PIM is specified using an individually customized version of UML, mostly based on UML stereotypes and UML profiles. Consequently, it is the first machine interpretable description of the software system and thus it can be automatically transformed to the next model level. Nevertheless, it still is not connected with a certain system platform (e.g., the .NET platform). Finally the PSM is generated directly by transforming and enriching the PIM by platform specific details, that is, an UML model of the software system is produced.

As mentioned above, the MDA inspired the development of the PDA. Unfortunately, it turned out during our two projects (Section 1) that a simple domain specific adoption of the MDA and UML would not fulfill the characteristic requirements of process oriented software development. The goals of the MDA are very generically to support the generation of any kind of software using the UML. However the UML is not perfectly suitable to describe complex and highly flexible medical processes, like the ones we encountered at the clinical application domain (Section 1.2). Several facts account for that. First, the UML currently does not directly support the design of an arbitrary graphical modeling notation. Second, a comparative analysis of several process modeling methodologies and languages revealed that the UML offers no concepts that are expressive enough to model processes precisely on a high abstraction level. Especially the UML 2.0 activity diagram currently does not provide satisfactory modeling constructs for our modeling needs. A project specific comparison of several process modeling methodologies and languages can be found in Müller, S. (2007) work. Third, the MDA's CIM is too imprecise to allow for a semiautomatic and thus effective and efficient transformation into the PIM.

As a result, we decided to develop a generic development method for process oriented applications based on the perspective oriented process modeling paradigm used for our PDL approach (Section 2). This allows us to build a perfectly integrated approach to efficiently derive an executable process oriented software system from a domain specific process description understandable for domain experts, such as the medical personnel.

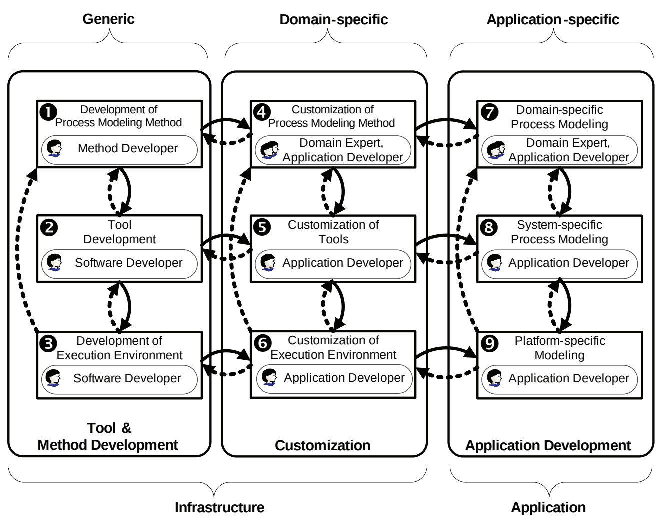
The PDA Approach

The PDA promotes a comprehensive approach for the development of process oriented software systems, which covers not only the application development, but also the prior development of adequate tools, methodologies, and their necessary customizations (Müller, 2007). The overall PDA development process is divided into three stages (Figure 6): tool and method development, customizing, and application development.

During the first stage the basic process modeling methodology, for example, the perspective oriented process modeling method, and associated tools are developed. This is done in three steps. First, a method developer, typically embodied by an industrial consortium (e.g., the OMG) or a researcher, designs a process modeling method (Figure 6, ❶). Second, software developers create specialized tools to be able to efficiently perform modeling tasks using the new process modeling method (Figure 6, ❷). Finally, an execution environment is developed that can implement processes modeled using the newly designed method (Figure 6, ❸). The tool and method development stage is generic by nature, as it is independent from any application domain and the results, that is, the methodology and the tools, can be used very widely. At the first sight, it seems to be plausible that the tool and method development stage is executed sequentially.

The second stage of the PDA approach adopts the results of the first stage to a certain application domain, for example, the clinical field of application. Yet, it is still independent of a con-

Figure 6. The multi layered PDA approach



crete application process and thus belongs to the infrastructure development. The customization stage is split into three steps as well. In the first step, the process modeling method is customized to the application domain (Figure 6, ❹). Typically this involves the cooperative design of modeling constructs that adhere to the terminology of the application domain. For example, the application independent modeling construct "organizational unit" may be specialized as physician, assistant medical technician, patient, and so forth. Actually, a new idiom of the modeling language, called the domain specific language (DSL) (van Deursen, Klingt, & Visser, 2000), is created that serves as communication basis between the domain experts and the application developers (Sections 1.1 and 2.1). This step is critical to many model driven software development approaches (Stahl & Völter, 2005) and thus has to be performed in close cooperation between experienced domain experts and application developers. The newly introduced modeling constructs and all other modifications of the modeling methodology subsequently have to be reflected in the modeling tools, which is done in the next step (Figure 6, ❺). The application developer is responsible to configure the used tools accordingly, that is, the tools have to be highly adaptable. The process execution environment has to be customized to fit into the application domain as well (Figure 6, ❻). The necessary actions at this step range from the installation of basic infrastructure, such as a Web server, to the provision of network access to required information systems, such as a HIS. Again, the customization steps are in practice interdependent and executed iteratively, if necessary. At the end of the customization stage, all tools and the modeling methodology are fitted to the intended application domain, that is, the development infrastructure is ready to be used for the "production" of process oriented applications.

For every process that is to be supported by our PDA approach, only the application development stage has to be passed. Three steps constitute the PDA application development. At first, a domain specific model ("clinical pathway") of the process to support is drafted cooperatively by the domain experts, for example, the medical personnel, and the application developers (Figure 6, ❼). This step is of utmost importance for the quality of the resulting process support application and has to be done very carefully. This ability to generate an application that is ready to run directly results from the comprehensive efforts during the customization stage. Directly after the domain specific process modeling is finished, the application developer takes care of system specific adoptions and enrichments of the process model by simply wiring the previously configured and prepared tools with the elements in the process model (Figure 6, ❽). All actions that have to be taken after this step may be automated and do not need any further human interaction if prepared well during the customization stage. The last application development step comprises the generation, compilation, and deployment of the application (Figure 6, ❾). What steps are exactly necessary depends on what kind of process execution environment was selected, that is, the generation of software is treated differently than the deployment of a workflow in a WfMS. Respectively, the degree of manual intervention strongly depends on the specific situation, especially if a well organized test and release cycle follows. The application development stage results in an executable process oriented application.

The next section gives a brief insight on how we applied our PDA approach within our two projects.

The Adaptation of the PDA

The PDA provides the structural framework for the development of process oriented application systems. In the following, we illustrate two of our major design decisions and briefly introduce the tools we used. As mentioned in Section 3.2, our first important decision regarding the tool and method development was to use the perspective oriented process modeling methodology for our projects. As a consequence, we chose the sole existing process modeling tool for perspective oriented process modeling, the integrated process modeler (i>PM) provided by ProDatO GmbH (ProDatO GmbH. (2006). This tool offers a convenient graphical editor and is easily adoptable to a new application domain mostly by configuration. When we started our project, i>PM offered the default graphical notation for perspective oriented modeling and several specialized notations (i.e., DSLs) for the automotive segment of industry. After the domain specific customization of the process modeling method, we added our DSL for the medical domain, called i>PM4Med, to i>PM and iteratively extended it. Thus we are able to perfectly support the system and domain specific modeling steps during the application development phase.

The second important strategic decision was to develop our own process execution environment. During the last years we pursued and compared two completely different approaches to support process execution in the sense of data logistics (Section 2.1). The first approach was a monolithic execution environment that is capable of interpreting XML-based process type definitions from i>PM4Med directly (Jablonski, Lay, Müller, Meiler, Faerber, Derhartunian et al., 2005). For this kind of application it is characteristic that new features have to be added to the central process interpretation and execution engine. Obviously this central process interpretation and execution engine is complex and it can easily become the bottleneck during process execution. Although we never experienced any serious problems with this kind of execution environment, we decided to try to develop a generative approach as alternative implementation. Thanks to the structured PDA development approach, we were able to introduce the new implementation step by step using clearly defined interfaces between the PDA stages and steps. Especially the users and the domain experts were nearly not affected by this transition.

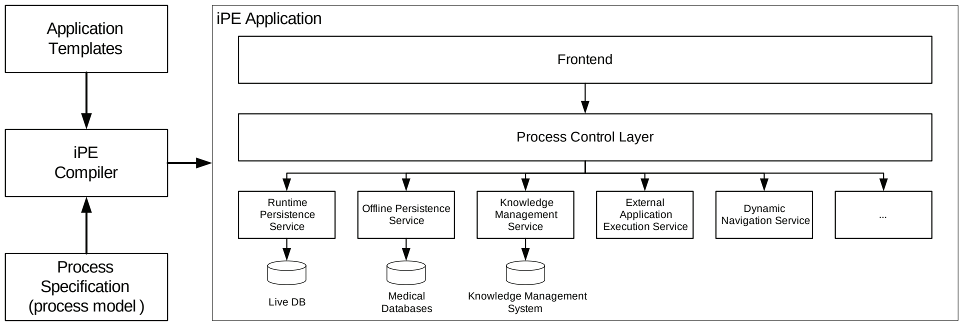
IPE: ARCHITECTURE AND IMPLEMENTATION

Modeling processes using a DSL helps to understand the procedures occurring within a medical institution. Beside this, processes can also practically support people; this is the application field of iPE, our solution for the execution of process models. The core component of iPE is the iPE compiler which we have developed for transforming process models into applications which can be interactively executed by a computer.

Function and Architecture of the iPE Compiler

The iPE compiler has been implemented with the concepts of the perspective oriented process modeling in mind. This means that the compiler can deal with arbitrary perspectives in such a way that their translation into executable code can be defined from the outside. The compiler itself reads a given process model and transforms it into a representation that is independent of the modeling language used for describing one process. After this conversion the model is passed on to single subprograms—called generators—which are responsible for producing the single components of the final application. It is the responsibility of each generator to decide which perspectives are interpreted and which are left out. Furthermore, the code the generator is producing out of the process model only depends on the generator and not on the compiler. This allows us to develop sets of generators which result – when used together – in one type of application (e.g., a Web application). With this in mind we also refer to the generators as templates for the mapping of a process to an application. Looking at the result

Figure 7. The iPE compiler generates the iPE application



of the compilation, an application has been built which is specific to the application domain and the process.

Figure 7 shows the use of the iPE compiler. The iPE compiler uses templates to transform the process model first into a modeling language neutral representation and then into an executable application. The architecture of the generated applications only depends on the templates and not on the compiler. Adding a new perspective to a process model means to provide in iPE a template for the compiler for code generation and changing the existing template for the language mapping into the neutral representation.

Architecture of an iPE Application

We have written a set of templates which generates a Web application for a given process. The architecture of these applications follows the classical model-view-controller (MVC) (Gamma, Helm, Johnson, & Vlissides, 1995) design pattern. The application has been split into three layers. The first layer is the front-end which is only used for the interaction of a process with its users if no other application for the execution of a process step is given. The second layer is the process control layer (PCL) which manages the interactions of the front-end and the third layer, the data layer. The front-end corresponds with the *view*, the PCL with the *controller*, and the data layer with the *model* in the MVC pattern. The data layer is again divided into separate services which add functionality to the iPE application being generated. In Figure 7 the following components are shown:

- **Runtime persistence service (RPS):** The RPS stores all the data gathered during the execution of a process.
- **Offline persistence service (OPS):** The OPS only considers data which is of special interest (e.g., patient records). At specific points during the execution, the OPS writes or reads data to/from medical databases.
- **Knowledge management service (KMS):** The KMS supports the user of a process by displaying links to additional background information, for example, about a special treatment or disease. Also during the execution new knowledge is created which can be written to a knowledge database through this service.
- **External application execution service (EAES):** This service is responsible for the interaction with external programs and/or hardware (e.g., a word processor for filling out special forms).

- **Dynamic navigation service (DNS):** The DNS decides for the next process in the line of execution depending on the users inputs.

These services can then be mapped onto the perspectives of the perspective oriented process modeling approach. The RPS and OPS together make up the data and data flow perspective, the KMS corresponds to a knowledge perspective (if introduced at all); the EAES is the direct counterpart of the operational perspective. The functional perspective is not directly visible. In an iPE application every process which did not specify the operational perspective is mapped onto a "default application" which is in our case simply a Web page displaying the incoming and outgoing data providing some functionality to manipulate it. The DNS implements more than one perspective. It considers causal dependencies in between processes (e.g., given through the functional perspective) as well as dependencies given through the data and data flow perspective respectively (a process can only be executed in case the data needed as input is available). Last but not least the organizational perspective is not shown; it is also implemented as a service attached to the PCL.

The PCL is the link in between the front-end and the data layer. It creates commands out of the inputs which manipulate data (RPS and OPS), determine the new process to be executed (DNS), and start the actual execution (EAES). Feedback is provided for the user through the front-end.

Integrating Data Logistics into a Clinical Application

An important issue with all clinical support systems is how they store data persistently. As described in the architecture section iPE implements a two staged persistence layer consisting of the runtime persistence layer (RPS) and the off-line persistence layer (OPS). The RPS has to store all data that are generated during a session. However the RPS only has to ensure that data are kept during process execution. After completing the process all medical data are committed to the OPS and removed from the RPS system.

After the execution of a path, the OPS has to make all medical data of the corresponding instance persistent. Two persistent storage systems are used: The HIS "Soarian" stores data used in the clinical routine, whereas the Erlanger Glaukomregister serves as central research database (Figure 1) for the SFB 539 (Section 1). Data are submitted to Soarian using HL7 messages containing glaucoma related data encoded in a newly developed CDA/SCIPHOX (Heitmann, Schweiger, & Dudeck, 2002) compliant document structure (Gerdsen, Müller, Jablonski, & Prokosch, 2006).

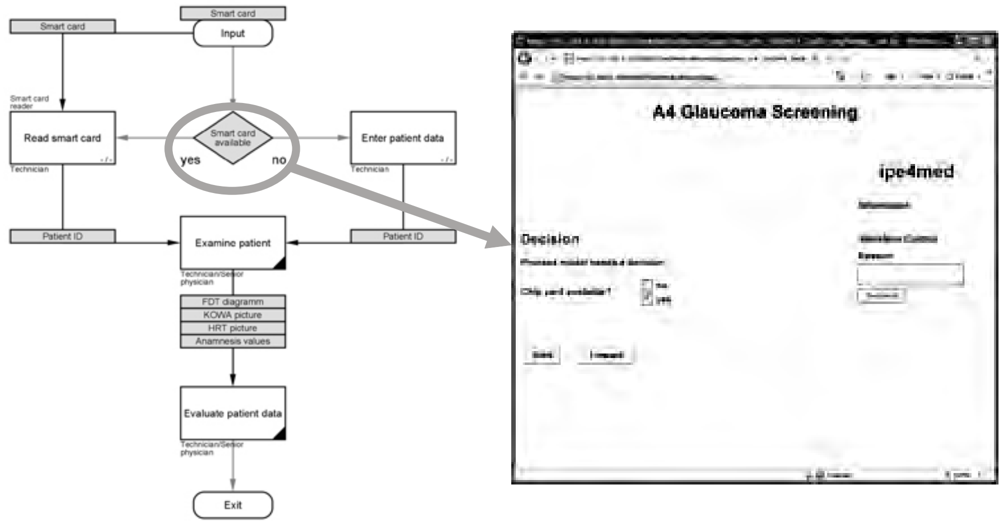
Execution Semantics and Application Example

The iPE engine is the central component as it generates the Web application which is our interface to the users. When processes are executed, three major concerns have to be reflected by the execution environment:

1. Be as restrictive as necessary so that no important steps in the process are left out.
2. Be as flexible as possible so that physicians and medical assistants have a high degree of freedom.
3. Provide all information that helps to fulfill the given task.

In the following section we explain how the execution of DL workflows is going to take place and how the three identified concerns are implemented. As example, we use a glaucoma screening process that is used at the Ophthalmic Department of the University of Erlangen for the early detection of the glaucoma disease. The process shows the following global structure.

Figure 8. Decision support during process execution



At first a patient ID has to be selected and the patient's master record has to be created. Then the patient is examined and the medical findings are transferred into the database. Finally a physician has to evaluate the patient data and a medical findings letter is printed.

It is quite common during the execution of a process that the user comes to some decisions. Figure 8 displays a situation at the beginning of the process where the patient's master record has to be entered into the system. In the German health system this record is stored on the smart card provided by the health insurance and can usually be read electronically by a smart card reader. However, this process is also used in a research environment where users do not have to provide insurance data obligatorily. Since the system cannot determine whether a smart card is available, iPE creates a Web page where a medical technician can provide information about the availability of the smart card (Figure 8).

In Figure 9 the execution of the process step "Review Patient Data" is depicted. First of all, an HRT2 picture of the right eye is presented to the physician; this data and further data to assess the state of the patient's eye are shown in the GUI (not all of them are visible in Figure 9). Below each section of the GUI (i.e., HRT picture or the patient's anamnesis) a physician can enter medical findings. As these findings are standardized and coded by numbers, a physician can select an appropriate entry out of a list; this feature increases the consistency of data entry, since only eligible data are offered. Later in the process these number codes are automatically transformed into readable text and printed in a findings letter.

In the information section of the GUI (right side) additional background information is provided that supports the user during his daily work. Here, links to Web pages in the clinic's intranet, operating instructions, or best practices, as well as hints can be given. Also the integration with knowledge management systems is possible and a first prototype exists that demonstrates the integration.

After each step of the process the DNS decides for the next process in the line of execution. As

Figure 9. Reviewing the patient’s data

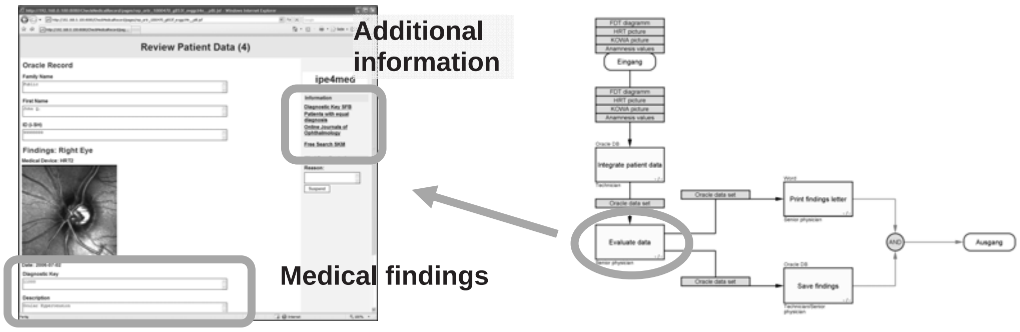


Figure 10. Dynamic navigation during process execution

described in Section 4.1, causal, organizational, and data flow dependencies are analyzed as well as the steps that already have been completed to determine which steps can be executed. The list of executable process steps is then displayed in the “Dynamic Navigation” section in the Web page (see Figure 10). Before a user can complete a certain process step, the user has to select which step has to be executed next. This way the control flow of the application can be determined by the user, whereby the boundaries of the process definition have to be regarded (control and data flow). The DNS is the central component in implementing our requirement for medical process execution engines to be as restrictive as necessary, but also to be as flexible as possible.

Even though this simple scenario only gives a glimpse on the functionality iPE offers, it demonstrates how flexible processes can be executed by iPE. Through the template based architecture of iPE, the generated applications can be tailored to the specific requirements of an application domain.

CONCLUSION

This contribution presents an integrated approach for process and data integration in healthcare applications. We provide a process based solution for data and process integration of healthcare applications. Long term studies in real clinical applications—some parts of them are discussed in this article—prove the feasibility of the approach.

One of the major contributions of our approach is the integration of multiple concepts and techniques from IT into a powerful process and data integration mechanism. Starting with process management, incorporating data management concepts, and completing the approach with the PDA, a powerful and comprehensive modeling, development, and execution environment for healthcare applications is constituted.

Currently we are working at the incorporation of a knowledge management component during the process execution. It should be used to capture knowledge about DL workflow execution to be later offered to new clinical personnel in order to better and faster made familiar with the clinical processes. We expect that this will further leverage the supportive effects of our data logistics approach for the medical personnel.

ACKNOWLEDGMENT

This work was mainly supported by funds provided by the German Research Foundation ("Deutsche Forschungsgemeinschaft [DFG]")

REFERENCES

Anderson, J. G. (1997). Clearing the way for physicians' use of clinical information systems. *Communications of the ACM, 40*(8), 83-90.

Berg, M. (1999). Patient care information systems and healthcare work: a sociotechnical approach. *Int J Med Inf, 55*, 87-101.

Berg, M., & Toussaint, P. (2003). The mantra of modeling and the forgotten powers of paper: A sociotechnical view on the development of process-oriented ICT in healthcare. *Int J Med Inf, 69*.

Beyer, M., Kuhn, K. A., Meiler, C., Jablonski, S., & Lenz, R. (2004). Towards a flexible, process oriented IT architecture for an integrated healthcare network. Special track on computer applications in healthcare (COMPAHEC 2004). In *Proceedings of the Annual ACM Symposium on Applied Computing (SAC'04).*

Dadam, P., Reichert, M., & Kuhn, K. (2000). Clinical workflows: The killer application for process-oriented information systems? In W. Abramowicz & M.E. Orlowska (Eds.), *BIS 2000: Proceedings of the 4th International Conference on Business Information Systems* (pp. 36-59). Poznan, Poland: Springer.

Dolin R. H., Alschuler, L., Beebe, C., Biron, P. V., Boyer, S. L., Essin, D., et al. (2001, November-December). The HL7 clinical document architecture. *J Am Med Inf Assoc., 8*(6), 552-569.

Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1995). *Design patterns.* Addison-Wesley Professional.

Gerdsen, F., Müller, S., Jablonski, S., & Prokosch, H.-U. (2006). Standardized exchange of clinical documents: Towards a shared care paradigm in glaucoma treatment. Special Issue of Meth Inf Med of the 50th ann. Conf. of the German Society of Medical Informatics, Biometry and Epidemiology (GMDS) combined with the 12th Conf. of the German Working Group of Epidemiology (DAE), Spec. Iss. GMDS/DAE 2005.

Hammer, M., & Champy, J. (1993). *Reengineering the corporation: A manifesto for business revolution*. New York: Harper Collins Publishers, Inc.

Handler, J. A., Feied, C. F., Coonan, K., Vozenilek, J., Gillam, M., Peacock, P. R., et al. (2004). Computerized physician order entry and online decision support. *J Acad Emerg Med, 11*(11), 1135-1141.

Haux, R., Seggewies, C., Baldauf-Sobez, W., Kullmann, P., Reichert, H., Luedecke, L., et al.

(2003), Soarian: Workflow management applied for healthcare. *Meth Inf Med, 42,* 25-36.

Health Level Seven Inc. (2006). Health level 7. Retrieved May 27, 2007, from http://www.hl7.org/

Heitmann K. U., Schweiger, R., & Dudeck, J. (2002). Discharge and referral data exchange using global standards - the SCIPHOX project in Germany. *Stud Health Technol Inform., 90,* 679-84.

Hellmann, W. (2003), Praxis Klinischer Pfade, Ecomed Verlagsgesellschaft, Landsberg/Lech.

Ireson, C. L. (1997). Critical pathways: Effectiveness in achieving patient outcomes. *J Nursing Adm, 27*(6).

Jablonski, S., & Bußler, C. (1996). *Workflow management: Modeling concepts, architecture and implementation.* London: International Thomson Computer Press.

Jablonski, S., Heinl, P., Horn, S., Neeb, J., Stein, K., & Teschke, M. (1999, February). A comprehensive approach to flexibility in workflow management systems. In *Proceedings of the International Joint Conference on Work Activities Coorindation and Collaboration (WACC'99),* San Francisco.

Jablonski, S., Lay, R., Müller, S., Meiler, C., Faerber, M., Derhartunian, V., et al. (2005). Building a generic platform for medical screening applications based on domain specific modeling and process orientation. In B. Ludäscher & L. Raschid (Eds.), *Second international workshop on data integration in the life sciences (DILS 2005)* (pp. 257-265). San Diego.

Kleppe, A., Warmer, J., & Bast, W. (2003). MDA explained. The model driven architecture: Practice and promise. *Object technology series.* Addison-Wesley.

Koppel, R., Metlay, J.P., Cohen, A., Abaluck, B., Localio, A.R., Kimmel, S.E., et al. (2005, March 9). Role of computerized physician order entry systems in facilitating medication errors. *JAMA, 293*(10), 1197-203.

Korycan, T. (1995). Achieving cost and quality benefits in critical pathway implementation. *J Qual Mgm in Healthcare, 3*(2).

Kuhn, K.A., Giuse, D.A., Bakker, A.R., Ball, M.J., & Gell, G. (2001). Challenges in deploying health information systems. In V.L. Patel, R. Rogers, & R. Haux (Eds.), *Medinfo 2001, Proceedings of the 10th World Con Med Inf* (p. 1551).

Lenz, R., Buessecker, F., Herlofsen, H., Hinrichs, F., Zeiler, T., & Kuhn, K. A. (2005). Demand-driven evolution of IT-systems in healthcare: A case study for improving interdisciplinary processes. *Meth Inf Med,* 44, 4-10.

Lenz, R., Elstner, T., Siegele, H., & Kuhn, K. A. (2002). A practical approach to process support in health information systems. *J Am Med Inform Assoc., 9*(6), 571-585.

Lenz, R., & Kuhn, K. A. (2004). Towards a continuous evolution and adaption of information systems in healthcare. *I J Med Inf, 73*, 75-89.

Meiler, C. (2005). *Modeling, planning, and execution of clinical paths.* Unpublished doctoral thesis, University of Erlangen-Nuremberg.

Müller, S. (2007). *Model based IT support of knowledge-intensive processes.* Unpublished doctoral thesis, University of Erlangen-Nuremberg.

ProDatO GmbH. (2006). *ProcessManager.* Retrieved September 23, 2006, from http://www.prodato.de

SFB 539. (2007). Sonderforschungsbereich 539: Glaukome einschließlich pseudoexfoliations-syndrom (PEX). Retrieved February 4, 2007, from http://www.sfb539.forschung.uni-erlangen

Stahl, T., & Völter, M. (2005). Modellgetriebene Softwareentwicklung. Techniken, Engineering, Management. 1. Auflage 2005, dpunkt.verlag, Heidelberg.

UML. (2007). Object management group (OMG): Unified modelling language (UML). Retrieved February 4, 2007, from http://www.uml.org

van Deursen, A., Klingt, P., & Visser, J. (2000, June). Domain-specific languages: An annotated bibliography. *ACM SIGPLAN Notices, 35*(6), 26-36.

This work was previously published in International Journal of Healthcare Information Systems and Informatics, Vol. 2, Issue 4, edited by J. Tan, pp. 1-21, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.25
Approximate Processing for Medical Record Linking and Multidatabase Analysis

Qing Zhang
CSIRO ICT Centre, Australia

David Hansen
CSIRO ICT Centre, Australia

ABSTRACT

In this article we investigate how approximate query processing (AQP) can be used in medical multidatabase systems. We identify two areas where this estimation technique will be of use. First, approximate query processing can be used to preprocess medical record linking in the multidatabase. Second, approximate answers can be given for aggregate queries. In the case of multidatabase systems used to link health and health related data sources, preprocessing can be used to find records related to the same patient. This may be the first step in the linking strategy. If the aim is to gather aggregate statistics, then the approximate answers may be enough to provide the required answers. At least they may provide initial answers to encourage further investigation. This estimation may also be used for general query planning and optimization, important in multidatabase systems. In this article we propose two techniques for the estimation. These techniques enable synopses of component local databases to be precalculated and then used for obtaining approximate results for linking records and for aggregate queries. The synopses are constructed with restrictions on the storage space. We report on experiments which show that good approximate

results can be obtained in a much shorter time than performing the exact query.

INTRODUCTION

There is an increasing amount of data being captured in health systems. The data are generally spread among many different data systems. Primarily the data are used for patient treatment; however secondary use of this data is important for improving the health service and medical treatments. These data might be used by the health service to ensure that there is adequate service provision. The medical community may use these data for gathering important information on diagnosis, treatment, and outcome for evidence based practice and future research. With the data spread between so many data repositories, multidatabase systems (Stead, Miller et al., 2000) provide one way to access the data. However, performing queries across large distributed data sets can be expensive, both in time and computing resources.

A typical multidatabase query is to validate patient records across separate local databases which refer to the same person. This can be difficult when, as is often the case, there is no global identifier to provide a single key. For example, suppose one database records the diagnosis information of patients who have been diagnosed with a particular cancer and another database records the information of patients who have received chemotherapy. If no unique patient identifier exists for these two databases, identifying information must be compared from the two databases to identify records which refer to the same patient. Based on the linking results, other analytic queries can then be issued to analyze diagnosis and treatment details.

The linking result can be represented as a *link table* that lists all matched records along with some measure of accuracy of the match. The number of records in the link table is known as the *selectivity* of the link table. The global query optimizer of a multidatabase system uses this selectivity to decide proper strategies on dissecting global query, transferring data among local databases, and so forth. However, a naïve selectivity estimation requires linking two local databases first, which may be very time consuming and even unaffordable due to the large amount of medical data or network failure possibilities. This provides a first motivation for developing an effective selectivity estimation method in a health multidatabase.

Our second motivation comes from the exploratory nature of some queries. In cases, such as decision support systems, users may not be particularly interested in the exact result (Chakrabarti, Garofalakis et al., 2000). An example of this situation may be the case of investigating the relationships between two diseases. Users may first wish to know that the number of patients linked in the two disease databases is big enough to do any further analysis. A large amount of time can be saved if the system can provide an approximate result early. Users then decide if a fuller investigation is warranted.

Motivated by these, we have developed a method for link table selectivity estimation based on constructing database synopses of large and complicate health databases. Since the size of the synopses is strictly limited to the storage requirements of the system, the main challenge in our method is to construct a "good" synopsis which can provide a close estimation to the exact result. We test our approximation methods on various synthetic databases. Experiments show that our techniques can provide close approximation in a much quicker response time than getting the exact result through the naïve query-and-count way.

The rest of the article is organized as follows. The following section presents the background of the selectivity estimation and the structure of our selectivity estimation module. The third section presents our estimation method based on wavelet transformation. The next section presents the estimation method based on histogram synopses. The following section evaluates our selectivity estimation techniques with various experiments. The sixth section demonstrates how to use our method for more complex multidatabase aggregate queries. The next to last section lists related works and the last section remarks our work and states possible future research issues.

BACKGROUND

In this section, we briefly review two traditional selectivity estimation techniques used in relational databases, the histogram and wavelet. Selectivity estimation is the use of precomputed synopses of the original data to estimate the number of records satisfying the selection criteria. Since such synopses usually occupy a small space compared with the original data, the quick response of the approximate answer would be naturally expected. However, constructing an efficient synopsis is a trade-off between space and accuracy. Moreover, in an environment where original data keep changing dynamically, updating the constructed synopses accordingly is an important issue. In this article, we assume that the original data are relatively static and hence the synopses can be updated by a periodically reconstructing procedure.

Formally, given a relation **R** and an attribute **X** of **R**, the domain **D** of **X** is the set of all possible values of **X**, and a finite set **V** ($\subseteq$ **D**) denotes the distinct values of **X** in an instance *r* of **R**. Let **V** be ordered increasingly. An instance *r* of **R** can be denoted by **T,** known as r's *data distribution:* **T** = $\{(v_1, f_1), ..., (v_n, f_n)\}$. In **T**, every (v_i, f_i) is called the value-frequency pair. Each v_i is unique and is called a *value* of **T**. The f_i of the pair is the occurrence of v_i in **T** and is called the *frequency* v_i. Selectivity of values between v_a and v_b is to count the number of tuples in **R** that satisfy $v_i \leq$ **R.X** $\leq v_b$. In other words, it is to calculate the summation of the frequencies of values between v_a and v_b. A synopsis **S** of the instance *r* is a compression of **T** and its size is much less than that of **T**. Many techniques have been proposed to compute **S** from **T** in literature. We focus on two of them in the followings: the histogram synopses and the wavelet synopses.

Constructing Histogram Synopses

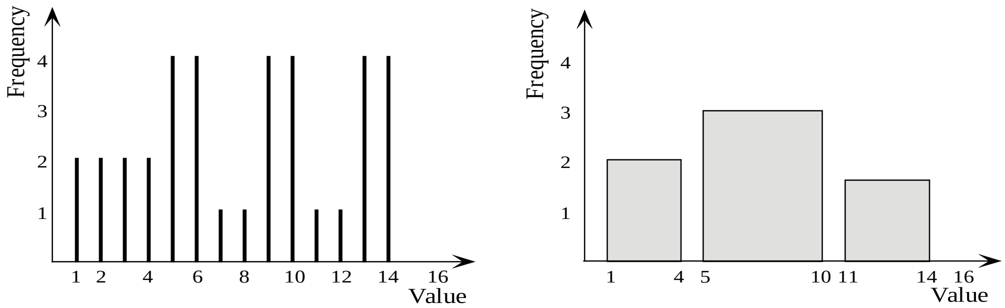
Histogram techniques have been used for selectivity estimation in commercial database management systems for few decades. Although estimation through histogram is fairly straightforward, it is not easy to construct a good histogram due to various data and query distributions.

We now provide a data distribution **T** as a running example. **T** and one of its histogram synopses are depicted in Figure 1.

The techniques for building the histogram synopses focus on *inner bucket estimation* and *bucket partition*. The common goal is to minimize the difference between the original and the approximate data. The inner bucket estimation is to approximate the distribution of value-frequency pairs that fall in one bucket. A generally adopted approximation is to assume the uniform value distribution and average frequency representation. That is values in one bucket are assumed to be spaced evenly. The frequencies in one bucket are represented uniformly by their average.

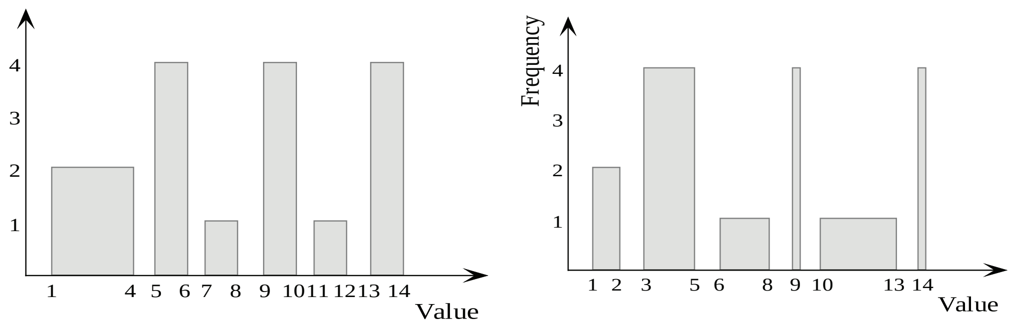
The bucket partition studies how to effectively partition the original data distribution, with a given bucket number. Traditional partition methods, including equiwidth and equidepth, have been widely used in commercial database environments (Kooi, 1980; Piatetsky-Shapiro & Connell, 1984). Other partition techniques, such as max-diff and v-optimal have been developed during recent years (Ioannidis, 1993; Jagadish, Jin et al., 2001). Techniques for constructing a nontrivial multidimensional histogram, that is, a histogram against multiple related attributes, have

Figure 1. Data distribution and histogram



a. Data disribution *b. Histogram with 3 buckets*

Figure 2. Various histograms



a. A perfect 6 buckest histogram *b. another 6 buckets histogram*

also appeared in literature since 1997 (Poosala & Ioannidis, 1997; Acharya, Poosala et al., 1999; Lee, Kim et al., 1999).

More buckets generally lead to less approximation error. As in Figure 2(a), the histogram with six buckets is considered a perfect one as it can accurately represent **T**. Under a fixed bucket number, different partition strategies also have great impacts on the histogram's performance. As in Figure 2(b), although the histogram also has six buckets, it is obviously much worse compared to the one in Figure 2(a).

Constructing Wavelet Synopses

Wavelet techniques have been used successfully in signal and image processing (Graps, 1995; Hubbard, 1996). Wavelet analysis combines the advantages of the Fourier transformation and windowed Fourier transformation. It adopts a wavelet prototype function, called a *mother wavelet*, to compress the original data to a synopsis. The simplest mother wavelet is the Haar wavelet which was proposed in 1909 by Alfred Haar. It has been used for selectivity estimation since 1998 and has been greatly developed thereafter.

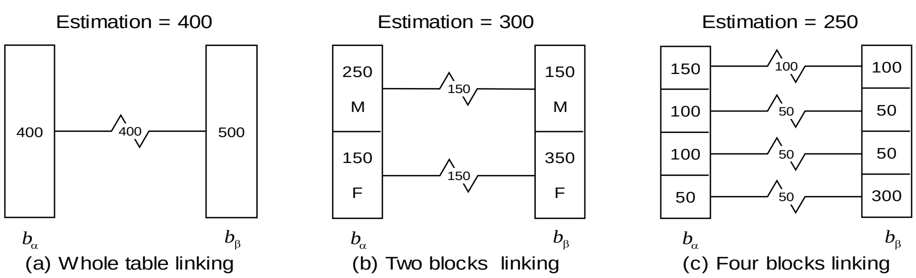
We still employ the data distribution **T** in our last example to show details of constructing Haar wavelet synopses. First, we apply Haar wavelet transformation on the frequency distribution of **T,** [2, 2, 2, 2, 4, 4, 1, 1, 4, 4, 1, 1, 4, 4, 0, 0]. Note that we add several 0 frequencies to make the number of **T** to be the power of 2. This original distribution will be considered having resolution level 4 (Figure 3). Note that the resolution represents the approximation of the original data. Data at higher resolution level preserve more original data information compared to data at lower level. The Haar wavelet transforms **A** by averaging the values of each pair, starting from the first two values of **A**, to get new *low-resolution* representations. It also records the difference of the pairs as coefficients on the same resolution level. By applying the transformation recursively, we reach the average value at resolution 0 with all the coefficient values are computed on the fly. Figure 3 shows details of the Haar wavelet transformation on **A**.

Thus the Haar wavelet transformation converts **A** to $\mathbf{W_A}$, [2.25, 0, −0.25, 0.25, 0, 1.5, 1.5, 2, 0, 0, 0, 0, 0, 0, 0, 0]. Each entry in $\mathbf{W_A}$ is called a *wavelet coefficient*. Second, we approximate $\mathbf{W_A}$ by keeping only those most significant coefficients to fit the space limitation of the synopses, and treat other coefficients as 0. For instance, to compress $\mathbf{W_A}$ with four coefficients space limit, we record a synopsis **S** as **S** = [(2.25, 1), (1.5, 6), (1.5, 7), (2, 8)]. The second number of the pairs in **S** indicates the non-zero coefficient's position. Applying inverse Haar wavelet transformation, we can get the approximation of the original frequency distribution **A**.

Figure 3. Haar wavelet decomposition

Resolution	Averages	Detail Coefficients
4	[2,2,2,2,4,4,1,1,4,4,1,1,4,4,0,0]	-
3	[2,2,4,1,4,1,4,0]	[0,0,0,0,0,0,0,0]
2	[2,2.5,2.5,2]	[0,1.5,1.5,2]
1	[2.25,2.25]	[-0.25,0.25]
0	[2.25]	[0]

Figure 4. Approximate link results on different partition



LINKING ESTIMATION THROUGH WAVELET

In this section, we introduce a record linking estimation method based on wavelet techniques. We start with a naïve approximation method. Then we propose our wavelet linking estimation technique.

A Naïve Approximation

This method is motivated by the folklore *blocking technique* used in the record matching. The basic idea is to subdivide records to be linked into a set of mutually exclusive subsets under same criteria that guarantees no matches occurring across blocks. The possible linking can therefore only happen between corresponding blocks. Given two such blocks, b_α and b_β, to link. Let $S(b_\alpha)$ and $S(b_\beta)$ represent the number of records. If there are no duplicate patient records in each block, we simply estimate the selectivity of the possible linked records with $\min(S(b_\alpha), S(b_\beta))$. If however there exists duplicate records, we estimate the linking selectivity with $\max(S(b_\alpha), S(b_\beta))$. For simplicity reason, we assume no record duplication in this article.

The blocking estimation method can be further tuned when we partition the records into more detail buckets. We show this in the following example.

Example 1. *Two blocks b_α and b_β with $S(b_\alpha) = 400$ and $S(b_\beta) = 500$. Suppose the exact linked records between b_α and b_β is 200. A rough selectivity estimation of linking is 400 as shown in Figure 4(a). After partitioning each table into two blocks according to patient's gender, as shown in Figure 4(b), we can improve our estimation to 300. Figure 4(c) shows further partitions make a closer 250 selectivity estimation.*

From the above observation, we conclude the following property.

Property 1: *Given two patient records to link. More partitions on the record guarantees a monotonic approach to the exact linking result.*

Based on this property, we develop the naïve approximation method. Formally, to estimate the linking between two patient records R_α and R_β, we partition them using a hash function *H*, which guarantees that no records will be linked across different hash bucket. Let $B_H(R_\alpha)$ and $B_H(R_\beta)$ represent the partitioned buckets. Then for each corresponding bucket pair, $B_H(R_\alpha)$ and $B_H(R_\beta)$, we choose the one with the smaller size as the selectivity estimation. Details are presented in Algorithm 1. Suppose the bucket number of the partition is *N*. The naïve method's time complexity is *O(N).*

Algorithm 1 Naïve Estimation
Input:
Records R_α, R_β and hash function H
Output:
Selectivity estimation of linked records
Description:
1: $B_H(R_\alpha)$:= Partition R_α by H; $B_H(R_\beta)$:= Partition R_β by H;
2: **for** each bucket $b_\alpha \in B_H(R_\alpha)$ **do**
3: find the bucket $b_\beta \in B_H(R_\beta)$, which has same hash key as b_α;
4: estimation + = $\min(S(b_\alpha), S(b_\beta))$;
5: **end for**
6: Output estimation

Applying the naïve method, we can partition the patient records of the local databases beforehand and store those partitions as data synopses for future linking estimation. It is obvious from Property 1 that the accuracy of the naïve method largely depends on the number of buckets. However, more buckets require more storage space, which is strictly limited by the available space in the global multidatabase server. We normally can not afford enough space for the nave method. This motivates us to develop the following wavelet method. For simplicity reason, we will use the bucket number to represent the size of available storage space hereafter.

Constructing Wavelet Synopses

To construct the wavelet synopses with a given bucket number B, we first partition the patient records **T** with the hash function *H*, as in the naïve method. Counting the number of records falling in each bucket, we organize them into a one-dimensional array $A(T)$. Transferring this one-dimensional array using the Haar wavelet, we get the coefficient array. For each coefficient *i*, suppose its resolution is *j*, we normalize the coefficient by . Then we choose the largest number of B/2 coefficients, in absolute normalize value, as well as their positions to store as the data synopses. In fact this is provably optimal with respect to minimizing the overall root-mean-squared error (i.e., L^2-norm average error) in data compression (Stollnitz, DeRose et al., 1996).

The procedure of wavelet estimation is quite similar to that of the naïve method. Details are presented in Algorithm 2. Note that we omit the procedures of constructing the wavelet synopses and reconstructing the original data from the synopses as they are covered in the background section.

Algorithm 2 Wavelet Estimation
Input:
Wavelet synopses W_α of records R_α and W_β of records R_β
Output:
Selectivity estimation of linked records
Description:
1: $B_H(R_\alpha)$:= Compute approximate original buckets of R_α from W_α;
2: $B_H(R_\beta)$:= Compute approximate original buckets of R_β from W_β;
3: **for** each bucket $b_\alpha \in B_H(R_\alpha)$ **do**
4: find the bucket $b_\beta \in B_H(R_\beta)$, which has same hash key as b_α;
5: estimation += min($S(b_\alpha)$, $S(b_\beta)$);
6: **end for**
7: Output estimation

Given a naïve hash partition with N buckets and a wavelet synopsis with B buckets. The time complexity on approximating the number of records of any hash bucket from the wavelet synopses is log *B*. Thus the time complexity of Algorithm 2 is $O(N \log B)$, where $N \gg B$.

The wavelet estimation can still be very time consuming while *N* is very large. This can be prohibitive in some online query processing situation. Thus in the next section, we propose our histogram estimation which always guarantees a fixed estimation time.

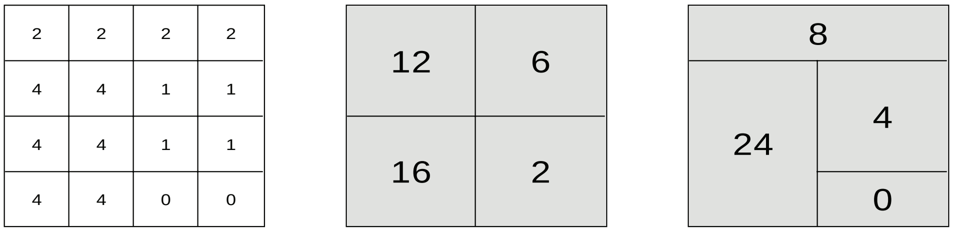
LINKING ESTIMATION THROUGH HISTOGRAM

In this section, we propose our linking estimation method based on the multidimensional histogram techniques. We first introduce constructing the histogram from the original patient records. Then we present how to estimate the selectivity of the linking records from the histogram.

A Multidimensional Histogram

Similar to the partition-and-linking idea, we construct the histogram through partitioning the original patient records first. However, instead of using only one hash function to partition the records into many buckets, we use several hash functions to partition the records into a multidimensional grids. That is, we partition the records such that no linking happens across grids. Examples of these hash functions can be the date of birth, the gender, and so forth. Counting the number of records falling in each grid, we get a multidimensional array. Due to the multidatabase server's space limitation, we can not store the whole array as our data synopses.

Figure 5. The 4 buckets histograms



(a) Original grid *(b) One 4 bucket histogram* *(c) Another 4 bucket histogram*

Instead we approximate the array through a B buckets histogram. In each histogram bucket, we adopt the uniform distribution assumption. That is, the number of records of each grid that belongs to the bucket will be represented by the average number of records in the bucket. Figure 5(a) is an example of patients records partitioned into 16 grids. The numeric number represents the number of records in that grid. Suppose we want to construct a four buckets histogram. Figure 5(b) is one possible histogram. Figure 5(c) is another possible histogram. Note that the number in each bucket represents the total number of records in that bucket.

Since we use the uniform distribution assumption, the more uniform intra-bucket distribution, the better estimation from the histogram. We formalize the problem as follows. Given a B buckets histogram, considering a bucket b_i, $1 \leq i \leq B$, suppose b_i includes n grids and the numbers in those grids are defined as $g_1,..., g_n$. The skew of bucket b_i is defined as: , here is the average of all g_j, $(1 \leq j \leq n)$. The skew of the histogram is the summation of each bucket's skew: . To acquire an accurate estimation, the goal is to find a histogram which has minimal skew, under a given number of buckets. For example, the histogram in Figure 5(c) is the best four buckets histogram since its skew is 0.

Unfortunately, constructing such a minimal skew histogram is a difficult problem that is provably NP-hard for even simple instances (Muthukrishnan, Poosala et al., 1999). Many approximate approaches have been proposed. In this article, we adopt a binary space partition techniques, the *min-skew* partition (Acharya et al., 1999). It is a heuristic partition method. At each step, it partitions one of the existing buckets into two sub-buckets. This partition is guided by minimizing the current total histogram skew. We demonstrate the min-skew partition through constructing a four bucket histogram on former example. Initially, consider the histogram has only one big bucket that holds all those grids. Its skew is $sk_1 = 35$ as shown in Figure 6(a). To divide it into two buckets, we have all together six choices: three different horizontal partition and three different vertical partitions, as pointed by the arrows in Figure 6(a). After calculating, we choose the one that can reach maximal skew decrease as shown in Figure 6(b). Continuing with similar steps, we reach the final four buckets histogram. The histogram partitioned by min-skew has a total skew 2, which is greater than the optimal partition in Figure 5(c).

After partition, we store the bucket information to form the histogram synopses. For each bucket, we only store its range, that is, the two coordinate values on each dimension, and the number of records in that bucket.

Estimation Through Histogram

The selectivity estimation through histogram synopses is much quicker compared with wavelet synopses. As mentioned before, we assume the average distribution in each histogram bucket. To estimate the number of possibly linked records of two buckets, we only need to multiply the space of the overlapping area with the smaller density of the two buckets. That is, given two buckets, b_α whose range is v_α with n_α number of records, and b_β whose range is v_β with n_β number of records. The number of linked records of these two buckets is estimated as . The algorithm of linking selectivity estimation through histogram synopses is as follows.

Algorithm 3 Histogram Estimation
Input:
Histogram H_α of records R_α and H_β of records R_β
Output:
Selectivity estimation of linked records
Description:
1: **for** each bucket $b_\alpha \in H_\alpha$ **do**
2: **for** each bucket $b_\beta \in H_\beta$ **do**
3: **if** b_α overlaps with b_β **then**
4: estimation +=
5: **end if**
6: **end for**
7: Output estimation

It is easy to see that, given B buckets histogram, the estimation's time complexity is $O(B^2)$. This means that the estimation through histogram synopses can always guarantee a quick response no matter how large the original record size is.

EXPERIMENTS

In this section, we study the performance of our methods on linking selectivity estimation of generic and health data. We implement our two types of synopses, wavelet and histogram, respectively. The wavelet synopsis is a one-dimension synopsis. The histogram synopsis is a multidimension synopsis. We first evaluate these two synopses against the generic data. Then we compare them against the health data. All algorithms are implemented through GCC 4.0.2. The experiments are performed on a Pentium IV 3.2 Ghz CPU with 1 GB memory computer.

The data used in our experiment is synthetic data. Specifically, we generate two sets of data. One includes generic person record tables with the person's name, date of birth (DOB), and gender. The other is the lung cancer patient records which includes patients' details as well as the treatments received.

Experiment Results on Generic Data

In this experiment, we generate two classes of tables to link, A and B. Each class includes three tables with different distribution on a person's

Figure 6. The MinSkew partition

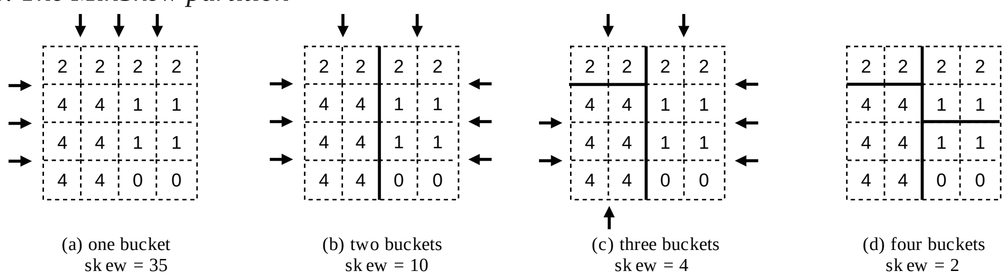


Figure 7. Linking estimation on generic data

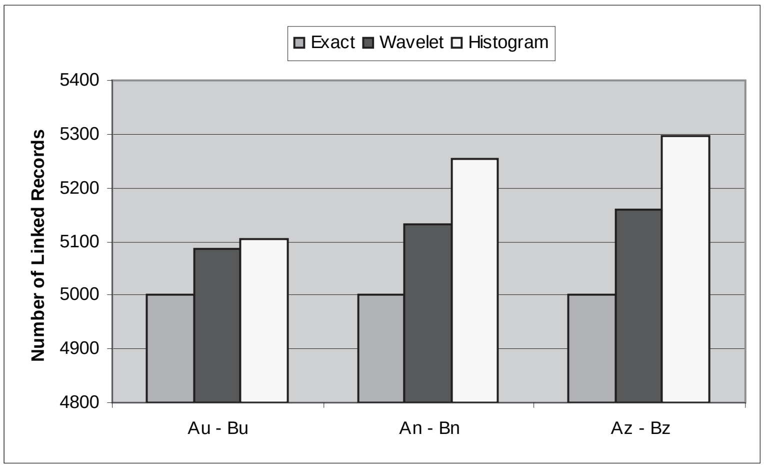


Figure 8. Linking estimation time used by various synopses

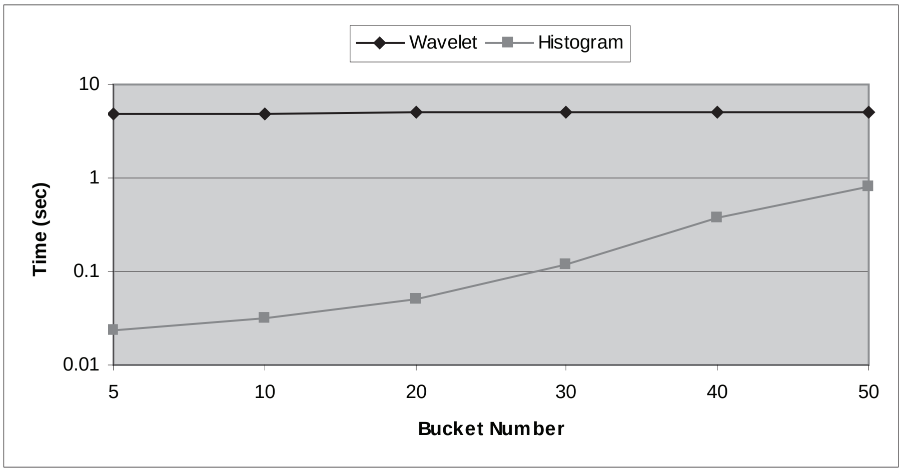
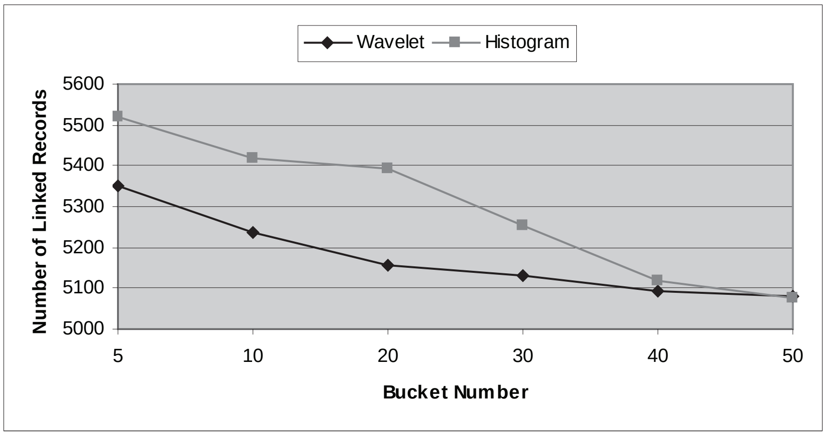


Figure 9. Linking estimation for various bucket numbers



DOB ranging from 1900-1-1 to 1999-12-31. We use A_u to represent the table in class A with uniform DOB distribution. A_n represents the table with standard normal DOB distribution. A_z represents the table with zipf DOB distribution. Here we set the zipf parameter $z = 1$, which means a medium skew distribution. For tables in class B, we also generate three similar tables and adopt the same naming convention as in class A. Each table has 10,000 tuples, which represents 10,000 different people recorded in that table. To compare our methods, we assign a uniform storage space to both the wavelet and histogram synopses. We use the bucket number B of the histogram synopses to measure the assigned space. Since we construct a four-dimensional histogram, we need to store nine numbers in each bucket of our histogram synopses which includes every two extents on each dimension and the number of records in that bucket. Note that in the wavelet synopses, we keep both the value and the position if storing a coefficient. Thus when we mention the synopses size B in our experiment, we mean constructing a B buckets four-dimensional histogram synopses or a wavelet synopses with coefficients.

Linking the six tables from two classes, we get nine linking results under each linking method accordingly. Figure 7 shows selected results of linking three pairs of tables that have same DOB distribution. The synopses size is $B = 30$. Note that we omit the linking results between other tables as those do not indicate significant difference. From the experiment results, the wavelet and histogram synopses both demonstrate good approximate results with reasonably small errors. Of all the three linking between different distribution types depicted in Figure 7, the approximation for linking between uniform distributed data generates the smallest error. This is because that when data have skew distribution, such as normal or zipf distribution, a large estimation error will occur while approximating the selectivity of buckets that has many people's records. We also note that the wavelet synopses normally generate better estimation results compared to the histogram synopses. This is partly due to rough approximation techniques, the uniform assumption, used in estimating the values inside a histogram bucket.

However, as mentioned before, the estimation time based on histogram synopses is much shorter compared to that on wavelet synopses. Figure 8 shows the comparisons between the estimation time on linking tables A_n and B_n. It is obvious that estimation based on histogram synopses is remarkably quicker than that based on wavelet synopses.

Our last experiment on the generic data is to evaluate the influences of synopses size on the estimation accuracy. Again, we test the linking estimation between the tables, A_n and B_n, through various wavelet and histogram synopses whose sizes range from 5 to 50. Figure 9 shows the result. It is interesting to note that the storage space changing has more influences on the estimation accuracy of the histogram synopses.

Experiment Results on Synthetic Health Data

In this part, we generate health data to emulate the real clinic and hospital patients' records. To imitate real situations, we also generate some errors and typographical errors in the person's records of the health data. For example, a patient's DOB may be recorded as 6019-3-82, which should be 1960-3-28. A patient's gender information may be lost. We generate two pairs of such patient databases for linking. The first pair includes database HA and HB, which has 3000 and 667 records individually. This represents the situation when a large number of linked records exists in the databases. The second pair includes tables HC and HD, which has 32 and 24 records individually. This represents the situation when a small number of linked records exists in the databases. To acquire the exact linking results, we use our health data integrator as introduced by Hansen, Pang et al. (2005). We apply 40 buckets to construct our wavelet and histogram synopses. Figure 10 shows the results. It indicates that when the number of actually linked records is large, both wavelet and histogram synopses can provide good estimations. The approximation quality is however decreased when the number of linked records is small.

CASE STUDY

In this section, we demonstrate how the presented link selectivity estimation techniques can facilitate estimate more complex aggregate queries in a health multidatabase system. We start by assuming our two synthetic health databases, HA and HB, as two component databases of a health system. Database HA records the information of patients having cancer A at any of the three stages, from I to III. Database HB records similar information of patients having cancer B at any of the two stages, from I to II. In both databases, information such as patient name, date of birth, gender, diagnosed stage, treatment results, and so forth, is recorded.

An investigation about the potential relationships between cancer A and B naturally leads to queries involving both databases. A typical analysis is to find the number of patients having both cancers at various stage combinations. This incurs a group of link and aggregate queries. Namely we link databases HA and HB first to find patients who are recorded in both databases. Then we issue six aggregate queries to acquire detail numbers. Figure 11 shows the exact query results, that is, the detail distribution of the 647 (refer to Figure 10) linked patients.

There exist many approximate techniques processing complicate aggregate queries across several databases. The unique global ID for each record is a de facto assumption in those

Figure 10. Linking estimation on health data

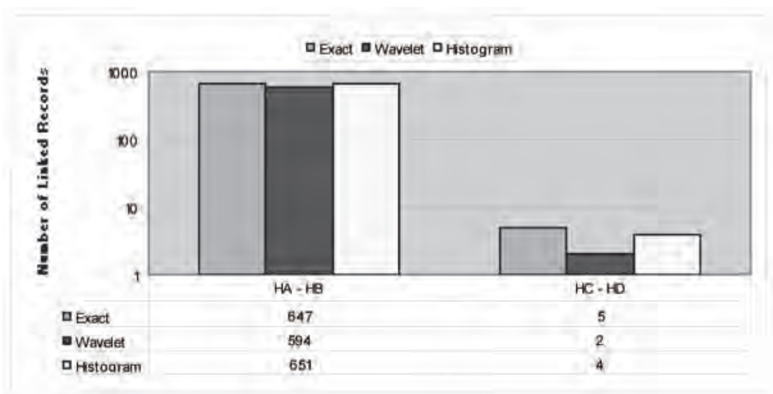


Figure 11. Exact answer for aggregate query

Cancer A / Cancer B	Stage I	Stage II	Stage III
Stage I	225	117	95
Stage II	111	50	49

techniques. However this assumption is usually not held in a multidatabase system, which makes those existing techniques no more applicable. Now combining our selectivity estimation techniques, we can process multidatabase aggregate queries as follows:

- Compute the approximate linking radio δ for each database, which is defined as the number of approximate linked results divided by the number of records, through our link estimation technique.
- Compute the initial approximate results of the aggregate query q_i by assuming unique patient record, through any existing approximate query processing technique.
- Compute the final approximation results q_f, $q_f = \delta_* q_i$.

Figure 12 shows the approximate results of the aggregate query involving database HA and HB.

RELATED WORKS

Medical record linking links patient records across organizations. Current techniques are those for data linkage problems, which are also defined as object identification, data matching, and so forth. Much research has been done for the data linkage problem (Fellegi & Sunter, 1969; Newcombe, 1967; Newcombe, Kennedy et al., 1959; Winkler, 1999). Existing techniques can be classified into two groups (Cochinwala, Dalal et al., 2001):

- *searching techniques* involve searching for potentially linkable pairs of records. Typical techniques include a nested-loop join, band join, blocking, and so forth (DeWitt, Naughton et al., 1991; Newcombe, 1988).
- *matching techniques* involve deciding whether or not a given record pair is correctly matched. Typical models are probabilistic and nonprobabilistic (Dey, Sarkar et al., 1998; Monge & Elkan, 1996; Newcombe, 1967).

CONCLUSION AND FUTURE WORK

In this article, we present two approximate techniques for the estimation in a multidatabase system. We first provide a naïve technique on the selectivity estimation. Based on this, we develop two techniques, namely wavelet and histogram. Given a limited storage space, our method can construct efficient wavelet or histogram data synopses. Experiment results show that our approximation method can quickly provide good estimations. We also conclude from our experiments that although wavelet synopses are generally more accurate than histogram ones, the approximation time used by wavelet synopses is much longer. Thus histogram techniques may be more promising when used in an online environment, where prompt answers are expected.

Currently we assume that original data is in a relatively static environment. Thus whenever change happens, we just reconstruct new data

Figure 12. Approximate answer for aggregate query

Cancer A / Cancer B	**Stage I**	**Stage II**	**Stage III**
Stage I	210	132	88
Stage II	104	55	62

synopses and discard old ones. However, how to efficiently maintain an up-to-date synopsis in a dynamic environment will be one of our future works. Another possible future work is to investigate how to process nonaggregate approximate queries, such as join, in a multidatabase system.

ACKNOWLEDGMENT

We would like to thank Marilla O'Dwyer for providing the real health data and helpful discussions.

REFERENCES

Acharya, S., V. Poosala, et al. (1999). Selectivity estimation in spatial databases. In *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data*.

Chakrabarti, K., Garofalakis, M. N., et al. (2000). Approximate query processing using wavelets. *The International Journal on Very Large Data Bases.*

Cochinwala, M., Dalal, S., et al. (2001, March). *Record matching: Past, present and future*.

DeWitt, D., Naughton, J., et al. (1991). An evaluation of non-equijoin algorithms. In *Proceedings of the 17th Conference on Very Large Databases.* Los Altos, California/Barcelona: Morgan Kaufman.

Dey, D., Sarkar, S., et al. (1998). Entity matching in heterogeneous databases: A distance based decision model. In *Proceedings of the Thirty-First Annual Hawaii International Conference on System Sciences-Volume 7 (HICSS '98)*, Washington, DC. IEEE Computer Society.

Fellegi, I. P., & Sunter, A. B. (1969). A theory for record linkage. *Journal of the American Statistical Association, 64*, 1183-1210.

Graps, A. (1995). An introduction to wavelets. *IEEE Computational Science and Engineering, 2*, 50-61.

Hansen, D. P., Pang, C., et al. (2005). *HDI: Integrated services for health data.* Paper presented at the International Conference on Machine Learning and Cybernetics (ICMLC).

Hubbard, B.B. (1996). *The world according to wavelets.* A.K.Peters.

Ioannidis, Y. (1993). Universality of serial histograms. In *Proceedings of the 19th International Conference on Very Large Data Bases.*

Jagadish, H. V., Jin, H., et al. (2001). Global optimization of histograms. In *Proceedings of the 2001 ACM SIGMOD International Conference on Management of Data*.

Kooi, R. P. (1980). *The optimization of queries in relational database*. Case Western Reserver University.

Lee, J.-H., Kim ,D.-H., et al. (1999). Multi-dimensional selectivity estimation using compressed histogram information. In *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data*.

Monge, A. E., & Elkan, C. (1996). *The field matching problem: Algorithms and applications. Knowledge discovery and data mining.*

Muthukrishnan, S., Poosala, V., et al. (1999). On rectangular partitionings in two dimensions: Algorithms, complexity, and applications. *Lecture Notes in Computer Science, 1540*, 236-256.

Newcombe, H. B. (1967). Record linkage: The design of efficient system of linking records into individual and family histories. *American Journal of human Genetics, 19*, 335-359.

Newcombe, H. B. (1988). *Handbook of record linkage: Methods for health and statistical studies, administration, and business.* Oxford University Press.

Newcombe, H. B., Kennedy, J. M., et al. (1959). Automatic Linkage of Vital Records. *Science, 130*(3381), 954-959.

Piatetsky-Shapiro, G., & Connell, C. (1984). Accurate estimation of the number of tuples satisfying a condition. In *Proceedings of the 1984 ACM SIGMOD International Conference on Management of Data,* Boston.

Poosala, V., & Ioannidis, Y. E. (1997). Selectivity estimation without the attribute value independence assumption. *The International Journal on Very Large Data Base*s.

Stead, W. W., Miller, R. A., et al. (2000). Integration and beyond. *Journal of American Medical Informatics Association, 7*(2), 135-145.

Stollnitz, E. J., DeRose, T. D., et al. (1996). *Wavelets for computer graphics.* Morgan Kaufmann.

Winkler, W. (1999). *The state of record linkage and current research problems.*

This work was previously published in International Journal of Healthcare Information Systems, Vol. 2, Issue 4, edited by J. Tan, pp. 59-72, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.26
Data Mining Medical Information:
Should Artificial Neural Networks be Used to Analyse Trauma Audit Data?

Thomas Chesney
Nottingham University Business School, UK

David Chesney
Freeman Hospital, UK

Kay Penny
Napier University, UK

Nicola Maffulli
Keele University School of Medicine, UK

Peter Oakley
University Hospital of North Staffordshire, UK

John Templeton
Keele University School of Medicine, UK

Simon Davies
University of Birmingham Research Park, UK

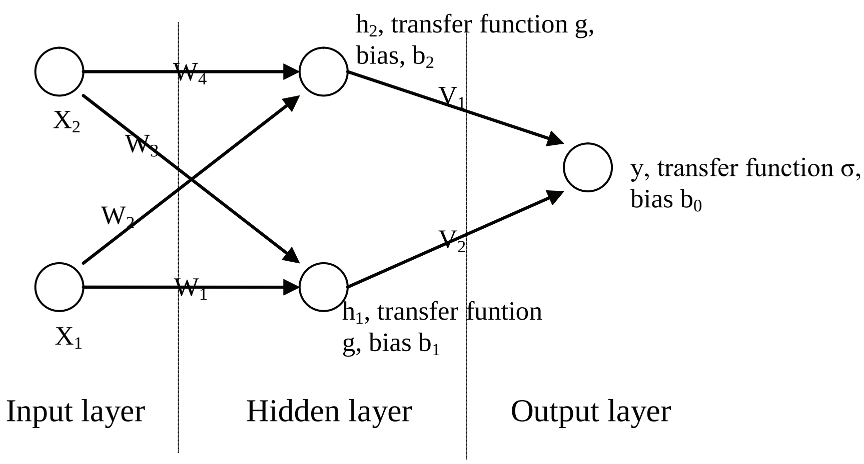
ABSTRACT

Trauma audit is intended to develop effective care for injured patients through process and outcome analysis, and dissemination of results. The system records injury details such as the patient's sex and age, the mechanism of the injury, various measures of the severity of the injury, initial management and subsequent management interventions, and the outcome of the treatment including whether the patient lived or died. Ten years' worth of trauma audit data from one hospital are modelled as an Artificial Neural Network (ANN) in order to compare the results with a more traditional logistic regression analysis. The output was set to be the probability that a patient will die. The ANN models and the logistic regression model achieve roughly the same predictive accuracy, although the ANNs are more difficult to interpret than the logistic regression model, and neither logistic regression nor the ANNs are particularly good at predicting death. For these reasons, ANNs are not seen as an appropriate tool to analyse trauma audit data. Results do suggest, however, the usefulness of using both traditional and non-traditional analysis techniques together and of including as many factors in the analysis as possible.

INTRODUCTION

An Artificial Neural Network (ANN) attempts to model human intelligence using the neurons in a human brain as an analogy. ANNs have been described numerous times (Lee & Park, 2001; Bose & Mahapatra, 2001; Setiono, Thong, & Yap, 1998; Lee, Hung Cheng, & Balakrishnan, 1998), but a brief description is that the network accepts a series of factors as input, which it processes to output a probability that the input belongs to a certain class. For example, in the case of the trauma data analysed in this study, the characteristics of the trauma are the input to the ANN, which then outputs the probability that the patient will die. The processing is done by layers of neurons (called hidden layers) which apply a weight to each input factor according to how important that factor is in calculating the classification probability. The weight is learned by the network during its training. In training, a series of input factors to which the correct classification is known is fed into the ANN. The ANN then adjusts its weights to minimise the error between its predicted classification and the known correct class. A pictorial representation of an ANN is shown in Figure 1.

An ANN has the potential to discriminate accurately between patients who will live and those who will die, and can capture complex relationships between factors that traditional analysis methods may miss. However, there are two potential problems with using ANNs to analyse trauma data. First, they are affected by imbalances in the data (Fu, Wang, Chua, & Chu, 2002). A common characteristic of medical data is its imbalance (Cios & Moore, 2002). What this means is that the attribute of interest to data miners is likely to be present only in a minority of records in the dataset. In the case of the trauma data discussed here, a much higher percentage of patients lived than died. The second disadvantage with neural networks is that it is very difficult to explain and to justify the model. In other words, after train-



Output of h_1, $h_{1out} = g(w_1x_1 + w_3x_2 + b_1)$
Output of h_2, $h_{2out} = g(w_2x_1 + w_4x_2 + b_2)$
Output of node y, the output layer which uses the sigmoid function and is the probability of a certain class, for instance DEATH = 1, given the input vector $\overline{x}$,
$p(DEATH = 1 \mid \overline{x}) = s(v_1(h_{1out}) + v_2(h_{2out}) + b_0)$

ing, it is difficult to explain why one neuron is weighted in a particular way. The ANN often does not have an underlying probabilistic model (Giudici, 2003), and evaluation of the model usually is done by examining its output — its predictive accuracy. In mining medical data, consideration is also usually given to the model's sensitivity and specificity and less frequently to the model's positive and negative predictive value.

Despite this, ANNs frequently have been used in healthcare. For example, Lee and Park (2001) applied neural networks in order to classify and predict the symptomatic status of HIV/AIDS patients; Baxt, Shofer, Sites, and Hollander (2002) used an ANN to aid in the diagnosis of acute myocardial infarction; Lapuerta, Azen, and LaBree (1995) used neural networks to predict the risk of coronary artery disease. DiRusso, Sullivan, Holly, Nealon Cuff, and Savino (2000) used ANNs to predict survival in trauma patients that was similar to the application in the current study. ANNs are just one of a number of techniques that are known collectively as machine learning. Others long have been used in medical applications to improve decision making (Konomenko & Kular, 1995). Lavrac (1999) lists some applications: diagnosis and prognosis in oncology (Bratko & Konomenko, 1987), liver pathology (Lesmo, Saitta, & Torasso, 1982), neuropsychology (Muggleton, 1990) and gynecology (Nunez, 1990). Breault, Goodall, & Foss (2002) used a decision tree to examine a diabetic data warehouse. Imberman, Damask, and Thompson (2002) identified association rules in a head trauma dataset. The aim of this article is to examine how analysing trauma data with an ANN compares with a more traditional logistic regression analysis.

Trauma Audit and Research Network

In 1991, in response to a report from the Royal College of Surgeons of England, the government supported a pilot system for treating trauma patients at the North Staffordshire Hospital (NSH) in Stoke-on-Trent in the UK (Oakley, MacKenzie, Temppleton, Cook, & Kirby, 2004). At the same time, a system for collecting trauma data from a set of core hospitals in England and Wales was established at the University of Manchester. This project, the Major Trauma Outcome Study (UK), grew in size with more hospitals becoming involved. By the late 1990s, one-half of all trauma-receiving hospitals in England and Wales returned data to the now renamed Trauma Audit & Research Network (TARN). NSH served as a core hospital and has contributed trauma audit data to TARN since that time.

Trauma is the most common cause of death in those under 40, and many of these are preventable (TARN, 2004). Trauma audit is intended to develop effective care for injured patients through process and outcome analysis and through dissemination of results. The system records injury details, such as the patient's sex and age, the mechanism of the injury, various measures of the severity of the injury, initial management and subsequent management interventions, and the outcome of the treatment, including whether the patient lived or died. Several analyses that used traditional statistical techniques have been performed on the data and published (Oakley et al., 2004; Templeton et al., 2000). Oakley et al. (2004) analysed six years of NSH trauma audit data from April 1992 through March 1998 in order to identify factors related to mortality and to examine whether there was a longitudinal trend in mortality. Multiple logistic regression analysis evidenced eight factors as determinants of mortality. In order of importance, these were age, head Abbreviated Injury Score (AIS), chest AIS, abdominal AIS, calendar year of admission, external injury AIS, mechanism of injury, and primary receiving hospital. However, the data from NSH never have been modelled as an ANN.

This article continues as follows. The next section describes the method of preparing and analysing the data by ANN and logistic regression. The results are then presented, compared, and discussed.

METHOD

The study concerned trauma audit data from patients treated at the North Staffordshire Hospital from 1993 to 1999 and from 2001 to 2004 (the gap was due to a lack of resources during this period, which affected data collection). The data were limited to those most severely injured patients, since there should be most to learn from such patients. Patients with an Injury Severity Score (ISS) greater than 15 were included. All the injury scores (Abbreviated Injury Scores, ISS, and Glasgow Coma Scores) were assigned, checked, and entered into the database by trauma audit staff under the guidance of three clinicians (three of the authors) as soon as possible after the injury. All were well-trained in the use of the scoring systems. (For an explanation of all of these scoring techniques, see Baker [1974] and Teasdale and Jennett [1974]) Very little data were missing, as the summary in Table 1 shows.

The criteria for inclusion just described left 1,658 records in the data set with which to train and test different ANN models. The factors shown in Table 1 were included as input, with the output being whether the patient lived or died. If the data are well-balanced with approximately equal numbers of patients who lived and died, learning algorithms will have a better chance of finding patterns that distinguish the two groups. As has been explained already, the trauma audit data are very imbalanced — 79% of patients lived, 21% died. Machine learning techniques usually are biased toward the majority class, as these dominate during training (Fu et al., 2002). For example, a classifier could achieve 79% accuracy with this data by always predicting "live." A number of techniques has been examined to deal with this (Guo & Viktor, 2004): undersampling the majority class (Kubat & Matwin, 1997), oversampling the minority class (Ling & Lee, 1998), doing both together (Chawla, Bowyer, Hall, & Kegelmeyer, 2002), and boosting the learning algorithm (Schapire, 2001). The data were balanced by adjusting the proportion of the patients who died by a factor of 3.685. However, it has been suggested that record balancing has no impact on neural network models (Fu et al., 2002).

The data mining package SPSS Clementine 7.0 was used to create the ANNs. This software allows the user to specify any ANN model (the model consists of the number of hidden layers of neurons and the number of neurons in each hidden layer), and the learning rates (alpha and eta, which specify how the weights change during training). All the neurons are fully connected and each is a feed-forward, multi-layer perceptron that uses the sigmoid transfer function (Watkins, 1997). The learning technique used is back propagation. This means that, starting with the given topology, the network is trained, then a sensitivity analysis is performed on the hidden units, and the weakest are removed (Watkins, 1997). This training/removing is repeated for a set length of time. A k-fold cross-validation technique was used to test the ANN models, with k set to five. This is good practice when building neural networks with medical data (Cunningham, Carney, & Jacob, 2000). Using this technique, the data were split into five groups. When splitting the dataset, those patients who lived were selected randomly and independently of those patients who died in order to keep the same proportion of live/die outcomes in each set. Four of the sets were used to train each ANN model; the fifth was used to test it. This then was repeated another four times so that each group was used to test the model once. The overall performance of the model was then the average performance of the five tests. Four different ANN models were tested in this way. To decide which ANN models to test, a range of models was trained with a random sample of two-thirds of the data and tested on the remaining one-third. Again, the split was made to keep the same proportion of live/die

Table 1. NSH trauma audit data (note that high AIS relates to an increase in severity, whereas high GCS relates to a decrease in severity)

Input factors		Percentage complete	Count (Percentage)	
Sex	Male	100	1244 (75)	
	Female		414 (25)	
Type of injury	Blunt	99.94	1608 (97)	
	Penetrating		50 (3)	
Referred from another hospital (yes or no)		100	464 (28) referred	
Mechanism group		100	Motor vehicle crash	829 (50)
			Fall greater than 2m	249 (15)
			Fall less than 2m	282 (17)
			Assault	50 (3)
			Other	249 (15)
Age group		99.94	0-15 years	149 (9)
			16-25 years	381 (23)
			26-35 years	265 (16)
			36-50 years	315 (19)
			51-70 years	332(20)
			71 years and over	216 (13)
Year of admission		100	1992	83 (5)
			1993	166 (10)
			1994	83 (5)
			1995	166 (10)
			1996	216 (13)
			1997	166 (10)
			1998	66 (4)
			2001	166 (10)
			2002	182 (11)
			2003	232 (14)
			2004	116 (7)
			2005	17 (1)

continued on following page

Table 1. (continued)

Month of admission	100	Jan	149 (9)
		Feb	116 (7)
		Mar	133 (8)
		Apr	133 (8)
		May	166 (10)
		Jun	133 (8)
		Jul	149 (9)
		Aug	133 (8)
		Sept	116 (7)
		Oct	149 (9)
		Nov	133 (8)
		Dec	149 (9)
Day of admission	100	Sun	249 (15)
		Mon	216 (13)
		Tues	232 (14)
		Wed	199 (12)
		Thurs	264 (16)
		Fri	249 (15)
		Sat	249 (15)
Time of admission group	100	0:00-3:00	431 (26)
		4:00-7:00	50 (3)
		8:00-11:00	116 (7)
		12:00-15:00	481 (29)
		16:00-19:00	265 (16)
		20:00-23:00	348 (21)

continued on following page

Table 1. (continued)

AIS Score		0	1	2	3	4	5	6
AIS Head	100	348 (21)	50 (3)	99 (6)	199 (12)	415 (25)	564 (34)	0 (0)
AIS Face	100	1111 (67)	265 (16)	232 (14)	66 (4)	0 (0)	0 (0)	0 (0)
AIS Neck	100	1625 (98)	17 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
AIS Chest	100	1011 (61)	50 (3)	50 (3)	249 (15)	216 (13)	83 (5)	0 (0)
AIS Abdomen	100	1376 (83)	50 (3)	83 (5)	83 (5)	50 (3)	17 (1)	0 (0)
AIS Spine	100	1310 (79)	0 (0)	166 (10)	116 (7)	33 (2)	17 (1)	0 (0)
AIS Upper limbs	100	1127 (68)	83 (5)	282 (17)	166 (10)	0 (0)	0 (0)	0 (0)
AIS Lower limbs	100	1111 (67)	66 (4)	116 (7)	282 (17)	66 (4)	0 (0)	0 (0)
AIS External	100	1227 (74)	381 (23)	17 (1)	0 (0)	17 (1)	17 (1)	0 (0)
AIS CSpine	100	1509 (91)	0 (0)	66 (4)	50 (3)	17 (1)	0 (0)	0 (0)
AIS TSpine	100	1509 (91)	0 (0)	83 (5)	33 (2)	17 (1)	17 (1)	0 (0)
AIS LSpine	100	1525 (92)	83 (5)	33 (2)	17 (1)	0 (0)	0 (0)	0 (0)
Glasgow Coma Score		1	2	3	4	5	6	
GCS Eye	88.24	483 (33)	102 (7)	132 (9)	746 (51)	N/A	N/A	
GCS Verbal	88.24	395 (27)	205 (14)	59 (4)	205 (14)	600 (41)	N/A	
GCS Motor	88.24	219 (15)	73 (5)	73 (5)	102 (7)	263 (18)	732 (50)	
Outcomes								
Alive or dead	98.97	344 (21) dead, 1296(79) alive						

outcomes in each set. These models differed on the number of hidden layers, neurons, and learning rates. The four best performing models were selected. Specific guidelines on how to choose these parameters were unavailable, as this was dependent upon the underlying data generating process (Berardi, Patuwo, & Hu, 2004), which is unknown, so experimentation like this is a reasonable approach.

The multiple logistic regression model was created using SPSS 11.0, and the factors shown in Table 1 were considered for inclusion. Such a model normally would not be tested using k-fold cross validation. In this case, the logistic regression model was created using the same random two-thirds of the data from before and tested on the remaining one-third.

In many data mining applications, the evaluation criterion is the percentage of correct classifications made by an algorithm. However, in medical data mining, consideration also must be given to the percentages of false positive and false negative classifications. The evaluation criteria for comparing the algorithms here are sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The sensitivity is the proportion of actual deaths that were correctly predicted to die, whereas the specificity is the proportion of actual survivors who were correctly predicted to survive. The positive predictive value (PPV) is the probability that a subject for whom the model predicts as dying actually will be a true death, and the negative predictive value (NPV) is the probability that a subject predicted as surviving actually will survive.

RESULTS

The number of hidden layers, the number of neurons in each layer, and the learning rates of the four ANN models are shown in Table 2. The results of the k-fold cross validations are shown in Table 3. Also shown in Table 3 are the results of the logistic regression analysis. The logistic regression model itself is shown in Table 4. This model shows that the mechanism group, age group, GCS-motor, Head AIS, Abdomen AIS, and External AIS all are associated independently with survival during a hospital stay. Those patients who have suffered a fall or assault have decreased odds of death compared to those involved in a motor vehicle crash. Those suffering a fall greater than two meters in height compared to those involved in a motor vehicle crash have much decreased odds of death (odds ratio = 0.27; 95% confidence interval is 0.13 to 0.53). A less severe GCS motor score is associated with decreased odds of death compared to a more severe injury. Those suffering abdominal or external injuries have increased

Table 2. Model characteristics

	ANN1	ANN2	ANN3	ANN4
Number of hidden layers	2	1	3	3
Number of neurons in layer 1	30	30	30	30
Number of neurons in layer 2	20	N/A	20	20
Number of neurons in layer 3	N/A	N/A	10	10
Alpha	0.5	0.5	0.3	0.4
Initial eta	0.5	0.5	0.5	0.5
High eta	0.9	0.9	0.8	0.8
Low eta	0.01	0.01	0.05	0.05
Eta decay	30	30	20	20

odds of death during a hospital stay. Correcting for the other factors included in the model, those with a head AIS of five or six have an odds ratio of 1.68 of death compared to those with a score of 0; however, those with head AIS scores of 1-4 appear to have decreased odds of death compared to those with a score of 0. Increasing age is associated with increased odds of death, especially for those over 70 years old. A cut-point of 0.2 was used in the logistic regression in order to increase the sensitivity (and thereby decrease the specificity).

All of the ANNs achieved roughly 80% overall predictive accuracy and slightly outperformed the logistic regression model in this respect, although this finding is examined more closely in the next section. There is no one agreed procedure for presenting ANN architectures with their weights, and in this case, it is not possible to show the four neural networks, since each was trained five times, and for each training session, the final weights were different. However, in an attempt to illustrate the ANNs, Table 5 shows the 10 most important factors in each when trained with two-thirds of the data used to create the logistic regression model. Apart from ANN2, these appear very similar.

DISCUSSION AND CONCLUSION

Implications for Practice

Interpreting Table 5 is difficult. While three of the four ANN models show Abdomen AIS to be the most important factor in determining whether a patient lives or dies, it should be remembered that it is a network of factors that calculates the output, and Abdomen AIS can be used only to make a prediction in conjunction with all the other factors. As the following demonstrates, the ANN is not designed to analyse individual factors in the way that logistic regression can. Apart from ANN2, the models show reasonable agreement in the most important factors. ANN2, however, suggests that time, day, and month of admission are important determinants of death. This appears counter-intuitive, although when these factors are removed as inputs and then the ANNs are retrained, the predictive accuracy of each drops to around 73%. Clearly, therefore, these factors do have an impact on building the classifier. It is possible to suggest ideas about why this might be (staff will be affected by tiredness at difference times of the day, staff and patients may be affected by the seasons, and staff turnover may have an impact

Table 3. K-fold cross validation results (note that for the analysis, death was considered the positive outcome)

	ANN1	ANN2	ANN3	ANN4	LR
True positives	44	41	41	44	84
True negatives	219	219	221	219	286
False positives	38	38	37	39	86
False negatives	26	30	29	26	23
Sensitivity	63	57	58	62	79
Specificity	87	85	86	85	77
Predictive accuracy	80	79	80	80	77
Positive predictive value	54	53	53	55	49
Negative predictive value	89	88	88	89	92

Table 4. Logistic regression model (note that high AIS relates to a increase in severity, whereas high GCS relates to a decrease in severity, and odd ratios refer to odds of death as opposed to survival)

Factor	No. of Patients	Odds Ratio (95% CI)	p-value
Mechanism			
Motor Vehicle Crash	497	1.00	0.004
Fall < 2m	145	0.68 (0.36 – 1.29)	0.234
Fall > 2m	159	0.27 (0.13 - 0.53)	0.000
Assault	33	0.34 (0.09 - 1.30)	0.115
Other	130	0.60 (0.31 - 1.15)	0.124
Agegroup			
0 - 15 years	89	1.00	<0.001
16 - 25 years	221	4.25 (1.45 – 12.50)	0.009
26 - 35 years	153	6.46 (2.09 - 20.00)	0.001
36 - 50 years	177	5.83 (1.90 – 17.86)	0.002
51 - 70 years	193	18.42 (6.06 - 56.02)	<0.001
71 years or over	131	87.27 (26.86 – 283.55)	<0.001
Arrival GCS - Motor			
1	151	1.00	<0.001
2	46	0.47 (0.21 – 1.02)	0.055
3	51	0.14 (0.06 – 0.32)	<0.001
4	70	0.07 (0.03 – 0.16)	<0.001
5	170	0.08 (0.04 – 0.15)	<0.001
6	476	0.03 (0.02 – 0.07)	<0.001
AIS – Head			
0	199	1.00	<0.001
1 – 2	84	0.31 (0.10 – 0.99)	0.048
3	116	0.28 (0.12 – 0.66)	0.004
4	247	0.38 (0.18 – 0.77)	0.008
5 – 6	318	1.68 (0.90 – 3.12)	0.101
AIS - Abdomen			
0	797	1.00	
1 – 5	167	2.31 (1.35 – 3.96)	0.002
AIS - External			
0	721	1.00	
1 – 6	243	1.85 (1.17 – 2.94)	0.009

Table 5. Factors important in each model

Order of importance	ANN1	ANN2	ANN3	ANN4
1	Abdomen AIS	Head AIS	Abdomen AIS	Abdomen AIS
2	Age group	Neck AIS	Chest AIS	Chest AIS
3	Chest AIS	Face AIS	Age group	External AIS
4	Arms AIS	Chest AIS	External AIS	Cervical spine AIS
5	External AIS	Abdomen AIS	Cervical spine AIS	Age group
6	Cervical spine AIS	Spine AIS	Arms AIS	Arms AIS
7	Mechanism group	Year of admission	Mechanism group	Motor GSC
8	Motor GSC	Month of admission	Referred from another hospital	Verbal GCS
9	Verbal GCS	Time of admission	Face AIS	Mechanism group
10	Referred from another hospital	Day of admission	Penetrating injury	Face AIS

from year to year). It seems far-fetched to suggest that these have a real impact on mortality, and in fact, when chi-squared tests were performed to examine the relationship between these factors and death, no significant association was found. So, it is known that these factors have some influence, but it is not clear what that influence is. Perhaps there are complex cross-correlations between these factors and others that can be modelled by the ANN but not explained. Discussions on how to deal with peculiarities like this are scarce in the ANN literature. One way of trying to examine these cross-correlations is to list the meaningful factors with which day, month, and year could be correlated. In this instance, there do not appear to be any. This goes back to one of the two big disadvantages of using ANNs: it is not possible to show why a neuron is weighted in a certain way after training. Without this, the usefulness of ANNs is severely impaired, and, in the case of trauma audit data, illustrates the limitations of an ANN as an analysis tool.

While ANNs may not be suitable alone for analysing trauma audit data, they may have a place alongside more traditional techniques. The results demonstrate the value of recording as much data about the trauma as possible and including it in an exploratory analysis. An analysis by ANN may be valuable for identifying factors that previously have not or normally would not be identified as having an impact on outcomes, such as time of day, which was identified in this study by an ANN but not by logistic regression. These factors then can be investigated by traditional techniques in order to examine their impact.

The implication of this is that the goal of using ANNs in analysing trauma data can only ever be one of prediction. The ANN cannot be used to analyse the individual factors that are important in making this prediction. If the goal is to analyse these factors, then logistic regression is more suitable. If an ANN is used for survival prediction, this will have implications for treatment (e.g., if carers believe a patient will die, their care changes from one of treatment to one of making the patient's last moments as comfortable as possible). Such systems already exist and are currently in use. For instance APACHE is used to predict an intensive care patient's risk of dying, although its advocates claim that it is used only to help carers consider the issue of whether their treatment is making a difference (http://www.

openclinical.org/aisp_apache.html, retrieved October 17, 2005) rather than deciding on what treatment to give.

Comparison with Logistic Regression

The ANN models appear to support the findings of Oakley et al. (2004) and the current logistic regression model about which factors have an impact on mortality, but they also demonstrate that outcomes are dependent on a complex interplay among a large number of factors and cannot be predicted easily or very accurately. While the ANN overall predictive accuracy is better than that of the logistic regression model, the logistic regression model is better at predicting death than all four ANN models; the sensitivity is 79% for the logistic model and ranges between 57% and 63% for the ANN models. Although the logistic regression model outperforms all the ANNs in sensitivity, it has a lower specificity (this is probably due to the cut-point being moved from the default of 0.5 to 0.2). The positive predictive value of the logistic regression model is slightly lower than the ANNs, showing that fewer than half of those predicted to die actually did die during their hospital stay. The results of analysing the trauma audit data show that the ANN is a relatively poor predictor of death and is of limited use in examining the determinants of death.

Concluding Remarks

The management of multiple-injured patients remains difficult with input from a number of surgical and medical specialities. Trauma scoring systems are a useful research tool, permitting comparison of multiple-injured patients in different centres around the world. Mortality remains a favoured outcome to study in trauma data mining, but others may be considered. Length of stay in hospital is one, although this really measures the process of care rather than anything specific about the injury. Disability as a result of the injury is another worthwhile outcome to study, but it is difficult to measure, especially when comparing different injury patterns. Future research should be directed at producing a scoring method to measure disability, similar to the AIS. Research to follow up patients after discharge from a hospital also would be useful in order to measure their outcome in terms of both continuing morbidity and quality of life. Trauma audit systems such as TARN take time to evolve; the data grow at a slower rate than some other medical databases, which means that improvements in patient care cannot be expected immediately. The University Hospital of North Staffordshire is starting to see a reduction in deaths; hopefully, studies such as the present one will help to continue this trend.

REFERENCES

Baker, S. P. et al. (1974). The injury severity score: A method for describing patients with multiple injuries and evaluating emergency care. *J Trauma*, *14*, 187-196. Retrieved January 18, 2005, from http://www.trauma.org/scores/iss.html

Baxt, W. G., Shofer, F. S., Sites, F. D., & Hollander, J. E. (2002). A neural computational aid to the diagnosis of acute myocardial infarction. *Annals of Emergency Medicine, 39*(4), 366-373.

Berardi, V. L., Patuwo, B. E., & Hu, M. Y. (2004). A principled approach for building and evaluating neural network classification models. *Decision Support Systems, 38*(2), 233-246.

Bose, I., & Mahapatra, R. K. (2001). Business data mining — A machine learning perspective. *Information & Management, 39*(3), 211-225.

Bratko, I., & Konomenko, I. (1987). Learning diagnostic rules from incomplete and noisy data. In B. Phelps (Ed.), *AI methods in statistics* (pp. 67-83). London: Gower Technical Press.

Breault, J. L., Goodall, C. R., & Fos, J. (2002). Data mining a diabetic data warehouse. *Artificial Intelligence in Medicine, 26*(1, 2), 37-54.

Chawla, N., Bowyer, K., Hall, L., & Kegelmeyer, W. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16,* 321-357.

Cios, K. J., & Moore, G. W. (2002). Uniqueness of medical data mining. *Artificial Intelligence in Medicine, 26*(1, 2), 1-24.

Cunningham, P., Carney, J., & Jacob, S. (2000). Stability problems with artificial neural networks and the ensemble solution. *Artificial Intelligence in Medicine, 20*(3), 217-225.

DiRusso, S., Sullivan, T., Holly, C., Nealon Cuff, S., & Savino, J. (2000). An artificial neural network as a model for prediction of survival in trauma patients: Validation for a regional trauma area. *The Journal of Trauma Injury, Infection and Care, 49*(2), 212-223.

Fu, X., Wang, L., Chua, K. S., & Chu, F. (2002). Training RBF neural networks on unbalanced data. In *Proceedings of the 9th International Conference on Neural Information Processing, Volume 2.* Retrieved from http://sci2s.ugr.es/keel/monografia/unbalanced/01198214.pdf

Giudici, P. (2003). *Applied data mining.* Chichester: Wiley.

Guo, H., & Viktor, H. L. (2004). Learning from imbalanced data sets with boosting and data generation: The DataBoost-IM approach. *ACM SIGKDD Explorations Newsletter Archive, 6*(1), 30-39.

Imberman, S. P., Damask, B., & Thompson, H. W. (2002). Using dependency/association rules to find indications for computed tomography in a trauma dataset. *Artificial Intelligence in Medicine, 26*(1, 2), 55-68.

Konomenko, I., & Kular, M. (1995). Machine learning for medical diagnosis. In *Proceedings of the CADAM-95* (pp. 9-30), Bled.

Kubat, M., & Matwin, S. (1997). Addressing the curse of imbalanced training sets: One sided selection. In *Proceedings of the 14th International Conference on Machine Learning,* San Francisco, CA (pp. 179-186).

Lapuerta, P., Azen, S. P., & LaBree, L. (1995). Use of neural networks in predicting the risk of coronary artery disease. *Computers and Biomedical Research, 2*8(1), 38-52.

Lavrac, N. (1999). Selected techniques for data mining in medicine. *Artificial Intelligence in Medicine, 16*(1), 3-23.

Lee, A., Hung Cheng, C., & Balakrishnan, J. (1998). Software development cost estimation: Integrating neural network with cluster analysis. *Information & Management, 34*(1), 1-9.

Lee, C. W., & Park, J. (2001). Assessment of HIV/AIDS-related health performance using an artificial neural network. *Information & Management, 38*(4), 231-238.

Lesmo, L., Saitta, L., & Torasso, P. (1982). Learning of fuzzy production rules for medical diagnosis. In M. Gupta, & E. Sanchez (Eds.), *Approximate reasoning in decision analysis* (pp. 249-260). Amsterdam: Elsevier

Ling, C., & Li, C. (1998). Data mining for direct marketing: Problems and solutions. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-98)* (pp. 73-79).

Muggleton, S. (1990). *Inductive acquisition of expert knowledge.* Wokingham: Addison-Wesley.

Nunez, M. (1990). Decision tree induction using domain knowledge. In B. Wielinga, J. Boose, B. Gaines, G. Schreiber, & M. van Someren (Eds.), *Current trends in knowledge acquisition* (pp. 276-288). Amsterdam: IOS Press.

Oakley, P. A., MacKenzie, G., Templeton, J., Cook, A. L., & Kirby, R. M. (2004). Longitudinal trends in trauma mortality and survival in Stoke-on-Trent 1992-1998. *Injury, 35*(4), 379-385.

Schapire, R. E. (2001, March 19-29). The boosting approach to machine learning: An overview. In *Proceedings of the MSRI Workshop on Nonlinear Estimation and Classification*, Berkeley, CA. Retrieved from http://citeseer.ist.psu.edu/schapire02boosting.html

Setiono, R., Thong, J. Y. L., & Yap, C. (1998). Symbolic rule extraction from neural networks: An application to identifying organizations adopting IT. *Information & Management*, *34*(2), 91-101.

TARN. (2004). *The first decade*. Retrieved from http://www.tarn.ac.uk/introduction/FirstDecade.pdf

Teasdale, G., & Jennett, B. (1974). *LANCET*, (ii), 81-83. Retrieved from http://www.trauma.org/scores/gcs.html

Templeton, J., Oakley, P. A., MacKenzie, G., Cook, A. L., Brand, D., Mullins, R. J., et al. (2000). A comparison of patient characteristics and survival in two trauma centres located in different countries. *Injury*, *31*(7), 493-501.

Watkins, D. (1997). *Clementine's neural networks technical overview.* Retrieved from http://www.cs.bris.ac.uk/~cgc/METAL/Consortium/secure/neural_overview.doc(Note that high AIS relates to an increase in severity, whereas high GCS relates to a decrease in severity.)

This work was previously published in the International Journal of Healthcare Information Systems and Informatics, Vol. 1, Issue 2, edited by J. Tan, pp. 51-64, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.27
Medical Document Clustering Using Ontology-Based Term Similarity Measures

Xiaodan Zhang
Drexel University, USA

Liping Jing
The University of Hong Kong, China

Xiaohua Hu
Drexel University, USA

Michael Ng
Hong Kong Baptist University, China

Jiali Xia
Jiangxi University of Finance and Economics, China

Xiaohua Zhou
Drexel University, USA

ABSTRACT

Recent research shows that ontology as background knowledge can improve document clustering quality with its concept hierarchy knowledge. Previous studies take term semantic similarity as an important measure to incorporate domain knowledge into clustering process such as clustering initialization and term re-weighting. However, not many studies have been focused on how different types of term similarity measures affect the clustering performance for a certain domain. In this article, we conduct a comparative study on how different term semantic similarity measures including path-based, information-content-based and feature-based similarity measure affect document clustering. Term re-weighting of document vector is an important method to integrate domain ontology to clustering process. In detail, the weight of a term is augmented by the weights of its co-occurred concepts. Spherical k-means are used for evaluate document vector re-weighting on two real-world datasets: Disease10 and OHSUMED23. Experimental results on nine different semantic measures have shown that: (1) there is no certain type of similarity measures that significantly outperforms the others; (2) Several similarity measures have rather more stable performance than the others; (3) term re-weighting has positive effects on medical document clustering, but might not be significant when documents are short of terms.

INTRODUCTION

Recent research has been focused on how to integrate domain ontology as background knowledge to document clustering process and shows that ontology can improve document clustering performance with its concept hierarchy knowledge (Hotho et. al., 2003; Jing et. al., 2006; Yoo et. al., 2006). Hotho, Staab and Stumme (2003) employed WordNet synsets to augment document vector and achieves better results than that of "bag of words" model on public domain. Yoo, Hu, and Song (2006) applied MeSH domain ontology to clustering initialization and achieved promising cluttering results. Terms are first clustered by calculating semantic similarity using MeSH ontology (http://www.nlm.nih.gov/mesh/) on PubMed document sets. Then the documents are mapped to the corresponding term cluster. Last, mutual reinforcement strategy is applied. Varelas et al. (2005) integrated domain ontology using term re-weighting for information retrieval application. Terms are assigned more weight if they are semantically similar with each other. Jing et al. (2006) adopted similar technique on document clustering.

Although existing approaches rely on term semantic similarity, not many studies have been done on evaluating the effects of different similarity measures on document clustering for a specific domain. Yoo, Hu, and Song (2006) employed one similarity measure that calculates the number of shared ancestor concepts and the number of co-occurred documents. Jing et al. (2006) compared two ontology-based term similarity measure. Even though these approaches heavily relied on term similarity information and all these similarity measures are domain independent, however, to date, relatively little work has been done on evaluating measures of term similarity for biomedical domain (where there are a growing number of ontologies that organize medical concepts into hierarchies such as MeSH ontology) on document clustering. In our pervious study (Zhang et al., 2007), a comparative study is conducted on a selected PubMed document set. However, the conclusion on one dataset may not be very general. Moreover, the similarity score threshold applied in previous study brings unfairness to term re-weighting since the distribution of similarity scores are different in terms of different similarity measure. Therefore, for a fair comparison, we use the minimum path length between two documents as the threshold.

Clustering initialization and term re-weighting are two techniques adopted for integrating domain knowledge. In this article, term re-weighting is chosen because: (1) a document is often full of class-independent "general" terms, how to discount the effect of general terms is a central task. Term re-weighting is more possible to help discount the effects of class-independent general terms and thus aggravates the effects of class-specific "core" terms; (2) hierarchically clustering terms (Yoo, Hu, & Song, 2006) for clustering initialization is more computational, expensive and more lack of scalability than that of term re-weighting approach.

As a result, we evaluate the effects of different term semantic similarity measures on document clustering using term re-weighting, an important measure for integration domain knowledge. We examine four path-based similarity measures, three information content- based similarity measures, and two feature-based similarity measures for document clustering on two biomedical literature sets: Disease10 and OHSUMED23. The rest of the article is organized as follows: the "Term Semantic Similarity Measures" section describes term semantic similarity measures; the "Methodology" section shows document representation and defines the term re-weighting scheme. The "Datasets" section lists two biomedical data sets. In the "Experimental Results And Analysis" section, we present and discuss experiment results. The last section briefly concludes the article.

TERM SEMANTIC SIMILARITY MEASURES

Ontology-based similarity measure has some advantages over other measures. First, ontology is manually created by human beings for a domain and thus more precise; second, compared to other methods such as latent semantic indexing, it is much more computational efficient; Third, it helps integrate domain knowledge into the data mining process. Comparing two terms in a document using ontology information usually exploits the fact that their corresponding concepts within ontology usually have properties in the form of attributes, level of generality or specificity, and their relationships with other concepts (Pedersen et al., 2007). It is worth noting that there are many other term semantic similarity measures such as latent semantic indexing, but this is out of scope of this study and our focus is on term semantic similarity measure using ontology information. In the subsequent subsections, we classify the ontology-based semantic measures into the following three categories.

Path-Based Similarity Measure

Path-based similarity measure usually utilizes the information of the shortest path between two concepts, of the generality or specificity of both concepts in ontology hierarchy, and of their relationships with other concepts.

Wu and Palmer (1994) developed a similarity measure finding the most specific common concept that subsumes both of the concepts being measured. The path length from most specific shared concept is scaled by the sum of IS-A links from it to the compared two concepts.

$$S_{W\&P}(C_1,C_2)=\frac{2H}{N_1+N_2+2H} \qquad (1)$$

In the Equation 1, N_1 and N_2 is the number of IS-A links from C_1, C_2 respectively to the most specific common concept C, and H is the number of IS-A links from C to the root of ontology. The similarity score is between 1(for similar concepts) and 0. In practice, we set H to 1 when the parent of the most specific common concept C is the root node.

Li et al. (2003) combines the shortest path and the depth of ontology information in a non-linear function:

$$S_{Li}(C_1,C_2)=e^{-\alpha L}\frac{e^{\beta H}-e^{-\beta H}}{e^{\beta H}+e^{-\beta H}} \qquad (2)$$

where L stands for the shortest path between two concepts, α and β are parameters scaling the contribution of shortest path length and depth, respectively. The value is between 1(for similar concepts) and 0. In our experiment, we set α and β to 0.2 and 0.6, respectively for the best performance.

Leacock and Chodorow (1994) defined a similarity measure based on the shortest path $d(C_1, C_2)$ between two concepts and scaling that value by twice the maximum depth of the hierarchy, and then taking the logarithm to smooth the resulting score:

$$S_{L\&C}(C_1,C_2)=-\log\left(d(C_1,C_2)/2D\right) \qquad (3)$$

where D is the maximum depth of the ontology and similarity value. In practice, we add 1 to both $d(C_1, C_2)$ and $2D$ to avoid log (0) when the shortest path length is 0.

Mao and Chu (2002) presented a similarity measure using both shortest path information and number of descendents of compared concepts.

$$S_{Mao}(C_1,C_2)=\frac{\delta}{d(C_1,C_2)\log_2(1+d(C_1)+d(C_2))} \qquad (4)$$

where $d(C_1, C_2)$ is the number of edges between C_1 and C_2, $d(C_1)$ is the number of C_1's descendants, which represents the generality of the concept. Here, the constant δ refers to a boundary case where C_1 is the only direct hypernym of C_2, C_2 is the only direct hyponym of C_1 and C_2 has no hyponym. In this case, because the concepts C_1 and C_2 are very close, δ should be chosen close to 1. In practice, we set it to 0.9.

Information Content-Based Measure

Information content-based measure associates probabilities with concepts in the ontology. The probability is defined in Equation 5, where *freq(C)* is the frequency of concept C, and *freq(Root)* is the frequency of root concept of the ontology (Pedersen et al., 2007). In this study, the frequency count assigned to a concept is the sum of the frequency counts of all the terms that map to the concept. Additionally, the frequency counts of every concept includes the frequency counts of subsumed concepts in an IS-A hierarchy.

$$IC(C) = -\log\left(\frac{freq(C)}{freq(Root)}\right) \quad (5)$$

As there may be multiple parents for each concept, two concepts can share parents by multiple paths. We may take the minimum $IC(C)$ when there is more than one shared parents, and then we call concept C the most informative subsumer— $IC_{mis}(C_1, C_2)$. In another word, $IC_{mis}(C_1, C_2)$ has the least probability among all shared subsumer between two concepts.

$$S_{Resnik}(C_1, C_2) = -\log IC_{mis}(C_1, C_2) \quad (6)$$

$$S_{Jiang}(C_1, C_2) = -\log IC(C_1) - \log IC(C_2) + 2\log IC_{mis}(C_1, C_2) \quad (7)$$

Resnik (1999) defined a similarity measure that signifies that the more information two terms share in common, the more similar they are, and the information shared by two terms is indicated by the information content of the term that subsume them in the ontology. The measure reveals information about the usage within corpus of the part of the ontology queried. Jiang and Conrath (1998) included not only the shared information content between two terms, but also the information content each term contains.

Lin utilized both the information needed to state the commonality of two terms and the information needed to fully describe these two terms. Since $IC_{mis}(C_1, C_2) >= \log IC(C_1), \log IC(C_2)$ the similarity value varies between 1 (for similar concepts) and 0.

$$S_{Lin}(C_1, C_2) = \frac{2\log IC_{mis}(C_1, C_2)}{\log IC(C_1) + \log IC(C_2)} \quad (8)$$

Feature-Based Measure

Feature-based measure assumes that each term is described by a set of terms indicating its properties or features. Then, the more common characteristics two terms have and the less non-common characteristics they have, the more similar the terms are (Varelas et al., 2005). As there is no describing feature set for MeSH descriptor concepts, in our experimental study, we take all the ancestor nodes of each compared concept as their feature sets. The following measure is defined based on the discussion in Knappe et al. (2006) and Lin (1003):

$$S_{BasicFeature}(C_1, C_2) = \frac{|Ans(C_1) \cap Ans(C_2)|}{|Ans(C_1) \cup Ans(C_2)|} \quad (9)$$

where $Ans(C_1)$ and $Ans(C_2)$ correspond to description sets (the ancestor nodes) of terms C_1 and C_2 respectively, $C_1 \cap C_2$ is the join of two parent node sets and $C_1 \cup C_2$ is the union of two parent node sets.

Knappe et al. (2006) developed a similarity measure, as seen below, using the information of generalization and specification of two compared concepts:

$$S_{Knappe}(C_1,C_2) = p \times \frac{|Ans(C_1) \cap Ans(C_2)|}{|Ans(C_1)|} + (1-p) \times \frac{|Ans(C_1) \cap Ans(C_2)|}{|Ans(C_2)|} \quad (10)$$

where p's range is [0, 1] that defines the relative importance of generalization versus specialization. This measure falls between 1 (for similar concepts) and 0. In our experiment, p is set to 0.5.

METHODOLOGY

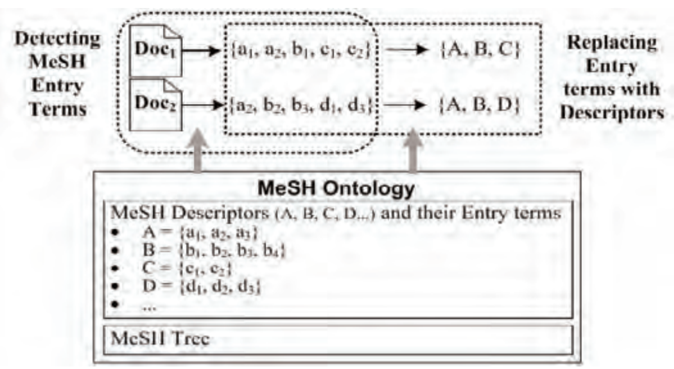
Given a document set, our clustering method is composed of the following steps: (1) apply ontology to index whole document set; each document is thus represented as a vector of terms; (2) each term's weight is re-calculated by the proposed term re-weighting method; (3) Spherical K-means is run the on the dataset.

MeSH Ontology

Ontology is very important to biomedical documents clustering. First, biomedical literature is usually composed of many complicated biomedical concepts with name variations containing usually more than one word. Second, bag-of-words model suffers from "the curse of dimension" and lacks interpretation power to clustering results. Therefore, MeSH ontology is applied to index biomedical literature in this article.

Medical Subject Headings (MeSH) [www.nlm.nih.gov./mesh] mainly consists of the controlled vocabulary and a MeSH Tree. The controlled vocabulary contains several different types of terms, such as Descriptor, Qualifiers, Publication Types, Geographics, and Entry terms. Descriptor terms are main concepts or main headings. Entry terms are the synonyms or the related terms to descriptors. For example, "Neoplasms" as a descriptor has the following entry terms {"Cancer," "Cancers," "Neoplasm," "Tumors", "Tumor", "Benign Neoplasm," "Neoplasm, Benign"}. As a result, descriptors terms are used in this research. MeSH descriptors are organized in a MeSH Tree, which can be seen as the *MeSH Concept Hierarchy*. In the MeSH Tree, there are 15 categories (e.g., category A for anatomic terms), and each category is further divided into subcategories. For each subcategory, corresponding descriptors are hierarchically arranged from most general to most specific. In addition to its ontology role, MeSH descriptors have been used to index PubMed articles. For this purpose, about 10 to 20 MeSH terms are manually assigned to each article (after reading full papers). On the assignment of MeSH terms to articles, about three to five MeSH terms are set as "MajorTopic" that primarily represent the article. This indicates that submitting Major

Figure 1. PubMed document indexing



Topic MeSH term query to PubMeD usually retrieves dataset with ground truth.

MeSH Concept Indexing

Terms in each document are mapped to the entry terms in MeSH and then maps the selected Entry terms into MeSH Descriptors to remove synonyms.

In detail, our indexing system matches the terms in each document to the entry terms in MeSH and then maps the selected entry terms into MeSH Descriptors. Instead of searching all entry terms in the MeSH against each document, we select 1- to 5-gram words as the candidates of MeSH Entry terms. Then, only those candidate terms are chosen that match with MeSH entry terms. We then replace those semantically similar entry terms with the Descriptor term to remove synonyms. Next, some MeSH Descriptors are filtered out that are too common or have nothing do with the contents of PubMed articles (e.g., ENGLISH ABSTRACT or Government supported); A stop term list is generated for this purpose by analyzing 10 years of PubMed documents (1995-2004). At the time of this writing, there are about 23,833 unique MeSH descriptor terms, 44,978 MeSH ontology nodes (one descriptor term might belong to more than one ontology nodes) and 593,626 MeSH entry terms.

MeSH entry term sets are detected from documents "Doc_1" and "Doc_2" using the MeSH ontology, and then the Entry terms are replaced with Descriptors based on the MeSH ontology. Both MeSH descriptors and entry terms are multi-grams.

Term Re-Weighting

A document is often full of class-independent "general" words and short of class-specific "core" words, which leads to the difficulty of document clustering. Steinbach et al. (2000) examined on the data that each class has a "core" vocabulary of words and remaining "general" words may have similar distributions on different classes. To solve this problem, we should "discount" general words and "emphasize" more importance on core words in a vector. Jiang and Conrath (1998) and Varelas et al. (2005) define the term re-weighting scheme as below:

$$\tilde{x}_{ji1} = x_{ji1} + \sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^{m} S\left(x_{ji1}, x_{ji2}\right) \cdot x_{ji2} \qquad (11)$$

where x stands for term weight, m stands for the number of co-occurred terms, and $S(x_{ji1}, x_{ji2})$ stands for the similarity score between two concepts. Through this re-weighting scheme, the weights of semantically similar terms will be co-augmented. Since we are only interested in re-weighting those terms that are more semantically similar with each other, it is necessary to set up a threshold value—the minimum similarity score or the minimum path length between compared concepts. In practice, document is first represented according to certain scheme such as TF-IDF. Then, each term's weight is augmented by Equation 11.

Table 1. The document sets and their sizes

	Document Sets	No. of Docs
1	Gout	642
2	Chickenpox	1,083
3	Raynaud Disease	1,153
4	Jaundice	1,486
5	Hepatitis B	1,815
6	Hay Fever	2,632
7	Kidney Calculi	3,071
8	Age-related Macular Degeneration	3,277
9	Migraine	4,174
10	Otitis	5,233

DATASETS

Disease10

Disease10 dataset is collected from PubMed (a Web interface of Medline documents) by submitting queries using "MajorTopic" tag along with the corresponding MeSH term of the disease name. For example, if the disease name's corresponding MeSH term is "Gout," then the query will become "Gout [Major Topic]." Table 1 shows the ten classes of document sets and their document numbers (24,566 documents). The document class name is the query name. The average document length for MeSH descriptor is 13 (as shown in Table 2). Compared to the average document length—81—when using bag of words representation, the dimension of clustering space is dramatically reduced. A general stop term list is applied to bag of words scheme.

OHSUMED23

OHSUMED consists of scientific abstracts collected from Medline, an online medical information database. The selected OHSUMED corpus contains 13,929 Medline abstract of the year 1991, each of which was assigned with one or multiple labels out of 23 cardiovascular diseases categories. Excluding abstracts with multiple labels, we indexed the rest 7,400 abstracts belonging to 23 classes.

EXPERIMENTAL RESULTS AND ANALYSIS

Evaluation Methodology

Cluster quality is evaluated by four extrinsic measures, *entropy* (Steinbach, Karypis & Kumar, 2000), *F-measure* (Larsen & Aone, 1999), *purity* (Zhao & Karypis, 2001), and *normalized mutual information (NMI)* (Banerjee & Ghosh, 2002). NMI is defined as the mutual information between the cluster assignments and a pre-existing labeling of the dataset normalized by the arithmetic mean of the maximum possible entropies of the empirical marginal, that is:

$$NMI(X,Y) = \frac{I(X;Y)}{(\log k + \log c)/2} \quad (12)$$

where X is a random variable for cluster assignments, Y is a random variable for the pre-existing labels on the same data, k is the number of clusters, and c is the number of pre-existing classes. NMI ranges from 0 to 1. The bigger the NMI is the higher quality the clustering is. NMI is better than other common extrinsic measures such as purity and entropy in the sense that it does not necessarily increase when the number of clusters increases. Purity can be interpreted as the classification rate under the assumption that all samples of a cluster are predicted to be members of the actual dominant class for that cluster. Entropy is a more comprehensive measure than purity since rather than just considering the number of objects "in" and "not in" the most frequent class, it considers the entire distribution. F-score combines the

Table 2. Document indexing schemes

Indexing Scheme	No. of term indexed	Avg. doc length
MeSH descriptor term	8829	13
Word	41,208	81

information of precision and recall which is the extensively applied in information retrieval, with values falling in [0, 1] and the larger is the value, the better is the cluster quality.

Experiment Settings

To improve the efficiency of the calculation of term-term similarity, a 44,978 term-term similarity matrix (including all MeSH descriptors) is constructed for each similarity measure before the document vector re-weighting.

The similarity score is disregarded between two terms of which minimal path length is larger than 3, since we are only interested in augmenting the weights of terms that are similar enough. This is better than setting a similarity score threshold and very important to evaluate different semantic similarity measures in a fair manner. The distributions of the similarity scores between documents are usually various in terms of different similarity measures. Setting one threshold to all similarity measures can make the results easily biased toward several measures and need time consuming tuning (Zhang et al., 2007). Therefore, we apply minimum length threshold instead of similarity score threshold. The minimum path length is defined as:

$$MinLen(C_1, C_2) = Dep(C_1) + Dep(C_2) - 2 \cdot Dep(C_1, C_2) \qquad (13)$$

where $Dep(C_1)$ indicates the depth of concept C_1 within the ontology and $Dep(C_1, C_2)$ is the depth of the nearest co-parent of concept C_1, C_2.

Apparently, the similarity score range of $S_{L\&C}$, S_{Resnik} and S_{Jiang} is not within [0, 1]. For a fair comparison, their similarity matrices are normalized before they are applied to re-weighting document vector. In detail, each similarity score is denominated by the row sum. In this study, each document is represented as TF-IDF vector since this scheme achieves much better performance than normalized term frequency and term frequency (Zhang, Zhou, & Hu, 2006). Each document vector is re-weighted using equation the equation by Pedersen, Pakhomov, Patwardhan, and Chute (2007) and the ontology term-term similarity matrix. Spherical K-means is used for documents clustering, for it is a well-known vector-based clustering algorithm. Documents are also indexed using unigram words for a more comprehensive comparison. Documents are not considered in our experiments if they contain fewer than five terms. The whole process is implemented using dragon toolkit (Zhou, Zhang, & Hu, 2006).

Table 3. Clustering results of Disease10

Type of Measure	Measure Name	Entropy	F-Score	Purity	NMI
Path-based	Wu & Palmer	0.348	0.858	0.874	0.779
	Li et al.	0.304	0.834	0.901	0.799
	Leacock	0.276	0.853	**0.923**	**0.811**
	Mao et al.	0.342	0.830	0.875	0.782
Information-Content-based	Resink	0.295	0.856	0.906	0.802
	Jiang	0.300	0.848	0.905	0.800
	Lin	0.342	0.845	0.882	0.782
Feature-based	Basic Feature	0.358	0.818	0.872	0.775
	Knappe	0.350	0.834	0.876	0.778
MeSH descriptor		0.341	0.772	0.867	0.776
Word		0.245	0.755	0.908	0.820

Result Analysis

Table 3 and 4 show the experimental results of document clustering on Disease10 and OHSUMED23 datasets, respectively. The nine ontology-based similarity measures are divided by their corresponding types including: path-based, information-content-based and feature-based. "MeSH descriptor" and "Word" indicate the type of document representation and they do not use term re-weighting scheme.

Comparison Between"Re-Weighting" and "None Re-Weighting"

The performance between re-weighting and none re-weighting varies in terms of the corresponding datasets. For Disease10 dataset, most similarity measures slightly outperform none re-weighting, that is, MeSH descriptor. For OHSUMED23 dataset, the results of different schemes are very close. Three measures including Li, Zuhair and McLean (2003), Leacock and Chodorow (1994) and Resnik (1999) have slightly better performances than none re-weighting scheme. These results show that the re-weighting scheme can slightly improve document clustering, but it is not very significant. They also show that term re-weighting as a method of integrating domain ontology to clustering might not be a very effective approach when the documents are short of terms (Table 2), because when all these terms are very important core terms for the documents, ignoring the effects of some of them by re-weighting can cause serious information loss. This is on the contrary to the experiment results (Jing et al., 2006) in general domain where document length is relatively longer.

Comparison Between Different Similarity Measures

Experimental results on two datasets show that, among the three types of term similarity measures, there is no certain type of measure that significantly outperforms others. Interestingly, information-content-based measures, with the support of corpus statistics, have very similar performances with the other two types of measure. This may indicate that the corpus statistics is fit with ontology structure of MeSH and thus does not have better performance than path-based measures. Two path-based measures including Leacock and Chodorow (1994) and Li, Zuhair, and

Table 4. Clustering results of OHSUMED23

Type of Measure	Measure Name	Entropy	F-Score	Purity	NMI
Path-based	Wu & Palmer	2.209	0.244	0.347	0.165
	Li et al.	**2.181**	**0.253**	**0.356**	**0.174**
	Leacock	2.199	0.241	0.351	0.168
	Mao et al.	2.183	0.255	0.354	0.173
Information- Content-based	Resnik	2.194	0.252	0.352	0.170
	Jiang	2.199	0.251	0.351	0.168
	Lin	2.234	0.239	0.341	0.158
Feature-based	Basic Feature	2.219	0.241	0.344	0.162
	Knappe	2.226	0.239	0.340	0.160
MeSH descriptor		2.193	0.248	0.353	0.170
Word		2.321	0.200	0.302	0.130

McLean (2003) achieve the best performance on both datasets, respectively. Both measures consider not only the shortest path and depth of two concepts. Judging from the overall performance on the two datasets, Li, Zuhair, and McLean (2003), Leacock and Chodorow (1994), Mao and Chu (2002), Resink (1999) and Jiang and Conrath (1998) have rather more stable performances than that of the other measures. Feature-based measures always have the worst performance. This shows that using parent concepts as features may have negative impact on term re-weighting.

Comparison Between Ontology-Based and Word-Based Document Representation

The performance of word scheme is significantly different on the two datasets. For Disease10 dataset, word scheme is slightly better than ontology-based scheme, but this is not significant. On OHSUMED23 dataset, word scheme performs significantly worse than the other schemes. The results show both advantage of ontology and the limitation of ontology. First, while keeping competitive or significantly better clustering results, not only the dimension of clustering space but also the computational cost are dramatically reduced especially when handling large datasets. Second, existing ontologies are under growing, they are still not enough for many text mining applications. For example, there are only about 44,000 unique MeSH descriptor terms for the time of writing. Third, there is also limitation of term extraction. So far, existing approaches usually use "exact match" to map abstract terms to entry terms. This will cause serious information loss. For example, when representing document as MeSH descriptor terms, the average document length is only 13 for Disease10, while the length of the corresponding word representation is 81. Finally, if taking advantage of both medical concept representation and informative word representation, the results of text mining application can be more convincing.

CONCLUSION AND FUTURE WORK

In this article, we evaluate the effects of nine semantic similarity measures with a term re-weighting method on document clustering of PubMed document sets. The spherical k-means clustering experiment shows that term re-weighting has some positive effects on medical document clustering, but might not be very significant. In detail, we obtain following meaningful findings by comparing nine semantic similarity measures three types: path-based, information-content-based and feature-based measure with two indexing schemes—MeSH descriptor and Word: (1) term re-weighting achieves very similar clustering results with none term re-weighting. Some of them outperform none re-weighting, some of them do not and neither of them is very significant, which indicates that term re-weighting can be effective in a very limited degree when documents are short of terms because when most of these terms are distinguishable core terms for a document, ignoring some of them by re-weighting will cause information loss; more developed ontology and advanced term extraction technique may help term re-weighting achieve better results; (2) There is no certain type of measure that is significantly better than others; the best performance are achieved by two path-based measures including Leacock and Chodorow (1994) and Li, Zuhair, and McLean (2003) that consider both the closeness and the depth of the compared concepts; feature-based measures have the worst overall performance, which shows that using parent concepts as feature set is not effective for this application; although information- content-based measures consider both ontology and corpus statistics, they do not achieve better results than the other measure types; (3) the performance of MeSH scheme is much better than that of word scheme on OHSUMED23 dataset and slightly worse than word scheme on Disease10 dataset, which demonstrates both the advantage and limitation of domain ontology; while keeping competitive or significantly better results, indexing using MeSH ontology dramatically reduces the dimension of clustering space and computational complexity; however, the limitation of ontology such as limited concepts and rough term extraction techniques can cause information loss easily and thus hurt the clustering performance. Furthermore, this finding indicates that there should be an approach taking advantage of both medical concept representation and informative word representation.

In our future work, we may consider other biomedical ontology such as Medical Language System (UMLS) and also expand this comparative study to some public domain.

ACKNOWLEDGMENT

This work is supported in part by NSF Career grant (NSF IIS 0448023), NSF CCF 0514679, PA Dept of Health Tobacco Settlement Formula Grant (No. 240205 and No. 240196), and PA Dept of Health Grant (No. 239667).

REFERENCES

Banerjee A., & Ghosh, J. (2002). Frequency sensitive competitive learning for clustering on high-dimensional hyperspheres. In *Proceedings of IEEE International Joint Conference on Neural Networks*, pp. 1590-1595.

Hotho, A., Staab, S., & Stumme, G. (2003). Wordnet improves text document clustering. In *Proceedings of the Semantic Web Workshop at 26th Annual International ACM SIGIR Conference*, Toronto, Canada.

Jiang, J., & Conrath, D. (1998). Semantic similarity based on corpus statistics and lexical taxonomy. In *Proceedings of the International Conference on Research in Computational Linguistic*, Taiwan.

Jing, J., Zhou, L., Ng, M., & Huang, Z. (2006). Ontology-based distance measure for text clustering. In *Proceedings of SIAM SDM workshop on text mining.* Bethesda, Maryland, USA.

Knappe, R., Bulskov, H., & Andreasen, T. (2006). Perspectives on ontology-based querying. *International Journal of Intelligent Systems.*

Larsen, B., & Aone, C. (1999). Fast and effective text mining using linear-time document clustering. *KDD-99* (pp. 16-22). San Diego, California.

Leacock, C., & Chodorow, M. (1994). Filling in a sparse training space for word sense identification. ms.

Li, Y., Zuhair, A., & McLean, D. (2003). An approach for measuring Semantic similarity between words using multiple information sources. *IEEE Transactions on Knowledge and Data Engineering, 15*(4), 871-882.

Lin, D. (1993). Principle-based parsing without over-generation. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics* (pp. 112-120). Columbus, Ohio.

Mao, W., & Chu, W. (2002). Free text medical document retrieval via phrased-based vector space model. In *Proceedings of AMIA'02*, San Antonio, TX.

Pedersen, T., Pakhomov, S., Patwardhan, S., & Chute, C. (2007). Measures of semantic similarity and relatedness in the biomedical domain. *Journal of Biomedical Informatics, 40*(3), 288-299.

Resnik, O. (1999). Semantic Similarity in Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity and Natural Language. *Journal of Artificial Intelligence Research, 11*, 95-130.

Steinbach, M., Karypis, G., & Kumar, V. (2000). *A comparison of document clustering techniques.*

Technical Report #00-034, Department of Computer Science and Engineering, University of Minnesota.

Varelas, G., Voutsakis, E., Raftopoulou, P., Petrakis, E., & Milios, E. (2005). Semantic similarity methods in WordNet and their application to information retrieval on the Web. *WIDM '05* (pp. 10-16). New York: ACM Press.

Wu, Z., & Palmer, M. (1994). Verb Semantics and lexical selection. In *Proceedings of the 32nd Annual Meeting of the Associations for Computational Linguistics* (pp. 133-138), Las Cruces, New Mexico.

Yoo I., Hu X., & Song I-Y. (2006). Integration of Semantic-based bipartite graph representation and mutual refinement strategy for biomedical literature clustering. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 791-796.

Zhao, Y., & Karypis, G (2001). *Criterion functions for document clustering: Experiments and analysi*s. Technical Report, Department of Computer Science, University of Minnesota.

Zhang X., Zhou X., & Hu X. (2006). Semantic smoothing for model-based document clustering. *In Proceedings of the sixth IEEE International Conference on Data Mining.*

Zhang X., Jing L., Hu X., Ng M., & Zhou X. (2007). A comparative study of ontology based term similarity measures on document clustering. Accepted in the *12th International conference on Database Systems for Advanced Applications (DASFFA2007).*

Zhou, X., Zhang, X., & Hu, X. *The Dragon Toolkit*, *Data Mining & Bioinformatics Lab.* iSchool at Drexel University. Retrieved from http://www.dragontoolkit.org/

This work was previously published in the International Journal of Data Warehousing and Mining, Vol. 4, Issue 1, edited by D. Taniar, pp. 62-73, copyright 2008 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.28
Ontology-Based Spelling Correction for Searching Medical Information

Jane Moon
Monash University, Australia

Frada Burstein
Monash University, Australia

ABSTRACT

There has been a paradigm shift in medical practice. More and more consumers are using the Internet as a source for medical information even before seeing a doctor. The well known fact is that medical terms are often hard to spell. Despite advances in technology, the Internet is still producing futile searches when the search terms are misspelled. Often consumers are frustrated with irrelevant information they retrieve as a result of misspelling. An ontology-based search is one way of assisting users in correcting their spelling errors when searching for medical information. This chapter reviews the types of spelling errors that adults make and identifies current technology available to overcome the problem.

INTRODUCTION

Medical terminology includes many words which nonprofessionals find difficult to spell. For novice computer users it can be extremely frustrating when, as a result of misspelling, they cannot find relevant information. The medical implications of errors arising as a result of misspelling a word is well documented in the literature (Lambert, 1997; Lilley, 1995). Reports show that there is a significant error rate observed in patient records—in particular, in discharge letters (Blaser et al., 2004). An intelligent system could provide the users with some suggestions in order to assist in using correct words when dealing with medical information. However; it seems we are still far from reaching this goal.

A word processing program such as Microsoft Word, auto corrects typos and provides advice on whether a word is correctly spelt through a process of "verification and offers users to spell correctly via a process of 'correction'" (Berghel & Andreu, 1998). However, word correction is based on text spelling alone and can cause more problems at the correction stage. For example, Microsoft Word does not distinguish homophones such as *heroin* or *heroine* (Jones & Martin, 1997). The approach it takes in correcting spelling is not context based (Fallman, 2002) or ontology based (Khan, McLeod, & Hovy, 2004; Patel, Supekar, Lee, & Park, 2003; Tijerino & Reza, 2005), it does not recognize blank space which wastes a lot of space (Nix, 1981) and does not exhibit much intelligence. Popular search engines such as Yahoo and Google do provide some spelling correction by prompting a list of words based on keywords. In most cases search method is based on simple word searches and frequency distributions and these do not capture the meaning behind the words. Mauldin (1991) calls this phenomenon *keyword barrier*. Breaking through the keyword barrier will require a system that understands the texts they process.

There have been many attempts to produce spelling correction programs. They focus on various approaches including word concatenation (Berghel & Andreu, 1998), spelling correctors such as *Talisman* (Berghel, 1998), and correcting misspellings that contain a single error form (Pollock & Zamora, 1984). Some of these systems are domain specific and operate in chemistry, such as ChemSpell (Mitton, 1996; Zamora, 1978, 1980).

With the recent explosion of portals and ever expanding health information searching via medical/health portals, the question of the effectiveness of these health portals remains to be answered. Despite major advances and efforts to make portals user friendly, the research shows (Moon, 2005; Moon & Burstein, 2005) that most portals still lack essential functions for assisting users with information retrieval and hence losing potential users from existing portals. Even though the medical portals analyzed had the answers to users' queries, the major problem lay in the portal's inability to handle misspelling, thereby lacking an essential provision for assisting users (Moon & Fisher, 2005).

Another issue is that some major search engines provide a list of words that are not, however, always relevant to the user's query. This is demonstrated in Figures 2-4. Recent advances in ontology-based search is suggested to be overcoming the shortcomings of keyword search (Khan et al., 2004). At least it eliminates lists of words that are not domain specific. It is highly domain specific and context based, thereby eliminating unnecessary retrieval of information. It is an efficient, time-saving method.

In this research we explored the basic patterns of adult spelling errors and researched the spelling error corrector to see what technique has been used. In addition, we present an evaluation of some existing medical spell correctors to test if they meet users' needs. We describe an ontology-based, spell checker architecture that would assist users with medical information retrieval.

Review of Spell Check Approaches

The literature on misspelling is diverse and much of it is based on studies of children. However, very little has been studied on adults. Review of papers show that spelling mistakes in adults are predictable (Yannakoudakis & Fawthrop, 1983). This section describes the types of spelling errors and the existing approaches to their correction.

General Pattern of Spelling Errors

A natural language such as English is not static. The English language is one of the most difficult languages to learn as it has the following properties:

- It is not a phonetic language; there are many words that are not spelled the way they sound.

- It has many borrowed words that are not of English origin.
- It has many suffixes and prefixes that serve the same or similar purposes.

It is natural to make spelling errors in a language as complicated as English. General spelling errors can occur for many reasons, two of which are most common, typographical errors and orthographic errors. These can be described as follows (Peterson, 1980).

Typographic error. The first one occurs when the user knows the correct spelling but enters wrong typing because of fatigue, memory lapse, distraction, carelessness, or inadvertent mistake. This type of error is not consistent but is predictable since it relates to the position of keys on the keyboard and is probably related to finger movements.

There are many different reasons for making typographic spelling errors. One can make mistakes due to keyboard dualities. For example, one may use a keyboard for more than one language where key settings vary; for example, the keyboard can be used for English as well as Spanish and could lead to some sort of confusion. Another reason can be forgetting to use shift key characters or forgetting to undo the capital locks. Phonetic similarities such as "ka" and "ga," which are similar in Japanese could lead to confusion. Visual similarity could also lead to mistakes. For instance if two letters are similar then one may mistype the character without realizing it.

Cognitive/Orthographic errors. This relates to the writer's knowledge, and to the difference between how a word sounds and how it should be spelled. The words with different forms, for example, verbs or nouns can be confusing depending on what context they are used, for example, *effect* and *affect.*

Consistent human spelling error is related to a systematic break of rule rather than exception. Examination of performance errors in spelling seems to indicate the spelling errors are random, but closer study of individual performers invariably reveals a pattern. Across individual spelling, errors are often idiosyncratic, but often there is a substantial degree of intra-individual consistency in the types of errors being made. Most of the spelling errors occur because of some type of systematic errors due to "logical" though incorrect reasoning.

Most spelling errors can be generated from correct spelling by a few simple rules. Damerau (1964) indicates that 80% of all spelling errors are the result of:

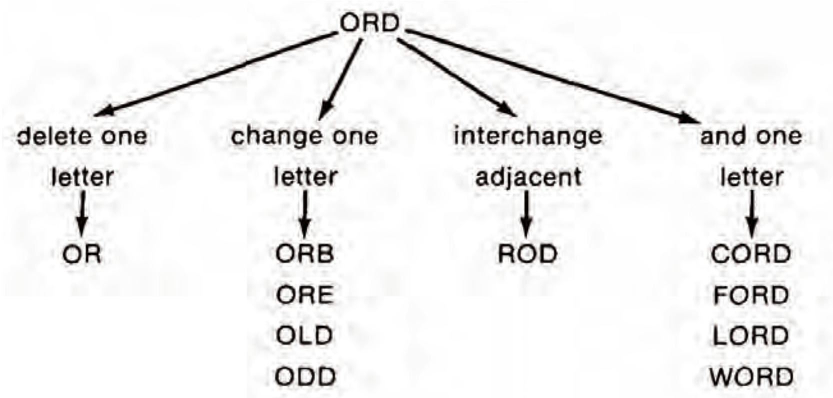
- transposition of two letters,
- insertion of a letter,
- missing a letter, or
- one wrong letter.

Yannakoudakis and Fawthrop (1983) confirmed that there are patterns in spelling errors; and continued Damerau's (1964) study and found that:

- many errors are made in the use of vowels;
- many errors are made in the use of the letters "H," "W," and "Y;"
- doubling and singling of any letters is common;
- few errors are made in the first character of a word;
- transposition of any two adjacent characters is common;
- missing or adding a character is common; and
- certain consonants are more frequently interchanged than others.

Based on Damerau's (1964) simple method, a SPELL check program can be written as in Figure 1.

Figure 1. Show Damerau's simple "arithmetic at" system based on deletion, insertion, transposition, and addition using ORD as example (Damerau, 1964)



Spelling Correctors

There are two types of spelling programs: *spell checkers* and *spell correctors.* Spell checkers simply look at the entered words and check whether they are correctly spelled. Spell correctors detect misspelled words and try to correct them. A spell corrector uses elements of pattern matching, appropriate algorithms, and coding theory.

Much work has been done on spelling correction algorithms, some based on phonetics, others are algorithms, clustering, and measuring distances which will be described in the following paragraphs.

Error Detection

The two most widely used error detection methods are n-gram analysis (Morris & Cherry, 1975; Zamora, Pollock, & Zamora, 1981) and dictionary look up (Aho and Corasick, 1975; Knuth, 1973; Peterson, 1980). These two approaches can be summarized as follows:

Trigram analysis is a method used to correct spelling by analyzing parts of words in three-letter sequences. It identifies three possible candidates for representing meaningful words. N-gram is used to perform trigram analysis:

- An *n-gram* is an n-letter sequence, extracted from a word, where *n* is usually 1, 2, or 3. In general, n-gram analysis techniques check each n-gram in an input string against a precompiled table of n-gram statistics to determine whether the n-gram can occur in a word. If it does, its frequency of occurrence in the words of the language is computed. Strings containing n-grams that do not occur in words or occur very infrequently are considered to be possible misspellings. For instance, the word HEPATITIS contains the trigrams: -HE, HEP, EPA, PAT, ATI, TIT, TIS, IS- where hyphen (-) represents the word boundary (Yannakoudakis & Fawthrop, 1983).
- *Dictionary lookup techniques* check whether an input string appears in a dictionary. It can be regarded as a special case of n-gram analysis where variable length n-grams delimited by blanks or punctuations are used (Zamora et al., 1981). Response time may become a problem as the size of the dictionary grows. To save the search time, most often the suffix is stripped before analysis. This could also present a problem, as meaning could sometimes be changed or lost without a suffix.

Error Correction

There has been much advance in the last decade on spelling error correction. Calculating the distance between strings has been a widely used method for error correction with a degree of success. Some of the error correction methods based on *distance between strings* are the *minimum edit distance*, *Hamming distance*, and *Levenshtein distance*. Another method of error correction that are based on learning from user's errors and improving and adapting from user's mistakes is called the *neural net technique*. The following paragraph describes approaches to error correction in detail.

Calculating distance between strings has been a commonly used option for modern spell checkers and seems to work well at least in English. In English spell correction word boundary problems such as in the word *forgot* (forgot – for got), run-ons such as in *form – inform*, or short words such as prepositions (to, or, at) are difficult. N-gram is used to correct context-independent problems. However, Asian languages such as Japanese and Chinese consist of run-on words, which are context dependent.

The latest word processor programs are capable of suggesting a replacement for a mistyped word. Spell checkers "know" how to evaluate distance between a misspelled word and the words in its files. Words whose evaluated distance is the smallest are suggested as candidates for replacement.

The minimum edit distance technique was first developed by Damerau (1964), to compute the minimum number of editing operations (such as insertion, deletion, transposition, and substitution) that are required to fix misspelled strings.

Hamming distance (H) is defined only for strings of the same length. Hamming distance is named after Richard Hamming; it is an algorithm that measures the number of *substitutions* required to change one string to the other. For strings s and t, H(s, t) is the number of places in which the two strings differ, have different characters.

For example (http://en.wikipedia.org/wiki/Hamming_distance):

- The Hamming distance between **1011101** and **1001001** is 2.
- The Hamming distance between **2143896** and **2233796** is 3.
- The Hamming distance between "***toned***" and "**roses**" is 3.

Hamming distance is used in telecommunications, signal distance, and cryptography. For more sophisticated systems where addition, subtraction are required, Levenshtein distance is more appropriate.

Levenshtein distance (or edit distance) is more sophisticated. It is defined for strings of arbitrary length. It counts the differences between two strings, where we would count a difference not only when strings have different characters but when one has a character and the other does not. It is the smallest number of insertions, deletions, and substitutions required to change one string or tree into another. An H(m x n) algorithm computes the distance between strings, where m and n are the lengths of the strings (http://www.nist.gov/dads/HTML/levenshtein.html).

The *simple correlation matrix technique* is a correlation technique. Each misspelled word is represented by an n-dimensional feature where Hamming distance of strongly correlated matches the most probably correct word (Cherkassky, Vassilas, Brodt, Wagner, & Fischer, 1974).

The *singular value decomposition* (SVD): Correlation Matrix Technique to apply matrix transformation techniques to simple correlation matrices in an effort to improve spelling correction accuracy (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990). The goal of SVD is to find the most relevant similarity in lexical space.

The correction of words rests on the basis of three common phenomena: nonword error detection, isolated word error correction, and context-dependent word correction (Kukich, 1992).

The *neural net technique* has the potential to adapt to specific error patterns of their user community, thus maximizing their correction accuracy for that population. This can have a user—adaptable chip that continuously monitors and learns from specific users or groups of users to improve spelling errors (Kukich, 1988).

Kukich did experiments to compare the effectiveness of spelling of the available three techniques to see the effects for different size word lexicons.

As can be seen from Table 1, slight improvements were noted with the neural network of approximately 15%. There was very little difference in performance regardless of numbers of words in the dictionaries.

In the past two decades there has been a lot of exploration of new techniques in word spelling errors (Nagata, 1996; Jones & Martin, 1997; Yannakoudakis & Fawthrop, 1983). There have also been many spelling corrector programs produced based on various algorithms. Table 2 lists some of such products and gives an overview of the strategies and characteristics of those products.

Impact of Spelling Errors on Medical Information Seeking

Medical terms are usually very hard to spell. Typing incorrectly or spelling phonetically can yield to inappropriate or irrelevant results, which causes a waste of time, resources or, at times, can give a false alarm if the users are not aware of the error. Blaster et al. (2004) found that spelling errors caused significant errors in patient discharge letters. Through earlier usability testing on Australian medical portals we found that the users did not want to come back to the sites that did not have spelling error correction facilities to assist their search (Moon & Burstein, 2005). In many instances the users could not find the information they were looking for from such portals. We have concluded then that spelling correction facility is one of the essential characteristics of an intelligent portal.

Most of the spelling correction techniques mentioned previously are developed for general English language. However, most of these techniques can be applied in a medical context too. In this chapter we investigate how a generic technique, such as Damerau (1964) method, can be suitably applied to medical settings to improve the system, to make a system more “intelligent” in assisting users in seeking information. Before looking at the generic architecture we analyze the functionality of two popular existing commercial medical spell checkers. The problem with using commercial products is high cost and lack of flexibility, which can be achieved with customized development.

ANALYSIS OF MEDICAL SPELL CHECKERS

We studied medical spell checkers to see how useful they were in dealing with correcting spell-

Table 1. Comparison of three spelling correction methods (Adopted from Kukich, 1992)

Techniques	521-Word Lexicon	1,142-Word Lexicon	1,872-Word Lexicon
Minimum Edit Distance	64%	62%	60%
Simple Correlation Matrix (Hamming Distance)	69%	68%	67%
Neural Network	75%	75%	?

ing errors in medical portals. The following two commercial products are available in assisting users with difficult medical terminology and names of medications. Spellex (www.spellex.com) and MediSpell (www.medispell.com) were identified as the most popular and comprehensive packages. Spellex searches for medical words of 670,000 words while MediSpell searches 170,000 words.

We compared the two products against the misspelling of medical terms to see what help the users would get. The words tested were chosen from the usability testing done in an earlier study of the effectiveness of Australian government medical portals (Moon & Burstein, 2005; Moon & Fisher, 2005). Heuristic usability terms were chosen as they are closer to natural language processing, that is, the users are dynamic and unpredictable

Table 2. Some spelling correction programs

Name of the product	Strategies	Advantages	Disadvantage	Reference
Spell	Based on addition, deletion, insertion and transposition. The list is matched against the dictionary	Short words correct well.	High percentage errors on long strings. Many orthographical errors fall outside of the scope of SPELL	Peterson, 1980
Speecop	Strings are converted to similarity keys. These similarity keys are blurred to mimic original words. The keys found within certain target keys are considered possible corrections.	Frequent problem of doubling, undoubling and transposition are all seen as original words.	Dependence on first few consonants. This causes problems especially on omission of words. Good with frequent typological errors.	Pollock & Zamora, 1984
Talisman	Designed for typing and knowledge errors.	In built to the system i.e. Microsoft.	Logic based (programmed to system) not semantic based.	Berghel & Andreu, 1988
Trigram Analysis	Use of trigrams to correct misspelling.	Works well on long words. Error position is not important.	Single error may disturb all or virtually all trigrams in a short word. Transposition disturb trigrams and thus difficult to correct.	Van Berkel and De Smedt, 1988
Spell Therapist	Linguistic method for correction of orthographical errors. Combines phonological code and checks dictionary for phonemic words.	Works well for words that are homophonous spellings.	Not suitable for typological errors and does not work for hard phonological differences, i.e., ‘managable’, ‘recommand’ doesn’t correct.	Van Berkel and De Smedt, 1988
The Penguin	Grammar and Spell Checking	Takes into account of idioms, colloquialism, names, and slang expressions. Dynamic, rather than prescriptive such as Microsoft Word.	Not contextual.	Fallman, 2002

in the way they spell the words especially when the word is long and difficult.

We have chosen words such as *schizophrenia*, *methotrexate*, and *Tamoxifen*. Schizophrenia is a common mental disease, Methotrexate is a common drug for sufferers of rheumatoid arthritis, and Tamoxifen is a common drug for breast cancer. These terms were chosen because of their wide usage in the community as found from usability testing (Moon & Fisher, 2005).

Both products give a list of synonyms as a pop-up box as seen in Figures 2-4.

A search for the query on skizophrenia produced the following lists:

- Sacciformis
- Schizophrenic ←
- Schizophrenics
- Shakespeareans

As can be seen, there were four words prompted and they were matched according to the sounds. The word *Shakespeareans* seems rather inappropriate in this list.

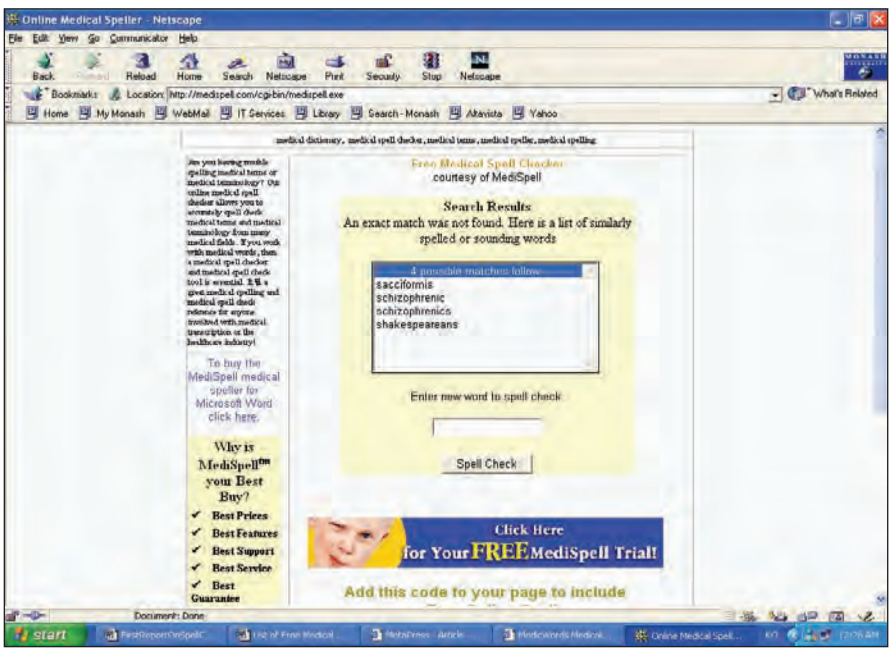
It would also be useful if those hits offered a description of what each word meant. For instance, what is sacciformis? Upon looking it up in a medical dictionary without suffix, *sacciform* is "bag-shaped" or "like a sac" (Davis, 1981), which is also useless in medical context.

Typing a common drug for breast cancer "Tamoxifen" spelled as "Tomoxifen" gave out the following options:

- Tensopin
- timespan
- TinCoBen
- Tamoxifen ←
- Temazepan

An intelligent guess is required to select Tamoxifen as the correct word, as there is no description of definition of any of the options offered. Tensopin is a drug for hypertension, TinCoBen

Figure 2. MediSpell search on "schizophrenia"



is a skin protectant for skin disease, temazepan is a cannabis. Timespan is an unreasonable list of words for selection.

In another search, typing Methotrexate, a common drug for arthritis, misspelt as "Metatrexate" gave the following options (www.spellex.com/cgi-bin/spellex.exe):

As the result the following list of drug names was produced:

- MEDdirect
- Matricide
- Matruchoti
- Maturest
- Methotrexate ←
- Metrizoate
- metrocyte

From the previous list, one could choose more than one selection as there is a lot of similarity with keywords Methotrexate:

- Matricide is a term used for killing a mother
- Matruchoti a type of species of Bacteria represented as *Bacterionema matruchoti*
- Maturest is an English word being a mature
- Metrizoate is a diagnostic radiopaque that usually occurs as the sodium slat
- Metrocyte refers to a neuron.

As can be seen from the list of words offered as alternatives for Methotrexate they have a completely different meaning for the users to choose. Offering words spelled on the basis of phonology as an alternative choice can be risky and at times dangerous, especially when dealing with large ethnic backgrounds. For instance, in Vietnamese the ending letter "t" is often not pronounced. If the user spells phonetically the user could easily type Methotrexate as Methotrax or Mesotrax.

As can be seen from Figures 2, 3, and 4 of screen dumps, phonetically based spelling alone is not very useful and can be extremely time consuming for the users to select individual choices and check on the Internet for their meaning. However, if the descriptions of words are provided to the users so they could select what they need rather than checking individually. Equally, if the list of words is domain specific, that is, for Rheumatoid Arthritis the hit list returns drugs pertaining to

Figure 3. Spellex search on "Tamoxifen"

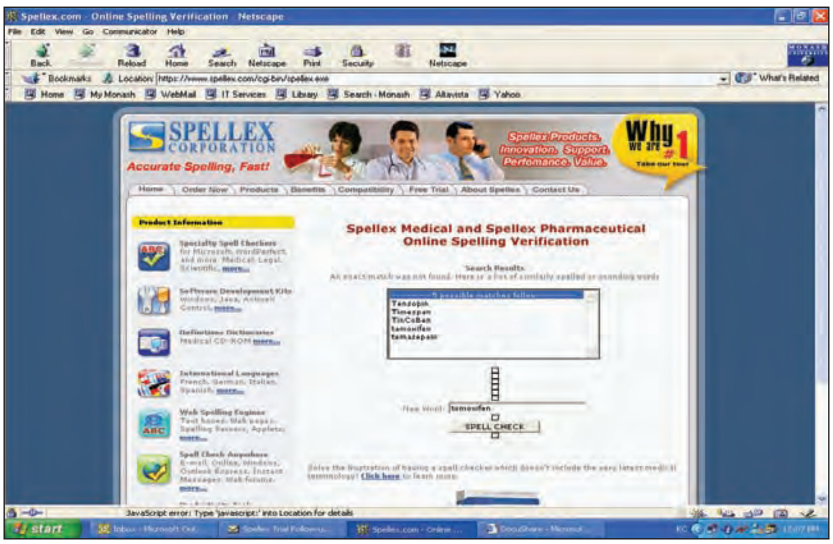
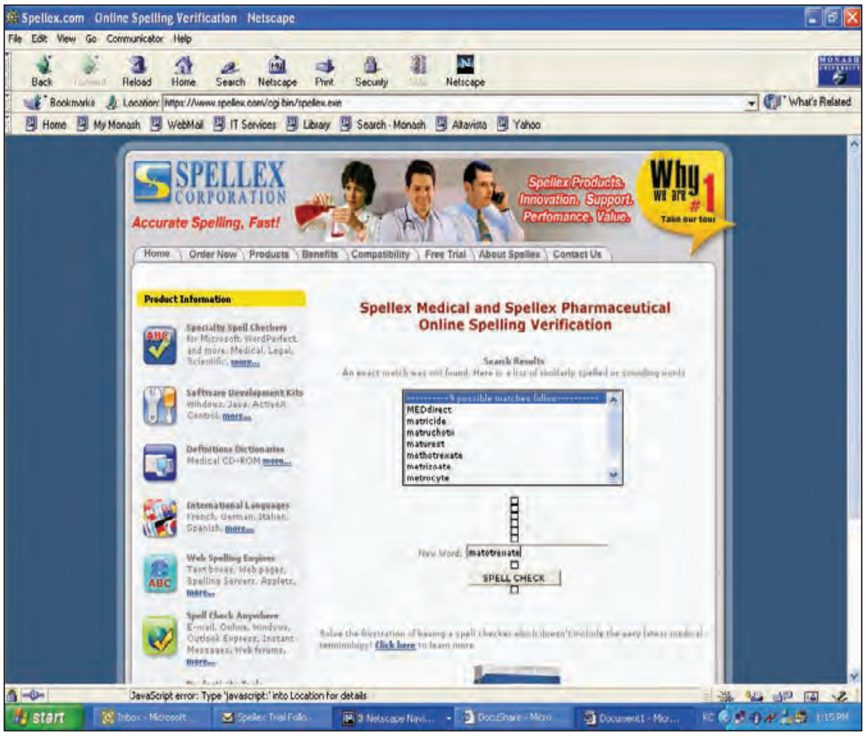


Figure 4. Spellex search on "Methotrexate"



the field would be helpful. The ontology-based search is aiming to solve this problem.

The following section describes steps that can be taken to improve current medical portals in particular correcting spelling mistakes and giving the users hit lists that are relevant to their field of search thereby reducing the time spent for searching, improving the user satisfaction by proving the choice for their selection, and helping them to find the relevant information.

Spelling Correction for Medical Information Retrieval

Generic spelling error detecting tools have been described in the earlier section of this article and their advantages and disadvantages described (Table 2). A few different types of error detection techniques, such as string-string edit distances, statistical packages such as SVD, Hamming distance, and the Levenshtein method have been described. These as well as statistic-based algorithms, where the search is based on probability cannot recognize the search domain. What is lacking in these spell checkers is that the searches are not context based. Spelling error correction based on phonetics combines with context-specific retrieval would eliminate the unnecessary list of words that are offered but not relevant to the user. In addition, ontology-based search can be utilized for searching words specific to the domain thus eliminating words that are not within the specific domain. In this section we consider a few approaches for dealing with phonetic errors in medical contexts, which have proven to be the most frequently encountered misspellings identified from our prior usability study of medical portals. We propose a system architecture, which is based on these approaches. The main advantage of this system is the ability to recognize terms relevant to the users' context as well as provide definitions for these terms.

The following section describes the phonetic and semantic-based search for such a system.

Damerau's Simple Detection Method

By using Damerau's (1964) simple method we can see that 80% of all spelling errors are the results of the following four methods. Damerau's method was developed for English but it can be applied in medical terms as in the following example. The word for liver disease *cirrhosis* can be wrongly written as follows:

1. insertion of a letter: cirrhossis;
2. deletion of a letter; cirhosis;
3. replacement of a letter by another one: cirrhocis; or
4. transposition of two adjacent letters: cirhrocis

In order for a medical portal to recognize these terms as related to the same word, the following methods can be used. These methods combine phonetic and context-based search to achieve more efficient information retrieval.

Phonetic-Based Spell Correction and Aspell

Lawrence Philips' *Metaphone Algorithm* is an algorithm which returns a rough approximation of how an English word sounds (http://aspell.net/metaphone). The original metaphone method was developed in 1990 for the C language, with more later upgrades to 4GI and for Perl, Java, and now Visual C++ and PHP. This method is used widely by Linux as an open source and matches English words phonetically rather than contextually (Jones & Martin, 1997). Lawrence Philips' methods superceded Soundex (developed by Robert Russell and Margaret Odell and was patented in 1918 and 1922) (http://patft.uspto.gov/netacgi/nph-Parser?patentnumber=1,262,167). Lawrence Philip's Metaphone combines the metaphone algorithm based on phonology and Aspell's open source. Aspell's open source spell checker is an advanced version of Ispell's which uses the basic strategy of inserting a space or hyphen, one letter deletion, one letter addition, and interchanging two adjacent letters when used for a correct term. It is an open source and it is free requiring only a little effort to download it.

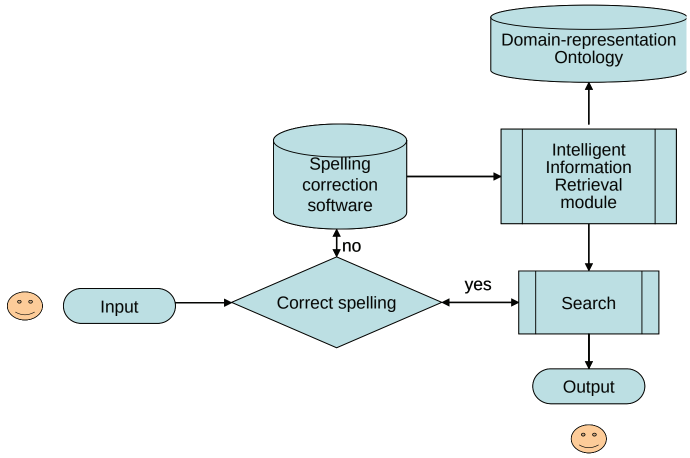
Semantic-Based Search

Kukich (1992) mentioned, in her much quoted review of technologies for automatic word correction, that research on context-based search was still in its infancy and that the task of fixing context-dependent text was an elusive one. She mentioned that 40-45% of spellings are contextual errors (Kukich, 1992a, 1992b). Natural language is complex and according to Zipf's law, "most events occur rarely and a few very common events occur most of the time" (Pedersen, 1996, p. 1). Indeed there has been much progress in parsing technology and grammar checking in English (Bernth, 1997). One example of analysis that incorporate semantic search is *latent semantic analysis (LSA)*.

The LSA proposed by Jones and Martin (1997) promises to be a step further from traditional spelling methods. In traditional spelling methods, the emphasis has been correcting the spelling rather than providing words that are relevant to contexts. Since many words are polysemous (have many meanings) and have many synonyms (words with similar meanings), trying to find the right word that matches the context has not been easy. LSA, with its use of a SVD matrix, has to some extent be able to correct the spelling errors in confusion sets (words such as *principle* and *principal*). However, it does not work yet on unedited texts (Jones & Martin, 1997).

Part of speech (POS). (Ruch, 2002) POS is about an experiment combining string-to-string

Figure 5. Ontology-based spelling correction architecture



distance, POS word + word language and found that POS word + word language improved search. Also spelling corrections using both left and right context outperformed the use of the left context.

Ontology-based query expansion using Unified Medical Language System (UMLS). UMLS is a full set of knowledge sources for the medical domain and is hierarchy based. The research shows that users often use the users use scenario-specific queries for information retrieval. The search terms are not keywords such as *hepatitis* or *cancer* but scenario-based search terms such as *lung, cancer treatment,* and *lung excision.* This is different to query expansion, which sought to search keywords in various documents and tried to link the words with documents. However, ontology-based query would look for keywords as well as the words in context. For instance, *lung excision* not only looks for *lung cancer* but also searches for *smoking, cigarette,* and *lymph node* (Liu & Chu, 2005).

SYSTEM DESIGN

In the system we propose that the user-entered term is first analyzed using phonetic approach as part of a spelling correction software. The terms that this process identifies as possible options are then compared with the domain ontology model. As a result of such comparison those not relevant to the user are excluded, the remaining terms are defined clearly using the definitions stored as part of the ontology. Such search procedure not only increases the relevance of the final results but improves the level of user satisfaction.

FUTURE TRENDS

In general, medical terminology is very challenging. Understanding the general spelling error patterns of adults and thus creating a system that understands the general algorithm of adult spellers could improve the system. For years scientists have worked on improving systems for information retrieval. Clearly a better system is required if users' needs are to be satisfied. Agent theory, use of artificial intelligence via fuzzy logic, artificial neutral networks, and Semantic Web-based systems are all being researched and refined. Ontology-based search has potential in that the listed hits are relevant to specific domain, thereby reducing the retrieval of irrelevant word lists.

CONCLUSION

Natural language processing is complicated, and English is a challenging language when it comes to spelling. Adult spelling errors are either typological or orthographical, and many different types of error detection and error correction methods have been developed. Although these studies have been done on English generally, the same techniques can be applied specifically to medical information retrieval. However, the challenge is that medical terminology or jargon often involves difficult, non-English words. From the usability study it was seen that users often spell phonetically and often use a scenario-based search rather than a keyword search.

Current spelling aid systems which are built on a phonetic basis might have high recall but are low in precision. Combinations of phonetic-based systems with a knowledge-based search are seen to improve searches by 5-10% (Liu & Chu, 2005). What is more, Liu and Chu suggest that user queries are often scenario-based rather than keyword based. Ontology-based searches using a medical corpus such as UMLS narrows the search to the medical context thereby reducing irrelevant information, that is, nonmedical information. With the use of query expansion it could prove to be more efficient.

REFERENCES

Aho, A. V., & Corassick, M. J. (1975). Efficient string matching: An aid to bibliographic search. *Communications of the ACM, 18,* 333-340.

Berghel, H. (1998). Talisman: A prototype expert system for spelling correction. *Proceedings of the 1998 ACM SIGSMALL/PC Symposium on ACTES*.

Berghel, H., & Andreu, C. (1998). Talisman: A prototype expert system for spelling correction. *Proceedings of the 1998 ACM SIGSMALL/PC Symposium on ACTES* (pp. 107-113).

Bernth, A. (1997). EasyEnglish: A tool for improving document quality. *Proceedings of the Fifth Conference on Applied Natural Language Processing.*

Blaser, R., Schnabel, M., Mann, D., Jancke, P., Kuhn, K. A., & Lenz, R. (2004). Potential prevention of medical errors in casualty surgery by using information technology. *Proceedings of the 2004 ACM symposium on Applied computing* (pp. 285-290).

Cherkassky, V., Vassilas, N., Brodt, G. L., Wagner, R. A., & Fischer, M. J. (1974). The string-to-string correction problem. *Journal of ACM, 21,* 168-178.

Damerau, F. J. (1964). A techniques for computer detection and correction of spelling errors. *Communications of the ACM, 7,* 171-176.

Davis, F. A. (1981). Sacciformis. *Taber's Cyclopedic Medical Dictionary.* Davis Company.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis, *JASIS, 41,* 391-407.

Fallman, D. (2002). The Penguin: Using the Web as a database for descriptive and dynamic grammar and spell checking. *CHI 2002* (pp. 616-617). Minneapolis, MN.

Goyal, R. D., & Nagaraja, G. (2002). Generalized brain-state-in-a-box based associative memory for correction words and images. *Proceedings of the 9th International Conference on Neural Information Processing (ICONIP'02), 1,* 291-295.

Jones, P. M., & Martin, H. J. (1997). Contextual spelling correction using latent semantic analysis. *Proceedings of the 5th Conference on Applied Natural Language* (pp. 166-173). Washington, DC.

Khan, L., McLeod, D., & Hovy, E. (2004). Retrieval effectiveness of an ontology-based model for information selection. *The International Journal on Very Large Data Bases, 13,* 71-85.

Kukich, K. (1988). Variations on a back-propagation name recognition net. *Proceedings of the Advanced Technology Conference 2*(2), 722-735.

Kukich, K. (1992). Spelling correction for the telecommunications network for the deaf. *Communications of the ACM, 35,* 80-90.

Liu, Z., & Chu, W. W. (2005). Knowledge-based query expansion to support scenario-specific retrieval of medical free text. *ACM,* 1076-1083.

Mauldin, M. L. (1991). *Conceptual information retrieval*. Kluwer Academic.

McKemmish, S., Burstein, F., Manaszewicz, R., & Fisher, J. (2002). *Towards meeting the decision support needs of a community via an "Intelligent portal": Breast cancer knowledge online*. Retrieved May 21, 2002 from http://sims.edu.au/research/eirg

Mitton, R. (1996). *English spelling and the computer.* Addison-Wesley.

Moon, J. (2005). Discussing health issues on the Internet. In S. Dasgupta (Ed.), *Encyclopedia of virtual communities and technologies.* Hershey, PA: Idea Group Inc.

Moon, J., & Burstein, F. (2005). Intelligent portals for supporting medical information needs. In A. Tatnall (Ed.), *Web portals: The new gateways to internet information and services.* Hershey, PA: Idea Group Inc.

Moon, J., & Fisher, J. (2005). *The effectiveness of Australian medical portals: Are they meeting the health consumers' needs?* Working paper.

Morris, R., & Cherry, L. L. (1975). Computer detection of typographical errors. *IEEE Transactions on Professional Communication, 18,* 54-63.

Nix, R. (1981). Experience with a space efficient way to store a dictionary. *Communications of the ACM, 24,* 297.

Patel, C., Supekar, K., Lee, Y., & Park, E. K. (2003). OntoKhoj: A Semantic Web portal for ontology searching, ranking and classification. *Proceedings of the 5th ACM international workshop on Web information and data management* (pp. 58-61). New Orleans.

Peterson, J. (1980). Computer programs for detecting and correcting spelling errors. *ACM, 23,* 676-687.

Pedersen, T. (1996). Fishing for exactness. *Proceedings of the South-Central SAS Users Group Conference*.

Pollock, J. J., & Zamora, A. (1984). Automatic spelling correction in scientific and scholarly text. *Communications of the ACM, 27,* 358-368.

Ruch, P. (2002). Information retrieval and spelling correction: An Inquiry into lexical disambiguation, *ACM,* 699-703.

Tijerino, Y. A., & Reza, S. (2005). OntoTEMAS: An ontology based teaching materials search engine. *Journal of Computing Sciences in Colleges*, *20,* 177-182.

Yannakoudakis, E. J., & Fawthrop, D. (1983). The rules of spelling errors. *Information Processing Management, 19,* 87-99.

Zamora, A. (1978). Control of spelling errors in large data bases. *The Information Age in Perspective, Proceedings of the ASIS, 15,* 364-367.

Zamora, A. (1980). Automatic detection and correction of spelling errors in a large data base. *Journal of the ASIS, 31,* 51-57.

Zamora, E. M., Pollock, J. J., & Zamora, A. (1981). The use of trigram analysis for spelling error detection. *Information Processing Management, 17,* 305-316.

KEY TERMS

Dictionary Lookup Techniques. Dictionary lookup techniques check whether an input string appears in a dictionary. Response time may become a problem as the size of the dictionary grows. To save the search time most often suffixes are stripped, and this can cause some problems.

Hamming Distance H. H is defined only for strings of the same length. For strings s and t, H(s,t) is the number of places in which the two string differ, or have different characters.

Levenshtein (or edit) Distance. Levenshtein distance is more sophisticated. It is defined for strings of arbitrary length. It counts the differences between two strings, where we would count a difference not only when strings have different characters but when one has a character whereas the other does not. The smallest number of insertions, deletions, and substitutions required to change one string or tree into another. A H(m x n) algorithm to compute the distance between strings, where m and n are the lengths of the strings (http://www.nist.gov/dads/HTML/levenshtein.html).

Minimum Edit Distance. Minimum edit distance is based on calculating the distance between a misspelled word and eh words in its files. Words whose evaluated distance is the smallest are offered as candidates for replacement.

N-Gram is an n-letter subsequence of a string, where *n* is usually 1,2, or 3. In general, n-gram analysis techniques check each n-gram in an input string against a precompiled table of n-gram statistics to determine whether the n-gram can occur in a word. If it does, its frequency of occurrence in the words of the language is computed. Strings containing n-grams that do not occur in words or occur very infrequently are considered to be possible misspellings.

Phoneme. Phoneme is the smallest unit of sound in a language which can distinguish two words, that is, *pan* and *ban* differ in that *pan* begins with /p/ and *ban* begins with /b/.

Simple Correlation Matrix Technique. Simple coorelation matrix technique is a correlation technique. Each misspelled word is represented by an n-dimensional feature where Hamming distance of strongly correlated matches the most probably correct word (Cherkassky et al., 1974).

Trigram Analysis. Trigram analysis is a method used to correct spelling. It simply means three possible candidates for representing words where n-gram is used to apply the analysis.

This work was previously published in Semantic Web Technologies and E-Business: Toward the Integrated Virtual Organization and Business Process Automation, edited by A. Salam and J. Stevens, pp. 384-404, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.29
A Bayesian Framework for Improving Clustering Accuracy of Protein Sequences Based on Association Rules

Peng-Yeng Yin
National Chi Nan University, Taiwan

Shyong-Jian Shyu
Ming Chuan University, Taiwan

Guan-Shieng Huang
National Chi Nan University, Taiwan

Shuang-Te Liao
Ming Chuan University, Taiwan

ABSTRACT

With the advent of new sequencing technology for biological data, the number of sequenced proteins stored in public databases has become an explosion. The structural, functional, and phylogenetic analyses of proteins would benefit from exploring databases by using data mining techniques. Clustering algorithms can assign proteins into clusters such that proteins in the same cluster are more similar in homology than those in different clusters. This procedure not only simplifies the analysis task but also enhances the accuracy of the results. Most of the existing protein-clustering algorithms compute the similarity between proteins based on one-to-one pairwise sequence alignment instead of multiple sequences alignment; the latter is prohibited due to expensive computation. Hence the accuracy of the clustering result is deteriorated. Further, the traditional clustering methods are ad-hoc and the resulting clustering often converges to local optima. This chapter presents a Bayesian framework for improving clustering accuracy of protein sequences based on association rules. The experimental results manifest that the proposed framework can significantly improve the performance of traditional clustering methods.

INTRODUCTION

One of the central problems of bioinformatics is to predict structural, functional, and phylogenetic features of proteins. A protein can be viewed as a sequence of amino acids with 20 letters (which is called the primary structure). The explosive growth of protein databases has made it possible to cluster proteins with similar properties into a family in order to understand their structural, functional, and phylogenetic relationships. For example, there are 181,821 protein sequences in the Swiss-Prot database (release 47.1) and 1,748,002 sequences in its supplement TrEMBL database (release 30.1) up to May 24, 2005. According to the secondary structural content and organization, proteins were originally classified into four classes: α, β, α+β, and α/β (Levitt & Chothia, 1976). Several others (including multi-domain, membrane and cell surface, and small proteins) have been added in the SCOP database (Lo Conte, Brenner, Hubbard, Chothia, & Murzin, 2002). Family is a group of proteins that share more than 50% identities when aligned, the SCOP database (release 1.67) reports 2630 families.

Pairwise comparisons between sequences provide good predictions of the biological similarity for related sequences. Alignment algorithms such as the Smith-Waterman algorithm and the Needleman-Wunsch algorithm and their variants are proved to be useful. Substitution matrices like PAMs and BLOSUMs are designed so that one can detect the similarity even between distant sequences. However, the statistical tests for distant homologous sequences are not usually significant (Hubbard, Lesk, & Tramontano, 1996). Pairwise alignment fails to represent shared similarities among three or more sequences because it leaves the problem of how to represent the similarities between the first and the third sequences after the first two sequences have been aligned. It is suggested in many literatures that multiple sequence alignment should be a better choice. While this sounds reasonable, it causes some problems we address here. The most critical issue is the time efficiency. The natural extension of the dynamic programming algorithm from the pairwise alignment to the multiple alignment requires exponential time (Carrillo & Lipman, 1988), and many problems related to finding the multiple alignment are known to be NP-hard (Wang & Jiang, 1994). The second issue is that calculating a distance matrix by pairwise-alignment algorithm is fundamental. ClustalW (Thompson, Higgins, & Gibson, 1994) is one of the most popular softwares for multiple-alignment problems. It implements the so-called progressive method, a heuristic that combines the sub-alignments into a big one under the guidance of a phylogenetic tree. In fact, the tree is built from a pre-computed distance matrix using pairwise alignment.

Many protein clustering techniques exist for sorting the proteins but the resulting clustering could be of low accuracy due to two reasons. First, these clustering techniques are conducted according to homology similarity, thus a preprocessing of sequence alignment should be applied to construct a homology proximity matrix (or similarity matrix). As we have mentioned, applying multiple sequence alignment among all proteins in a large data set is prohibited because of expensive computation. Instead, an all-against-all pairwise alignment is adopted for saving computation time but it may cause deterioration in accuracy. Second, most of the traditional clustering techniques, such as hierarchical merging, iterative partitioning, and graph-based clustering, often converge to local optima and are not established on statistical inference basis (Jain, Murty, & Flynn, 1999).

This chapter proposes a Bayesian framework for improving clustering accuracy of protein sequences based on association rules. With the initial clustering result obtained by using a traditional method based on the distance matrix, the strong association rules of protein subsequences for each cluster can be generated. These rules satisfying both minimum support and minimum confidence can serve as features to assign proteins to new clusters. We call the process to extract features from clusters the alignment-less alignment. Instead of merely comparing similarity from two protein sequences, these features capture important characteristics for a whole class from the majority, but ignore minor exceptions. These exceptions exist due to two reasons: the feature itself or the sequence itself. For the first reason, the feature being selected could be inappropriate and thus causes exceptions. Or, there does not exist a perfect feature that coincides for the whole class. The second reason is more important. The sequence causing the exception may be pre-classified into a wrong cluster; therefore, it should be re-assigned to the correct one. The Bayes classifier can provide optimal protein classifications by using the *a priori* feature information through statistical inference. As such, the accuracy of the protein clustering is improved.

The rest of this chapter is organized as follows. The background reviews existing methods relevant to protein clustering and the motivations of this chapter. The third section presents the ideas and the theory of the proposed method. The fourth section gives the experimental results with a dataset of protein sequences. The final section concludes this chapter.

BACKGROUND

Related Works

Many clustering techniques for protein sequences have been proposed. Among them, three main kinds of approaches exist, namely the hierarchical merging, iterative partitioning, and graph-based clustering. All of these methods use a pre-computed similarity matrix obtained by performing pairwise alignments on every pair of proteins. In the following we briefly review these approaches.

- **Hierarchical Merging:** The hierarchical merging clustering (Yona, Linial, & Linial, 1999, 2000; Sasson, Linial, & Linial, 2002) starts with a partition that takes each protein as a separate cluster, and then iteratively merges the two clusters that have the highest similarity from all pairs of current clusters. The similarity between two clusters is derived from the average similarity between the corresponding members. Thus, the hierarchical merging procedure forms a sequence of nested clusterings in which the number of clusters decreases as the number of iterations increases. A clustering result can be obtained by specifying an appropriate cutting-off threshold, in other words, the iterative merging procedure progresses until the maximal similarity score between any two clusters is less than the cutting-off threshold. The algorithm for hierarchical merging clustering is summarized in Figure 1.
- **Iterative Partitioning:** The iterative partitioning method (Jain & Dubes, 1988; Guralnik & Karypis, 2001; Sugiyama & Kotani, 2002) starts with an initial partition of *k* clusters. The initial partition can be obtained by arbitrarily specifying *k* proteins as cluster centers then assigning each protein to the closest cluster whose center has the most similarity with this protein. The next partition is obtained by computing the average similarity between each protein and all members of each cluster. The partitioning process is iterated until no protein changes its assignment between successive iterations. Although the iterative partitioning method has some variants like ISODATA (Ball & Hall, 1964) and *K-means* (McQueen, 1967), their general principles can be described as shown in Figure 2. A post-processing stage can be added to refine the clusters obtained from the iterative partitioning by splitting or merging the clusters based on intra-cluster

and inter-cluster similarity scores (Wise, 2002).

- **Graph-based Clustering:** These methods (Tatusov, Koonin, & Lipman, 1997; Enright & Ouzounis, 2000; Bolten, Schliep, Schneckener, Schomburg, & Schrader, 2001) represent each protein sequence as a graph vertex, and every pair of these vertices are connected by an edge with a label on it. The label denotes the similarity score between the two proteins represented by the vertices connecting to the corresponding edge. A partition of the graph can be generated by cutting off the edges whose labels are less than a specified similarity threshold, and each connected component of vertices corresponds to a cluster of proteins since the similarity scores between proteins from the same component are higher than those between proteins from different components. The general idea of the graph-based clustering is outlined in Figure 3. A post-refining process can be conducted by using the graph-based clustering result as the input to a cluster-merging algorithm which iteratively merges the nearest neighboring clusters if the relative entropy decreases with the merging of the clusters (Abascal & Valencia, 2002).

Motivations

The protein clustering result obtained by using the above mentioned approaches could be of low accuracy. This is partly due to the reason that these clustering methods use only pairwise sequence alignment information and partly because these clustering techniques often converge to local optima. Since it is computationally prohibitive to derive multiple sequence alignment information, an alternative is to calculate the statistics from the sequences directly by using data mining techniques. More precisely, the *association rules* between the amino acids in the protein sequences are mined within each cluster. Then the rules satisfying minimum confidence can serve as salient features to identify each cluster. Moreover, matching each protein sequence with the association rule provides a good estimate of the *a priori* probability that the protein satisfies the rule. The statistical inference can compensate the accuracy inadequacy of the clustering result due to the pairwise alignment and the local clustering technique. Therefore, we propose to improve the clustering accuracy by using the Bayes classifier with the conditional probabilities of association rules with each cluster and the alignment scores between protein sequences and these rules.

Figure 1. Summary of hierarchical merging clustering

```
        create.cluster(p);
Repeat
        Find clusters x, y such that similarity(x, y) is maximal;
        If similarity(x, y)>cutting_off_threshold then merge.cluster(x, y);
        Otherwise, terminate;
```

Figure 1. Summary of hierarchical merging clustering.

Figure 2. Summary of iterative partitioning method

Select *k* proteins as initial cluster centers
Repeat
    Compute the average similarity between each protein and each cluster;
    Generate a new partition by assigning each protein to its closest cluster;
Until no proteins change assignment between successive iterations

Figure 3. Summary of graph-based clustering

Represent each input sequence as a vertex, and every pair of vertices are connected by an edge.
Label each edge with the similarity score between the two connected vertices.
Create a partition of the graph by cutting off the edges whose similarity scores are less than a specified threshold.

METHODS

With the assistance of association rule mining and Bayes classifier, our system improves the clustering accuracy of traditional protein-clustering methods. The system overview is shown in Figure 4. First, a traditional protein-clustering approach (either one of the hierarchical merging, iterative partitioning, or graph-based clustering methods) is performed to obtain an initial clustering result of the input protein sequences. In general, the traditional protein-clustering approach consists of three steps as shown in the upper grey box: (1) perform the local alignment (such as BLAST with scoring matrix of BLOSUM 62) between each pair of protein sequences, (2) construct a distance matrix from the raw distance scores (such as the E-values produced by BLAST) of the local alignment, and (3) apply the clustering method with the distance matrix to get the clustering result.

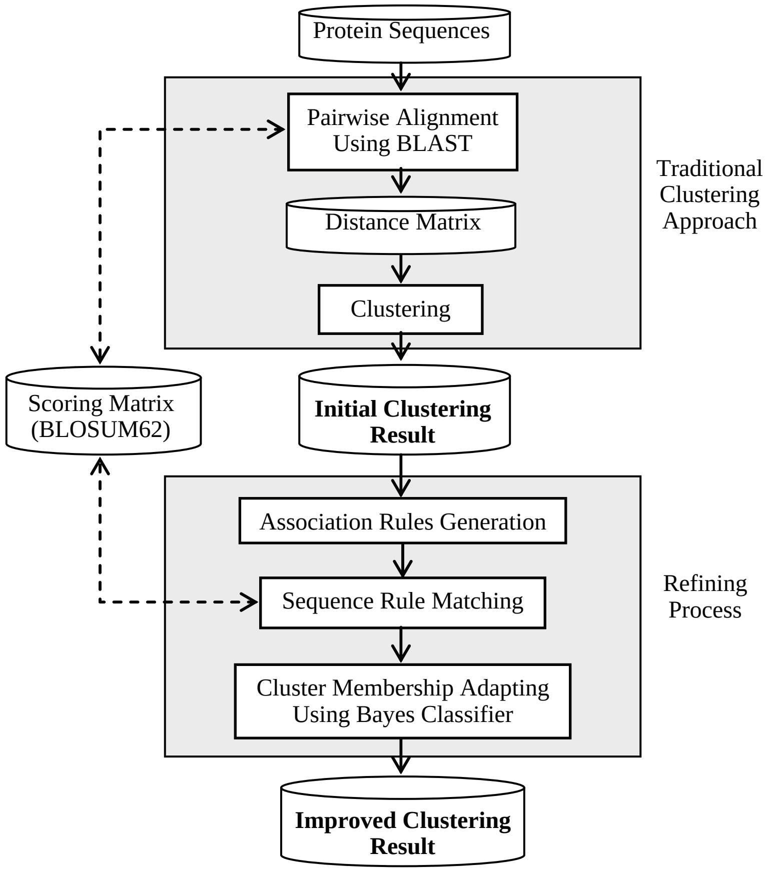
Second, the refining process (as shown in the lower grey box) is performed such that the clustering result is improved. The refining process also consists of three steps: (1) generate strong association rules which satisfy minimum confidence for each cluster, (2) perform sequence rule matching between each protein sequence and each association rule based on local alignment, and (3) adapt cluster membership of each protein sequence by using Bayes classifier and its matching score. The proposed refining process improves the clustering result based on statistical inference. The association rules used are statistically confident and the reassignment of protein sequences using Bayes classifier satisfies the maximum *a posteriori* criterion.

We now present the refining process in details.

Sequence Association Rule Generation

Association rule mining has been intensively used for finding correlation relationships among a large set of items and has delivered many successful

Figure 4. System overview



applications, such as catalog design, cross marketing, and loss-leader analysis (Han & Kamber, 2001). Association rule mining finds significant associations or correlation relationships among a large repository of data items. These relationships, represented as rules, can assist the users to make their decisions. Traditionally, association rule mining works with unordered itemset, that is, the order of the items appearing in the itemset does not matter (Agrawal, Imielinski, & Swami, 1993). However, the order of the amino acids appearing in a protein sequence reserves important phylogeny information among homologies and should be taken into account. The *sequence Apriori algorithm* (Agrawal & Srikant, 1994, 1995) which adapts the classical association rule mining algorithm for sets of sequences can be used for this purpose.

Given a cluster C of protein sequences, the algorithm finds sets of sequences that have support above a given *minimum support*. The *support* of a sequence X with respect to C, denoted by $\text{support}_C(X)$, is the number of sequences in C that contain X as subsequence. A sequence is called *frequent* if its support is greater than the given *minimum support*. To expedite the search for all frequent sequences, the algorithm enforces an iterative procedure based on the a priori property

Figure 5. An illustrative example of sequence association rule matching: (a) A substitution matrix, (b) Two possible alignments between the sequence and the association rule

	a	b	c	d	
a	2	-1 2	0		
b	-1 3	-	2	1	
c	2	-2 2	-	1	
d	0	1	-1 4		

(a)

Sequence:	*adabdacd*	Score
Association rule:	*dadd*	4+2+1+4=11
Association rule:	*dadd*	4+2 -1+4= 9

(b)

which states that any subsequence of a frequent sequence must be also frequent. It starts with the set of frequent 1-sequences which are the frequent sequences of length 1 and uses this set to find the set of frequent 2-sequences, then the set of frequent 2-sequences is used to find the set of frequent 3-sequences, and so on, until no more longer frequent sequences can be found. As the number of frequent sequences can be extremely large for a cluster of protein sequences having length of hundreds of amino acids, we retain the frequent sequences that are not contained in longer ones and restrict the search for frequent sequences of length 5 to 10. The reduced set of frequent sequences is sufficient for deriving similarity statistics in homology since longer frequent sequences of length more than 10 usually have lower support and a number of short frequent sequences within the specified range of lengths that the longer ones contain as subsequences can still be reserved.

With the set of found frequent sequences, we generate sequence association rules as follows. For a given frequent sequence *X*, it can be divided into two disjoint subsequences *A* and *B* with their position information attached. For example, let *X* be '*abcde*', one of the possible divisions could be *A* = '*a_c_e*' and *B* = '*_b_d_*'. A sequence association rule is of the form $A \Rightarrow B$ where both *A* and *B* contain at least one amino acid. *A* is called the rule *antecedent* and *B* is called the rule *consequent*. A sequence association rule is *strong* if it has a confidence value above a given *minimum confidence*. The confidence of a rule $A \Rightarrow B$ with respect to protein cluster *C*, denoted by $\text{confidence}_C(A \Rightarrow B)$, is defined as:

$$\text{confidence}_C(A \Rightarrow B) = \frac{\text{support}_C(A \cup B)}{\text{support}_C(A)} \quad (1)$$

where denotes the supersequence that is properly divided into *A* and *B*. We generate all strong sequence association rules for each cluster of protein sequences.

Sequence Association Rule Matching

In order to determine the cluster membership of each protein, we need to propose a measure which estimates the possibility that the evolution of a protein follows a particular association rule. Here we propose a rule matching scheme which is analogous to local sequence alignment without gaps. Given a sequence *t* and a strong association rule *r*, we compute the alignment score between *t* and *r* according to a substitution matrix with the constraint that no gap is allowed and the rule *antecedent* cannot be substituted in order to detect most similarities. For example, suppose we use the substitution matrix as shown in Figure 5 (a). Let the sequence be *'adabdacd'* and the association rule be '*da_d'* ⇒ '_ _*d_*', there are two possible alignments as illustrated in Figure 5 (b) and the best alignment score without gaps is 11.

Assume that we obtain a set of strong association rules $\Re = \{r_1, r_2, \ldots, r_n\}$ from the procedure of sequence association rule generation, the probability that the evolution of a protein *t* follows a particular association rule r_i can be estimated by:

$$p(r_i \mid t) = \frac{w(r_i, t)}{\sum_{h=1}^{n} w(r_h, t)} \qquad (2)$$

where $w(r_i, t)$ is the alignment score between r_i and *t*. As such, we can use the probabilities:

$$\left(p(r_1 \mid t),\ p(r_2 \mid t), \ldots,\ p(r_n \mid t)\right)$$

as the feature values of protein *t*, and determine its cluster membership by using the Bayes classifier.

Cluster Membership Adapting by Using Bayes Classifier

The Bayes classifier is one of the most important techniques used in data mining for classification. It predicts the cluster membership probabilities based on statistical inference. Studies have shown that the Bayes classifier is comparable in performance with decision tree and neural network classifiers (Han & Kamber, 2001). Herein we propose to predict the cluster membership of each protein by using the Bayes classifier with association rules.

Let the initial clustering result consists of *k* clusters, denoted by $\Im = \{C_1, C_2, \ldots, C_k\}$, from which a set of *n* strong association rules is derived. The *a priori* probability $p(C_i)$ that a protein belongs to cluster C_i can be calculated by counting the ratio of C_i in size to the whole set of proteins. The condition probability $p(r_i \mid C_j)$ that a protein satisfies association rule r_i given that this protein is initially assigned to cluster C_j is estimated by the average probability $p(r_i \mid t)$ for any $t \in C_j$. The conditional probability $p(r_i \mid t)$ that the evolution of protein *t* follows association rule r_i is estimated by using Equation (2). We then use the naïve Bayes classifier to assign protein *t* to the most probable cluster C_{Bayes} given by:

$$C_{Bayes} = \arg\max_{C_j \in \Im} \left\{ \prod_{i=1}^{n} \left(p(r_i \mid C_j)\, p(r_i \mid t) \right) p(C_j) \right\} \qquad (3)$$

In theory, the naïve Bayes classifier makes classification with the minimum error rate.

EXPERIMENTAL RESULTS

We validate our method by using protein sequences selected from SCOP database (release 1.50, Murzin, Brenner, Hubbard, & Chothia, 1995; Lo Conte et al., 2002) which is a protein classification created manually. SCOP provides a hierarchy of known protein folds and their detailed structure information. We randomly select 1189 protein sequences from SCOP and these sequences compose 388 protein clusters according to manual annotations on structure domains. The mean length of these protein sequences is 188 amino acids. Our method is implemented in C++ programming language and the experiments are conducted on a personal computer with a 1.8 GHz CPU and 512 MB RAM.

Accuracy Measures

To define the accuracy measures of protein classification, some notations are first introduced. For every pair of protein sequences, the predicted classification and the annotated classification have four possible combinations. *TP* (true positive) is the number of pairs predicted in the same domain given that they are in the same SCOP domain, *TN* (true negative) is the number of pairs predicted with different domains given that they are in different SCOP domains, *FP* (false positive) is the number of different SCOP-domain pairs that are predicted in the same domain, and *FN* (false negative) is the number of SCOP-domain pairs that are predicted in different domains. Two accuracy measures, namely *sensitivity* (S_n) and *specificity* (S_p), are defined as follows:

$$S_n = \frac{TP}{TP + FN} \tag{4}$$

and

$$S_p = \frac{TP}{TP + FP} \tag{5}$$

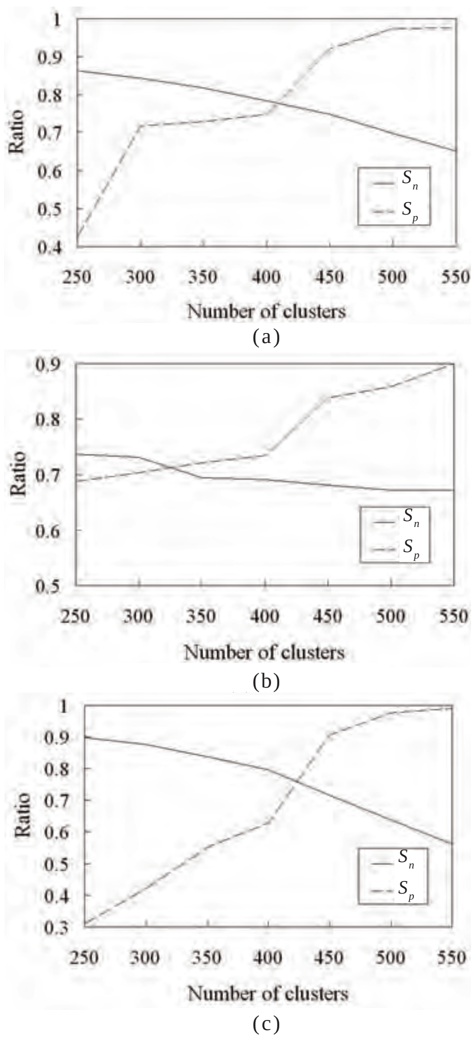
Sensitivity is the proportion of SCOP-domain pairs that have been correctly identified, and specificity is the proportion of pairs predicted in the same domain that are actually SCOP-domain pairs. Sensitivity and specificity cannot be used alone since perfect sensitivity can be obtained if all the pairs are predicted in the same domain, and specificity is not defined if all the pairs are predicted in different domains. To compare the performance between two competing methods, a sensitivity vs. specificity curve is usually used for evaluation. Or alternatively, the mean of sensitivity and specificity values can be used as a unified measure.

Performance Evaluation

We first evaluate the clustering performance using the traditional methods, in particular, we have implemented the hierarchical merging method (Sasson et al., 2002), the K-means algorithm (modified from Guralnik & Karypis, 2001, by changing the feature space to pair-wise E-value), and the graph-based clustering (Bolten et al., 2001). Various clustering thresholds are specified to obtain the performances at different specified numbers of clusters. Figures 6 (a) - 6 (c) show the variations of sensitivity and specificity as the number of clusters increases for hierarchical merging, K-means, and graph-based clustering, respectively. These curves are intuitive since the mean size of clusters is smaller if the partition with more clusters is obtained and, in general, the smaller the cluster-mean size is, the lower the sensitivity is, but the higher the specificity is. If we take the accuracy values obtained when the number of clusters is equal to the true number (388), the hierarchical merging method with S_n = 0.79 and S_p = 0.73 is superior to the K-means algorithm which produces S_n = 0.69 and S_p = 0.72, while the graph-based clustering has medium performance (S_n = 0.80 and S_p = 0.63).

Next we evaluate the improvement for clustering accuracy due to our Bayesian framework. From the clustering results obtained by traditional clustering methods, we apply the sequence association rule mining to generate strong association rules. Each protein sequence is matched with these rules and updates its cluster membership by using the Bayes classifier in order to improve the accuracy. Because the sensitivity values obtained by hierarchical merging, K-means, and graph-based clusterings with the true cluster number (388) are 0.79, 0.69, and 0.80, respectively, we compute the specificity improvement for these methods within a range of sensitivity close to these values. Figures 7 (a) - 7 (c) show the sensitivity vs. specificity curve for illustrating the improvement achieved. It is observed that the proposed refining process

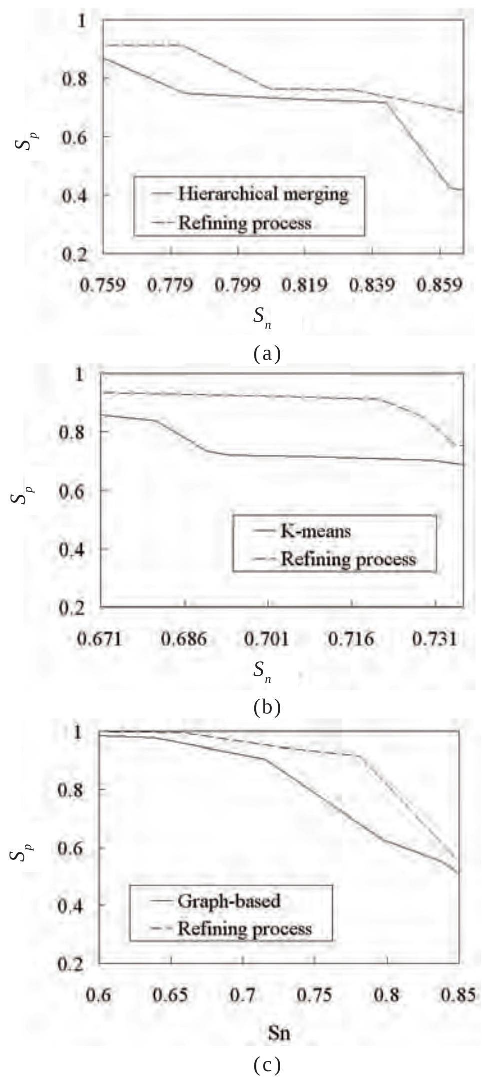
Figure 6. Performances of traditional clustering methods, (a) hierarchical merging clustering, (b) K-means clustering, (c) graph-based clustering



can significantly improve the specificity of the traditional methods. The average improvements in Sp are 0.09, 0.11, and 0.06 for hierarchical merging, K-means, and graph-based clustering, respectively. We also compute the accuracy improvement by using the unified measure of $(S_n + S_p)/2$ as shown in Figs. 8 (a) - 8 (c). The average improvements over various specified numbers of clusters are 0.03, 0.04, and 0.03 for hierarchical merging, K-means, and graph-based clustering, manifesting the robustness of the proposed framework.

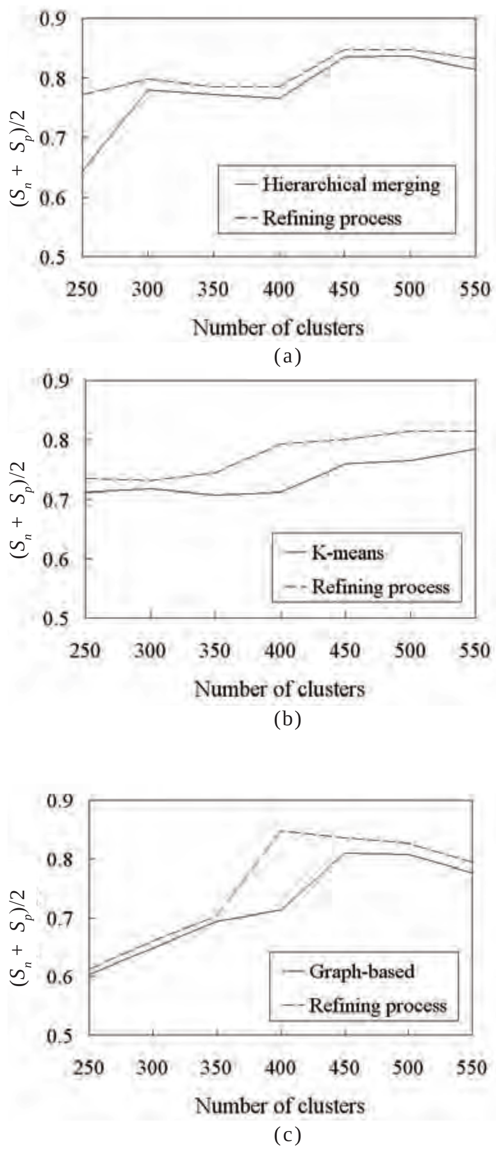
Table 1 shows the incurred computation time (in seconds) by each component of the proposed framework and the corresponding percentage to the whole for our collective database. We observe that the distance matrix computation using pairwise sequence alignment and the association rules generation using sequence Apriori algorithm are the most time-consuming components, and they

Figure 7. Performance improvement for specificity vs. sensitivity, (a) hierarchical merging clustering, (b) K-means clustering, (c) graph-based clustering



consume 35% and 33% of the total computation time, respectively. Further, the computation time for the refining process involving the last two components of Table 1 is about half of the whole time needed. That is, for our collective database, the proposed framework provides a considerable amount of accuracy improvement (as shown in Figures 7-8) for the traditional protein-clustering algorithms by doubling the computation time. In general cases, the time proportion for the refining process diminishes when the number of sequences in the database increases. Suppose there are m sequences in the database, and for simplicity, assume that their average length is r. The time complexity for calculating a distance matrix is $\Theta(m^2 r^2)$, for performing traditional clustering is $\Omega(m^2)$, and for generating the association rules of fixed length is $\Theta(mr)$. Rule matching and Bayesian classification depends on two factors: the number of features extracted and the number

Figure 8. Performance improvement for the unified measure, (a) hierarchical merging clustering, (b) K-means clustering, (c) graph-based clustering



of clusters classified (and remember that each rule is within a fixed length). The former can be regarded as a constant once the lengths of rules and the alphabet (which is 20 for amino acids) are fixed to constants. Therefore a rough calculation asserts that it takes $\Theta(mrc)$ where c is the number of clusters. As a summary, the time complexity for each component is $\Theta(m^2 r^2)$, $\Omega(m^2)$, $\Theta(mr)$, and $\Theta(mrc)$, respectively. Parameter r, the average length of peptide sequences, is around 300 and can be regarded as a constant. Parameter c grows mildly as the number of sequences increases. As a result, when the number of sequences in a database grows larger, the percentage of the computing time on the refining process becomes smaller.

Table 1. The computation time used by each component of the proposed method

	CPU Time (second)	Percentage
Distance Matrix Computation	162	35%
Traditional Clustering	6 4	14%
Association Rules Generation	157	33%
Rule Matching and Bayesian Classification	8 7	18%

SUMMARY

Protein sequence clustering is useful for structural, functional, and phylogenetic analyses. As the rapid growth in the number of sequenced proteins prohibits the analysis using multiple sequences alignment, most of the traditional protein-clustering methods derive the similarity among sequences from pairwise sequence alignment. In this chapter, we have proposed a Bayesian framework based on association rule mining for improving the clustering accuracy using existing methods. A selective dataset from SCOP has been experimented and the result manifests that the proposed framework is feasible. The main features of the proposed framework include:

- The proposed framework improves the accuracy of a given initial clustering result which can be provided by any clustering methods. Therefore, the user is still able to choose a particular clustering method which is suited to his/her own analysis of the final result.
- From the initial clustering result, the sequence association rules among amino acids are mined. These rules represent important relationships for the sequences belonging to the same cluster.
- The association rules serve as features for classification of proteins. Using Bayes classifier, the classification error can be minimized based on statistical inference.

Future research is encouraged in expediting the computation for distance matrix and association rule generation in order to extend the application of the proposed framework to large protein databases such as SWISSPROT.

REFERENCES

Abascal, F., & Valencia, A. (2002). Clustering of proximal sequence space for the identification of protein families. *Bioinformatics, 18*(7), 908-921.

Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *Proceedings of ACM-SIGMOD, International Conference on Management of Data*, Washington, DC (pp. 207-216).

Agrawal, R., & Srikant, R. (1994). Algorithms for mining association rules. *Proceedings of the 20th International Conference on Very Large Databases*, Santiago, Chile (pp. 487-499).

Agrawal, R., & Srikant, R. (1995). Mining sequential patterns. *Proceedings of the International*

Conference on Data Engineering (ICDE), Taipei, Taiwan.

Ball, G. H., & Hall, D. J. (1964). Some fundamental concepts and synthesis procedures for pattern recognition preprocessors. *Proceedings of International Conference on Microwaves, Circuit Theory, and Information Theory*, Tokyo.

Bolten, E., Schliep, A., Schneckener, S., Schomburg, D., & Schrader, R. (2001). Clustering protein sequences—structure prediction by transitive homology. *Bioinformatics, 17*(10), 935-941.

Carrillo, H., & Lipman, D. (1988). The multiple sequence alignment problem in biology. *SIAM Journal on Applied Mathematics, 48,* 1073-1082.

Enright, A. J., & Ouzounis, C. A. (2000). GeneRAGE a robust algorithm for sequence clustering and domain detection. *Bioinformatics, 16*(5), 451-457.

Guralnik, V., & Karypis, G., (2001, August 26). A scalable algorithm for clustering protein sequences. *Proceedings of Workshop on Data Mining in Bioinformatics,* San Francisco (pp. 73-80). ACM Press.

Han, J., & Kamber, M., (2001). *Data mining: Concepts and techniques.* San Francisco: Morgan Kaufmann Publishers.

Hubbard, T. J., Lesk, A. M., & Tramontano, A. (1996). Gathering them into the fold. *Nature Structure Biology, 4*, 313.

Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data*. Englewood Cliffs, NJ: Prentice Hall.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, *13*(3), 264-323.

Levitt, M., & Chothia, C. (1976). Structural patterns in globular proteins. *Nature, 261*, 552-558.

Lo Conte, L., Brenner, S. E., Hubbard, T. J. P., Chothia, C., & Murzin, A. G. (2002). SCOP database in 2002; refinements accommodate structural genomics. *Nucleic Acids Research, 30*, 264-267.

McQueen, J.B. (1967). Some methods of classification and analysis of multivariate observations. *Proceedings of Fifth Berkeleys Symposium on Mathematical Statistics and Probability* (pp. 281-297).

Murzin A. G., Brenner S. E., Hubbard T., & Chothia C. (1995). SCOP: A structure classification of proteins database for the investigation of sequences and structures. *Journal of Molecular Biology*, *247*, 536-540.

Sasson, O., Linial, N., & Linial, M. (2002). The metric space of proteins—comparative study of clustering algorithms. *Bioinformatics, 18*(1), S14-S21.

Sugiyama, A., & Kotani, M. (2002). Analysis of gene expression data by self-organizing maps and k-means clustering. *Proceedings of IJCNN* (pp. 1342-1345).

Tatusov, R. L., Koonin, E. V., & Lipman, D. J. (1997). A genomic perspective on protein families. *Science, 278*, 631-637.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Research, 22*, 4673-4680.

Wang, L., & Jiang, T. (1994). On the complexity of multiple sequence alignment. *Journal of Computational Biology, 1*, 337-348.

Wise, M. J. (2002). The POPPs: clustering and searching using peptide probability profiles. *Bioinformatics*, *18*(1), S38-S45.

Yona, G., Linial, N., & Linial, M. (1999). ProtoMap: Automatic classification of protein sequences, a hierarchy of protein families, and local maps of the protein space. *Proteins, 37*, 360-378.

Yona, G., Linial, N., & Linial, M. (2000). ProtoMap: Automatic classification of protein sequences, a hierarchy of protein families, and local maps of the protein space. *Nucleic Acids Research*, *28*, 49-55.

Chapter 7.30
Retrieving Medical Records Using Bayesian Networks

Luis M. de Campos
Universidad de Granada, Spain

Juan M. Fernández Luna
Universidad de Granada, Spain

Juan F. Huete
Universidad de Granada, Spain

INTRODUCTION

Bayesian networks (Jensen, 2001) are powerful tools for dealing with uncertainty. They have been successfully applied in a wide range of domains where this property is an important feature, as in the case of information retrieval (IR) (Turtle & Croft, 1991). This field (Baeza-Yates & Ribeiro-Neto, 1999) is concerned with the representation, storage, organization, and accessing of information items (the textual representation of any kind of object). Uncertainty is also present in this field, and, consequently, several approaches based on these probabilistic graphical models have been designed in an attempt to represent documents and their contents (expressed by means of indexed terms), and the relationships between them, so as to retrieve as many relevant documents as possible, given a query submitted by a user.

Classic IR has evolved from flat documents (i.e., texts that do not have any kind of structure relating their contents) with all the indexing terms directly assigned to the document itself toward structured information retrieval (SIR) (Chiaramella, 2001), where the structure or the hierarchy of contents of a document is taken into account. For instance, a book can be divided into chapters, each chapter into sections, each section into paragraphs, and so on. Terms could be assigned to any of the parts where they occur. New standards, such as SGML or XML, have been developed to represent this type of document. Bayesian network models also have been extended to deal with this new kind of document.

In this article, a structured information retrieval application in the domain of a pathological anatomy service is presented. All the medical records that this service stores are represented in XML, and our contribution involves retrieving records that are relevant for a given query that could be formulated by a Boolean expression on some fields, as well as using a text-free query on other different fields. The search engine that answers this second type of query is based on Bayesian networks.

BACKGROUND

Probabilistic retrieval models (Crestani et al., 1998) were designed in the early stages of this discipline to retrieve those documents relevant to a given query, computing the probability of relevance. The development of Bayesian networks and their successful application to real problems has caused several researchers in the field of IR to focus their attention on them as an evolution of probabilistic models. They realized that this kind of network model could be suitable for use in IR, specially designed to perform extremely well in environments where uncertainty is a very important feature, as is the case of IR, and also because they can properly represent the relationships between variables.

Bayesian networks are graphical models that are capable of representing and efficiently manipulating n-dimensional probability distributions. They use two components to codify qualitative and quantitative knowledge, respectively: first, a directed acyclic graph (DAG), G=(V,E), where the nodes in V represent the random variables from the problem we want to solve, and set E contains the arcs that join the nodes. The topology of the graph (the arcs in E) encodes conditional (in)dependence relationships between the variables (by means of the presence or absence of direct connections between pairs of variables); and second, a set of conditional distributions drawn from the graph structure. For each variable $X_i \in V$, we therefore have a family of conditional probability distributions $P(X_i \mid pa(X_i))$, where $pa(X_i)$ represents any combination of the values of the variables in $Pa(X_i)$, and $Pa(X_i)$ is the parent set of X_i in G. From these conditional distributions, we can recover the joint distribution over V.

This decomposition of the joint distribution gives rise to important savings in storage requirements. In many cases, it also enables probabilistic inference (propagation) to be performed efficiently (i.e., to compute the posterior probability for any variable, given some evidence about the values of other variables in the graph).

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid pa(X_i))$$

The first complete IR model based on Bayesian networks was the Inference Network Model (Turtle & Croft, 1991). Subsequently, two new models were developed: the Belief Network Model (Calado et al., 2001; Reis, 2000) and the Bayesian Network Retrieval Model (de Campos et al., 2003, 2003b, 2003c, 2003d). Of course, not only have complete models been developed in the IR context, but also solutions to specific problems (Dumais, et al., 1998; Tsikrika & Lalmas, 2002; Wong & Butz, 2000).

Structural document representation requires IR to design and implement new models and tools to index, retrieve, and present documents according to the given document structure. Models such as the previously mentioned Bayesian Network Retrieval Model have been adapted to cope with this new context (Crestani et al., 2003, 2003b), and others have been developed from scratch (Graves & Lalmas, 2002; Ludovic & Gallinari, 2003; Myaeng et al., 1998).

MAIN THRUST

The main purpose of this article is to present the guidelines for construction and use of a Bayesian-network-based information retrieval system. The source document collection is a set of medical records about patients and their medical tests stored in an XML database from a pathological anatomy service. By using XML tags, the information can be organized around a well-defined structure. Our hypothesis is that by using this structure, we will obtain retrieval results that better match the physicians' needs. Focusing on the structure of the documents, data are distributed between two different types of tags: on the one hand, we could consider fixed domain tags (i.e., those attributes from the medical record with a set of well-defined values, such as sex, birthdate, address, etc.); and on the other hand, free text passages are used by the physicians to write comments and descriptions about their particular perceptions of the tests that have been performed on the patients, as well as any conclusions that can be drawn from the results. In this case, there is no restriction on the information that can be stored. Three different free-text passages are considered, representing a description of the microscopic analysis, the macroscopic analysis, and the final diagnostic, respectively.

Physicians must be able to use queries that combine both fixed and free-text elements. For example, they might be interested in all documents concerning males who are suspected of having a malignant tumor. In order to tackle this problem, we propose a two-step process. First, a Boolean retrieval task is carried out in order to identify those records in the dataset, mapping the requirements of the fixed domain elements. The query is formulated by means of the XPath language. These records are then the inputs of a Bayesian retrieval process in the second stage, where they are sorted in decreasing order of their posterior probability of relevance to the query as the final output of the process.

The Bayesian Network Model

Since, for those attributes related to fixed domains, it is sufficient to consider a Boolean retrieval, the Bayesian model will be used to represent both the structural and the content information related to free-text passages. In order to specify the topology of the model (a directed acyclic graph, representing dependence relationships), we need to determine which information components (variables) will be considered as relevant. In our case, we can distinguish between two different types of variables: the set that contains those terms used to index the free-text passages, $T = \{T_1,\ldots,T_M\}$, with M being the total number of index terms used; and set D, representing the documents (medical records) in the collection. In this case, we consider as relevant variables the whole document D_k and also the three subordinate documents that comprise it: macroscopic description, D_{mk}; microscopic description, $D_{\mu k}$; and final diagnostic, D_{fk} (generically, any of these will be represented by $D_{\bullet k}$). Therefore,

$$D = \{D_1, D_{m1}, D_{\mu 1}, D_{f1},\ldots,D_N, D_{mN}, D_{\mu N}, D_{fN}\},$$

with N being the number of documents that comprise the collection[1].

Each term variable, T_i, is a binary random variable taking values in the set $\{\bar{t}_i, t_i\}$, where $\bar{t}_i$ stands for the term T_i is not relevant, and t_i represents the term T_i is relevant. The domain of each document variable, D_j, is the set $\{\bar{d}_j, d_j\}$, where, in this case, $\bar{d}_j$ and d_j mean the document D_j is not relevant for a given query, and the document D_j is relevant for the given query, respectively. A similar reasoning can be stated for any subordinate document, $D_{\bullet j}$.

In order to specify completely the model topology, we need to include those links representing the dependence relationships between variables. We can distinguish two types of nodes. The first type links between each term node $T_i \in T$ and each subordinate document node $D_{\bullet j} \in D$, whenever T_i belongs to $D_{\bullet j}$. These links reflect the dependence between the (ir)relevance values of this document and the terms used to index it and will be directed from terms to documents. Therefore, the parent set of a document node $D_{\bullet j}$ is the set of term nodes that belong $D_{\bullet j}$ to (i.e., $Pa(D_{\bullet j}) = \{T_i \in T \mid T_i \in D_{\bullet j}\}$).

The second type links by connecting each subordinate document $D_{\bullet j}$ with the node document D_j to which it belongs, reflecting the fact that the relevance of a document to a query will depend only on the relevance values of its subordinate documents. These links will be directed from subordinate to document nodes.

It should be noted that we do not use links between terms and documents, because we consider these to be independent, given that we know the relevance value of the subordinate documents. Consequently, we have designed a layered topology for our model that also represents the structural information of the medical records. Figure 1 displays the graph associated with the Bayesian network model.

Probability Distributions

The following step to complete the design of the model is the estimation of the quantitative components of the Bayesian network (i.e., the probability distributions stored in each node). For term nodes, and taking into account that all terms are root nodes, marginal distributions need to be stored. The following estimator is used for every term T_i: $p(t_i) = (1/M)$ and $p(\bar{t}_i) = (M-1)/M$. Therefore, the prior probability of relevance of any term is very small and inversely proportional to the size of the index.

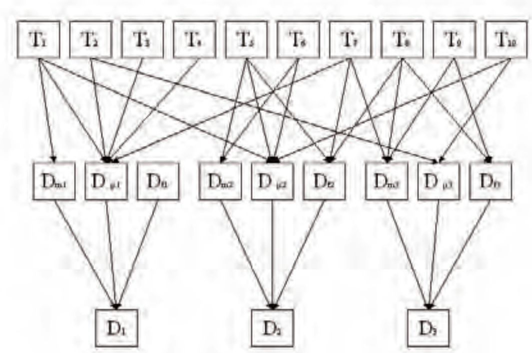
Considering now document and subordinate document nodes, it is necessary to assess a set of conditional probability distributions, the size of which grows exponentially with the number of parents. We therefore propose using a canonical model that represents the particular influence of each parent in the relevance of the node. In particular, given a variable X_j (representing a document or a subordinate document node), the probability of relevance given a particular configuration of the parent set $pa(X_j)$ is computed by means of

$$p(x_j \mid pa(X_j)) = \sum_{i \in pa(X_j)} w_{ij},$$

where the expression $i \in pa(X_j)$ means that only those weights where the value assigned to the ith parent of X_j in the configuration $pa(X_j)$ is relevant will be included in the sum. Therefore, the greater the number of relevant variables in $pa(X_j)$, the greater the probability of relevance of X_j.

The particular values of the weights w_{ij} are first, for a subordinate document, $D_{\bullet j}$, $w_{ij} = (tf_{ij} idf_i^2)/(\sum_{T_k \in Pa(D_{\bullet j})} tf_{kj} idf_k^2)$, with tf_{ij} being the frequency of the term ith in the subordinate document and idf_i the inverse document frequency of the term T_i in the whole collection; and second, for a document node, D_j, we use three factors $\alpha = w_{mj,j}$, $\beta = w_{\mu j,j}$ and $\delta = w_{fj,j}$, representing the influence of the macroscopic description, microscopic

Figure 1. Topology of the Bayesian information retrieval model



description, and final diagnosis, respectively. These values can be assigned by the physicians with the restriction that the sum $\alpha + \beta + \delta$ must be 1. This means, for example, that we can choose $\alpha = \beta = \delta = 1/3$, so we decide that every subordinate document has the same influence when calculating the probability of relevance for a document in general. Another example is to choose $\alpha = \beta = 1/4$ and $\delta = 1/2$, if we want the final diagnosis to obtain a higher influence by the calculation of the probability of relevance for a document in general.

Inference and Retrieval

Given a query Q submitted to our system, the retrieval process starts by placing the evidences in the term *subnetwork*—the state of each term T_{iQ} belonging to Q is fixed to t_{iQ} (relevant). The inference process is then run, obtaining, for each document D_j, its probability of relevance, given that the terms in the query are also relevant, $p(d_j \mid Q)$. Finally, documents are sorted in decreasing order of probability and returned to the user.

We should mention the fact that the Bayesian network contains thousand of nodes, many of which have a great number of parents. In addition, although the network topology is relatively simple, it contains cycles. Therefore, general-purpose propagation algorithms cannot be applied for reasons of efficiency. We therefore propose the use of a specific inference process (de Campos et al., 2003), which is designed to take advantage of both the topology of the network and the kind of probability function used at document nodes, but ensuring that the results are the same as those obtained using exact propagation in the entire network. The final probability of relevance for a document, therefore, is computed using the following equations:

$$p(d_k \mid Q) = \alpha \cdot p(d_{mk} \mid Q) + \beta \cdot p(d_{\mu k} \mid Q) + \delta \cdot p(d_{fk} \mid Q)$$

where $p(d_{\bullet k} \mid Q)$ can be computed as follows:

$$p(d_{\bullet j} \mid Q) = \sum_{T_i \in D_{\bullet j} \cap Q} w_{ij} + (1/M) \sum_{T_i \in D_{\bullet j} \setminus Q} w_{ij} = (1/\mathrm{M}) \sum_{T_i \in D_{\bullet j}} w_{ij} + (M-1)/M \sum_{T_i \in D_{\bullet j} \cap Q} w_{ij}$$

FUTURE TRENDS

Because of the excellent features offered by Bayesian networks for representing relationships between variables and their strengths, as well as their efficient inference mechanisms, these probabilistic models will be used in many different areas of IR.

Following the subject of this article (i.e., dealing with structured documents), one interesting line of research would be the introduction of decisions in the inference process. Instead of returning a ranking of documents, it might be very useful to give the user only those parts of the document that might be relevant, instead of the whole document. A first attempt has been made by de Campos, et al. (2004) using influence diagrams. This field is relatively new and is an open and promising research line. On the grounds of the basic methodology proposed in this article, an intuitive step in this line of work would be to open the field of research to the area of recommendation systems, where Bayesian networks also can perform well.

The Web is also a challenging context. As well as the large number of existing Web pages, we must consider the hyperlinks between these. A good treatment of these links by means of a suitable representation through arcs in a Bayesian network and by means of the conditioned probability distributions, which should include the positive or negative influences regarding the relevance of the Web page that is pointed to, should help improve retrieval effectiveness. Finally, another interesting point would be not to consider index terms independent among them, but to take into account relationships, captured by means of data mining techniques with Bayesian networks.

CONCLUSION

In this article, we have presented a retrieval model to deal with medical records from a pathological anatomy service represented in XML. The retrieval model operates in two stages, given a query: the first one employs an XPath query to retrieve XML documents, and the second, using Bayesian networks, computes a probability of relevance using IR techniques on the free text tags from records obtained in the previous step. This model ensures not only an accurate representation of the structure of the record collection, but also a fast mechanism to retrieve relevant records given a query.

ACKNOWLEDGMENT

(a) This work has been jointly supported by the Spanish Fondo de Investigación Sanitaria and Consejería de Salud de la Junta de Andalucía, under Projects PI021147 and 177/02, respectively; (b) we would like to thank Armin Stoll for his collaboration with the development of the software implementing the model presented in this article.

REFERENCES

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval*. Addison-Wesley.

Calado, P., Ribeiro, B., Ziviani, N., Moura, E., & Silva, I. (2001). Local versus global link information in the Web. *ACM Transactions on Information Systems, 21*(1), 42-63.

Chiaramella, Y. (2001). Information retrieval and structured documents. *Lectures Notes in Computer Science, 1980*, 291-314.

Crestani, F., de Campos, L.M., Fernández-Luna, J., & Huete, J.F. (2003a). A multi-layered Bayesian network model for structured document retrieval. *Lecture Notes in Artificial Intelligence, 2711*, 74-86.

Crestani, F., de Campos, L.M., Fernández-Luna, J., & Huete, J.F. (2003b). Ranking structured documents using utility theory in the Bayesian network retrieval model. *Lecture Notes in Computer Science, 2857*, 168-182.

Crestani, F., Lalmas, M., van Rijsbergen, C.J., & Campbell, L. (1998). Is this document relevant? Probably: A survey of probabilistic models in information retrieval. *Computing Survey, 30*(4). 528-552.

de Campos, L.M., Fernández-Luna, J., & Huete, J.F. (2003a). An information retrieval model based on simple Bayesian networks. *International Journal of Intelligent Systems, 18*, 251-265.

de Campos, L.M., Fernández-Luna, J., & Huete, J.F. (2003b). Implementing relevance feedback in the Bayesian network retrieval model. *Journal of the American Society for Information Science and Technology, 54*(4), 302-313.

de Campos, L.M., Fernández-Luna, J., & Huete, J.F. (2003c). The BNR model: Foundations and performance of a Bayesian network-based retrieval model. *International Journal of Approximate Reasoning, 34*, 265-285.

de Campos, L.M., Fernández-Luna, J., & Huete, J.F. (2003d). Improving the efficiency of the Bayesian network retrieval model by reducing relationships between terms. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 11*, 101-116.

Dumais, S.T., Platt, J., Hecherman, D., & Sahami, M. (1998). Inductive learning algorithms and representations for text categorization. *Proceedings of the ACM International Conference on Information and Knowledge Management.*

Graves, A., & Lalmas, M. (2002). Video Retrieval using an MPEG-7 based inference network. *Proceedings of the 25th ACM-SIGIR Conference.*

Jensen, F.V. (2001). *Bayesian networks and decision graphs*. Springer Verlag.

Ludovic, D., & Gallinari, P. (2003). A belief network-based generative model for structured documents. An application to the XML categorization. *Lecture Notes in Computer Science, 2734*, 328-342.

Myaeng, S.H., Jang, D.H., Kim, M.S., & Zhoo, Z.C. (1998). A flexible model for retrieval of SGML documents. *Proceedings of the 21th ACM—SIGIR Conference*.

Piwowarski, B., & Gallinari P. (2002). A Bayesian network model for page retrieval in a hierarchically structured collection. *Proceedings of the XML Workshop—25th ACM-SIGIR Conference*.

Reis, I. (2000). *Bayesian networks for information retrieval* [doctoral thesis]. Universidad Federal de Minas Gerais.

Tsikrika, T., & Lalmas, M. (2002). Combining Web document representations in a Bayesian inference network model using link and content-based evidence. *Proceedings of the 24th European Colloquium on Information Retrieval Research*.

Turtle, H.R., & Croft, W.B. (1991). Evaluation of an inference network-based retrieval model. *Information Systems, 9*(3), 187-222.

Turtle, H.R., & Croft, W.B. (1997). Uncertainty in information systems. In *Uncertainty management in information system: From needs to solutions* (pp. 189-224). Kluver.

Wong, S.K.M., & Butz, C.J. (2000). A Bayesian approach to user profiling in information retrieval. *Technology Letters, 4*(1), 50-56.

KEY TERMS

Bayesian Network: A directed acyclic graph where the nodes represent random variables and arcs represent the relationships between them. Their strength is represented by means of conditional probability distributions stored in the nodes.

Information Retrieval: A research field that deals with the representation, storage, organization, and accessing of information items.

Probability Distribution: A function that assigns a probability to each value that a random variable can take, fulfilling the Kolmogorov's axioms.

Recommendation System: Software that, given preferences expresses by a user, select those choices, from a range of them, that better satisfy these user's preferences.

Structured Document: A textual representation of any object, whose content could be organized around a well-defined structure.

XML: Acronym for Extensible Markup Language. A meta-language directly derived from SGML but designed for Web documents. It allows the structuring of information and transmission between applications and between organizations.

XPath: A language designed to access the different elements of an XML document.

ENDNOTE

[1] The notation T_i (D_j, respectively) refers to both the term (document, respectively) and its associated variable and node.

This work was previously published in Encyclopedia of Data Warehousing and Mining, edited by J. Wang, pp. 960-964, copyright 2005 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 7.31
Bayesian Network Approach to Estimate Gene Networks

Seiya Imoto
University of Tokyo, Japan

Satoru Miyano
University of Tokyo, Japan

ABSTRACT

In cells, genes interact with each other and this system can be viewed as directed graphs. A gene network is a graphical representation of transcriptional relations between genes and the problem of estimation of gene networks from genome-wide data, such as DNA microarray gene expression data, is one of the important issues in bioinformatics and systems biology. Here, we present a statistical method based on Bayesian networks to estimate gene networks from microarray data and other biological data. Because microarray data are measured as continuous variables and the relationship between genes are usually nonlinear, we combine Bayesian networks and nonparametric regression to handle continuous variables and nonlinear relations. Most parts of gene networks are still unknown, and we need to estimate them from observational data. This problem is equivalent to the structural learning of Bayesian networks, and we solve it from a Bayes approach. The main difficulty of gene network estimation is due to the number of genes involved in the network. Therefore, it leads to model overfitting to the observational data like microarray data. Hence, a combination of various kinds of biological data is a key technique to estimate accurate gene networks. We show a general framework to combine microarray data and other biological information to estimate gene networks.

INTRODUCTION

The microarray technology has produced a huge amount of gene expression data under various conditions such as gene knock-down, overexpression, experimental stressors, transformation, exposure to a chemical compound, and so forth. Along with this new data production, there have been considerable attempts to infer gene networks from such gene expression profile data, and several computational methods have been proposed together with gene network models such as Boolean networks, differential equation models, and Bayesian networks.

A Bayesian network is an effective method in modeling phenomena through the joint distribution of a large number of random variables. In recent years, some interesting works have been established in constructing gene networks from microarray gene expression data by using Bayesian networks. Friedman and Goldszmidt (1998) discretized the expression values and assumed multinomial distributions as the candidate statistical models. Pe'er, Regev, Elidan, and Friedman (2001) investigated the threshold value for discretizing. On the other hand, Friedman, Linial, Nachman, and Pe'er (2000) pointed out that the discretizing probably loses information of the data. In addition, the number of discretizing values and the thresholds are unknown parameters, which have to be estimated from the data. The resulted network strongly depends on their values. Then Friedman et al. (2000) considered fitting linear regression models, which analyze the data in the continuous variables (see also Heckerman & Geiger, 1995). However, the assumption that the parent genes depend linearly on the objective gene is not always guaranteed. Imoto et al. (2002) proposed the use of nonparametric additive regression models (see also Green & Silvermn, 1994; Hastie & Tibshirani, 1990) for capturing not only linear dependencies but also nonlinear structures between genes. In this chapter, we introduce a method for constructing the gene network by using Bayesian networks and the nonparametric regression, which is more suitable for estimating gene networks from microarray gene expression data than discrete type Bayesian networks.

Once we set the graph, we have to evaluate its goodness or closeness to the true graph, which is usually unknown. Hence, the construction of a suitable criterion becomes the center of attention of statistical genetic network modeling. Friedman and Goldszmidt (1998), used the BDe criterion, which was originally derived by Cooper and Herskovits (1992) for choosing a graph (see also Heckerman, Geiger, & Chickering, 1995). The BDe criterion only evaluates the Bayesian network model based on the multinomial distributions and Dirichlet priors. However, Friedman and Goldszmidt (1998) kept the unknown hyperparameters in Dirichlet priors and we only set up the values experimentally. We investigate the graph selection problem as a statistical model selection or evaluation problem and theoretically derive a new criterion for choosing a graph using the Bayes approach (see Berger, 1985). The proposed criterion automatically optimizes all parameters in the model and gives the optimal graph when we can score all candidate graphs.

The problem of finding an optimal Bayesian network is known to be NP-hard. The brute force method employing all computing resources in the world would even require time exceeding the lifetime of the solar system for finding an optimal Bayesian network of 30 genes from 100 microarray datasets. Our approach has made it possible to find optimal and near optimal Bayesian networks with respect to the score of the network in a reasonable time and has provided an evidence of the biological rationality in this computational approach (Ott & Miyano, 2003; Ott, Imoto, & Miyano, 2004; Ott, Hansen, Kim, & Miyano, 2005). For larger networks, we carefully employ the greedy hill-climbing algorithm for finding better gene networks.

The main drawback for the gene network construction from microarray data is that, while the gene network contains a large number of genes, the information contained in gene expression data is limited by the number of microarrays, their quality, the experimental design, noise, and measurement errors. Therefore, estimated gene networks contain some incorrect gene regulations, which cannot be evaluated from a biological viewpoint. In particular, it is difficult to determine the direction of gene regulation using gene expression data only. Hence, the use of biological knowledge, including protein-protein and protein-DNA interactions (Bader, Donaldson, Wolting, Ouellette, Pawson, & Hogue, 2001; Ideker, Ozier, Schwikowski, & Siegel, 2002; Ito, Chiba, Ozawa, Yoshida, Hattori, & Sakaki, 2001) sequences of the binding site of the genes controlled by transcription regulators (Lee, Rinaldi, Robert, Odom, Bar-Joseph, Gerber et al., 2002), literature, and so forth, are considered to be a key for microarray data analysis. The use of biological knowledge has previously received considerable attention for extracting more information from microarray data (Bannai, Inenaga, Shinohara, Takeda, & Miyano, 2002; Bussemaker, Li, & Siggia, 2001; Hartemink, Gifford, Jaakkola, & Young, 2002; Masys, 2001; Ong, Glasner, & Page, 2002; Pilpel, Sudarsanam, & Church, 2001; Segal, Barash, Simon, Friedman, & Koller, 2002; Segal et al., 2003a, Segal et al., 2003b). In this chapter, we also provide a general framework for combining microarray data and biological knowledge aimed at estimating a gene network by using a Bayesian network model.

We show the effectiveness of our approach by the analysis of *Saccharomyces cerevisiae* gene expression data obtained by singly disrupting 100 genes, which are almost all transcription factors together with binding site information. We also apply the method to the estimation of human endothelial cell gene networks by using 270 novel gene knock-downs and time-course of gene expression following treatment with the drug fenofibrate.

MODELING OF GENE NETWORKS

In gene network estimation, we regard a gene as a random variable representing the abundance of a specific RNA species, shown as a node in a graph, and the interaction between genes is represented by the direct edge between nodes. In this section, we show our basic model for constructing gene networks from microarray gene expression data. More precisely, we introduce a combination of Bayesian network and nonparametric regression to model gene networks.

Bayesian Networks with Continuous Variables

Suppose that $\{X_1, ..., X_p\}$ is the set of p random variables, and that the dependency among these p random variables is shown as a directed graph G. When the random variables X_j take discrete values, we can decompose the joint probability of $\{X_1, ..., X_p\}$ into the product of conditional probabilities as:

$$\Pr(X_1, ..., X_p) = \prod_{j=1}^{p} \Pr(X_j \mid Pa(X_j)) \qquad (1)$$

where $Pa(X_j)$ is the set of random variables corresponding to the direct parents of X_j in G. Because gene expression data take continuous variables, some discretization methods are required for using the Bayesian networks based on the discrete random variables described above. However, the discretization lead to information loss, and the number of categories and the threshold values are parameters to be optimized. To avoid the discretization, we rewrite the decomposition of joint probability given in (1) by using densities.

Suppose that we have the observational data $\boldsymbol{X} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)'$ of the set of p random variables $\{X_1, ..., X_p\}$, where $\boldsymbol{x}_i = (x_{i1}, ..., x_{ip})'$ and x_{ij} is the gene expression value of jth gene from ith microarray. That is, we measure p genes' expression values by n microarrays. We then rewrite the decomposition (1) by:

$$f(\boldsymbol{X} \mid \boldsymbol{\Theta}, G) = \prod_{i=1}^{n}\prod_{j=1}^{p} f_j(x_{ij} \mid \boldsymbol{pa}_{ij}, \boldsymbol{\theta}_j) \qquad (2)$$

where $\boldsymbol{\Theta} = (\boldsymbol{\theta}'_1, ..., \boldsymbol{\theta}'_p)'$ is a parameter vector, and $\boldsymbol{pa}_{ij}$ is the expression value vector of $Pa(X_j)$ measured by ith microarray. Hence, the construction of the graph G is equivalent to model the conditional probabilities f_j ($j = 1, ..., p$), that is essentially the same as the regression problem. A possible and simple solution of the construction of conditional density is given by using the linear regression as:

$$x_{ij} = \beta_0 + \beta_1 p_{i,1}^{(j)} + \cdots + \beta_{q_j} p_{i,q_j}^{(j)} + \varepsilon_{ij}$$

where β_j (j=1, ..., q_j) are parameters, ε_{ij} (i = 1, ..., n) are independently and normally distributed with mean 0 and variance σ^2_j, and $\boldsymbol{pa}_{ij} = (p_{i,1}^{(j)}, ..., p_{i,q_j}^{(j)})'$ with $q_j = |Pa(X_j)|$. This model assumes the relationships between genes are linear, and it is unsuitable to extract effective information from the data with complex structure. Also, the assumption that the parent genes depend linearly on the objective gene is not always guaranteed. To capture even nonlinear dependencies, (Imoto, Goto, & Miyano, 2002) proposed the use of the nonparametric additive regression model (Hastie & Tibshirani, 1990) of the form:

$$x_{ij} = m_{j1}(p_{i,1}^{(j)}) + \cdots + m_{jq_j}(p_{i,q_j}^{(j)}) + \varepsilon_{ij} \qquad (3)$$

where m_{jk} $(\cdot)$(k = 1, ..., q_j) are smooth functions from $\Re$ to $\Re$. We construct m_{jk} $(\cdot)$ by the basis function expansion method with B-splines (De Boor, 1978; Imoto et al., 2003):

$$m_{jk}(p) = \sum_{s=1}^{M_{jk}} \gamma_{sk}^{(j)} b_{sk}^{(j)}(p)$$

where $\gamma_{sk}^{(j)}$ (s = 1, ..., M_{jk}) are parameters, $\{b_{1k}^{(j)}(\cdot), ..., b_{M_{jk}k}^{(j)}(\cdot)\}$ is the prescribed set of B-splines, and M_{jk} is the number of B-splines. We then have the Bayesian network and nonparametric regression model:

$$f_j(x_{ij} \mid \boldsymbol{pa}_{ij}, \boldsymbol{\theta}_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left[-\frac{\{x_{ij} - \sum_k \sum_s \gamma_{sk}^{(j)} b_{sk}^{(j)}(p_{ik}^{(j)})\}^2}{2\sigma_j^2}\right]$$

Note that the Bayesian network model based on the linear regression is included in this model as a special case. To extend from the additive regression model to a general regression, (Imoto, Higuchi, Kim, Jeong, & Miyano, 2005) proposed a Bayesian network with moving boxcel median filtering to capture such interactions.

Information Criterion for Structural Learning

Once we set a graph, the statistical model based on the Bayesian network and nonparametric regression can be constructed and estimated by a suitable procedure. However, the problem that still remains to be solved is how we can choose the optimal graph which gives a best approximation of the system underlying the data. Notice that we cannot use the likelihood function as a model selection criterion, because the value of likelihood becomes large in a more complicated model. Hence, we need to consider the statistical approach based on the generalized or predictive error, Kullback-Leibler information, Bayes approach, and so forth (see e.g., Akaike, 1973; Burnham & Anderson 1998; Konishi, 1996; Konishi & Kitagawa, 1996, for the statistical model selection problem). In this section, we construct a criterion for evaluating a graph based on our model from Bayes approach.

The posterior probability of the graph $p(G \mid \boldsymbol{X})$ is given by:

$$p(G|\boldsymbol{X}) = \frac{p(G)\,p(\boldsymbol{X}|G)}{p(\boldsymbol{X})} \propto p(G)\,p(\boldsymbol{X}|G)$$

where $p(G)$ is the prior probability of the graph, $p(\boldsymbol{X} \mid G)$ is the likelihood of the data $\boldsymbol{X}$ conditional on G, and $p(\boldsymbol{X})$ is the normalizing constant and does not depend on the selection of G. The likelihood $p(\boldsymbol{X} \mid G)$ is computed by the marginization:

$$p(\boldsymbol{X} \mid G) = \int f(\boldsymbol{X} \mid \boldsymbol{\Theta}, G)\, p(\boldsymbol{\Theta} \mid \boldsymbol{\lambda}, G)\, d\boldsymbol{\Theta} \qquad (4)$$

where $p(\boldsymbol{\Theta} \mid \boldsymbol{\lambda}, G)$ is the prior distribution on the parameter $\boldsymbol{\Theta}$ specified by the hyperparameter $\boldsymbol{\lambda}$. We now suppose the decomposition:

$$p(\boldsymbol{\Theta} \mid \boldsymbol{\lambda}, G) \prod_{j=1}^{p} p_j(\boldsymbol{\theta}_j \mid \boldsymbol{\lambda}_j)$$

We also assume that the prior distribution on the parameter vector $\boldsymbol{\theta}_j$ is decomposed as:

$$p_j(\boldsymbol{\theta}_j \mid \boldsymbol{\lambda}_j) = \prod_{k=1}^{q_j} p_{jk}\,(\boldsymbol{\gamma}_{jk} \mid \lambda_{jk})$$

with $\boldsymbol{\gamma}_{jk} = (\gamma^{(j)}_{1k}, ..., \gamma^{(j)}_{M_{jk}k})'$. Each prior distribution $p_{jk}\,(\boldsymbol{\gamma}_{jk} \mid \lambda_{jk})$ is a singular M_{jk} variate normal distribution given by:

$$p_{jk}(\boldsymbol{\gamma}_{jk} | \lambda_{jk}) = \left(\frac{2\pi}{n\lambda_{jk}}\right)^{-(M_{jk}-2)/2} \left|K_{jk}\right|_+^{1/2} \exp\left(-\frac{n\lambda_{jk}}{2}\,\boldsymbol{\gamma}'_{jk} K_{jk}\,\boldsymbol{\gamma}_{jk}\right) \qquad (5)$$

where λ_{jk} is a hyper parameter, K_{jk} is an $M_{jk} \times M_{jk}$ matrix that holds:

$$\boldsymbol{\gamma}'_{jk} \boldsymbol{K}_{jk}\, \boldsymbol{\gamma}_{jk} = \sum_{l=3}^{M_{jk}} (\gamma^{(j)}_{lk} - 2\gamma^{(j)}_{l-1k} + 2\gamma^{(j)}_{l-2k})^2$$

and $|K_{jk}|_+$ is the product of $M_{jk} - 2$ nonzero eigenvalues of K_{jk}. Under the decomposition of the prior of parameters $p(\boldsymbol{\Theta} \mid \boldsymbol{\lambda}, G)$ and the decomposition of the joint density by the Bayesian network given in (2), the marginal likelihood (4) can be rewritten as:

$$p(\boldsymbol{X} \mid G) = \int f\,(\boldsymbol{X} \mid \boldsymbol{\Theta}, G)\,p(\boldsymbol{\Theta} \mid \boldsymbol{\lambda}, G)\,d\boldsymbol{\Theta} = \prod_{j=1}^{p}\left\{\prod_{i=1}^{n} f_j(x_{ij} \mid \boldsymbol{pa}_{ij}, \boldsymbol{\theta}_j)\, p_j(\boldsymbol{\theta}_j \mid \boldsymbol{\lambda}_j)\,d\boldsymbol{\theta}_j\right\}$$

$$\equiv \prod_{j=1}^{p} p_j(\text{sub}.G_j \mid \boldsymbol{X}),$$

where sub.G_j is a subgraph of G consisting of jth gene and its direct parents as the node and edges from these parents to jth gene.

The high-dimensional integral in $p_j(\text{sub}.G_j \mid \boldsymbol{X})$ can be asymptotically approximated with an analytical form by the Laplace approximation:

$$p_j(\text{sub.G}_j \mid \boldsymbol{X}) = \int \exp\{n l_\lambda\,(\boldsymbol{\theta}_j \mid \boldsymbol{X})\}\,d\boldsymbol{\theta}_j$$
$$= \frac{(2\pi/n)^{r_j/2}}{\left|J_\lambda(\hat{\boldsymbol{\theta}}_j)\right|^{1/2}} \exp\{n l_\lambda(\hat{\boldsymbol{\theta}}_j | X)\}\{1 + O_p(n^{-1})\} \qquad (6)$$

where r_j is the dimension of $\boldsymbol{\theta}_j$,

$$l_\lambda(\boldsymbol{\theta}_j \mid \boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n} \log f_j(x_{ij} \mid \boldsymbol{pa}_{ij}, \boldsymbol{\theta}_j) + \frac{1}{n}\log p_j(\boldsymbol{\theta}_j \mid \boldsymbol{\lambda}_j),$$

$$J_\lambda(\boldsymbol{\theta}_j) = -\frac{\partial l_\lambda(\boldsymbol{\theta}_j | \boldsymbol{X})}{\partial \boldsymbol{\theta}_j\,\partial \boldsymbol{\theta}'_j}$$

and $\hat{\theta}_j$ is the mode of $l_\lambda(\boldsymbol{\theta}_j | \boldsymbol{X})$. We also decompose the prior probability of the graph $p(G)$ as $\prod_j p(\text{sub}.G_j)$. By taking minus twice logarithm of $p(\text{sub}.G_j) \times (6)$, we then define:

$$\textbf{BNRC}_j = -2\log p(\text{sub}.G_j) - r_j \log(2\pi/n) + \log|J_\lambda(\hat{\theta}_j)| - 2n l_\lambda(\hat{\theta}_j \mid \boldsymbol{X})$$
$$= 2(q_j - 1) - \left(\sum_{k=1}^{q_j} M_{jk} + 1\right)\log\left(\frac{2\pi}{n}\right) + n\log(2\pi\hat{\sigma}_j^2) + n$$
$$+ \sum_{k=1}^{q_j}\{\log|\boldsymbol{\Lambda}_{jk}| - M_{jk}\log(n\hat{\sigma}_j^2)\} - \log(2\hat{\sigma}_j^2)$$
$$+ \sum_{k=1}^{q_j}\left\{(M_{jk} - 2)\log\left(\frac{2\pi\hat{\sigma}_j^2}{n\beta_{jk}}\right) - \log|K_{jk}| + \frac{n\beta_{jk}}{\hat{\sigma}_j^2}\hat{\boldsymbol{\gamma}}'_{jk} K_{jk}\hat{\boldsymbol{\gamma}}_{jk}\right\}$$

with

$$\boldsymbol{\Lambda}_{jk} = \boldsymbol{B}'_{jk}\boldsymbol{B}_{jk} + n\beta_{jk}\boldsymbol{K}_{jk}, \quad \beta_{jk} = \hat{\sigma}_j^2\lambda_{jk},$$
$$\boldsymbol{B}_{jk} = (\boldsymbol{b}_{jk}(p^{(j)}_{1k}), ..., \boldsymbol{b}_{jk}(p^{(j)}_{nk}))',$$
$$\boldsymbol{b}_{jk}(p^{(j)}_{ik}) = (b^{(j)}_{1k}(p^{(j)}_{ik}), ..., b^{(j)}_{M_{jk}k}(p^{(j)}_{ik}))',$$
$$\hat{\sigma}_j^2 = \frac{1}{n}\sum_{i=1}^{n}\left\{x_{ij} - \sum_{k=1}^{q_j}\hat{\boldsymbol{\gamma}}'_{jk}\,\boldsymbol{b}_{jk}(p^{(j)}_{ik})\right\}.$$

Here we approximate the Hessian matrix by:

$$l_\lambda(\boldsymbol{\theta}_j | \boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n}\log f_j(x_{ij} | \boldsymbol{pa}_{ij}, \boldsymbol{\theta}_j) + \frac{1}{n}\log p_j(\boldsymbol{\theta}_j | \boldsymbol{\lambda}_j),$$

$$\log\left|J_\lambda(\hat{\boldsymbol{\theta}}_j)\right| \approx \sum_{k=1}^{q_j}\log\left|-\frac{\partial^2 l_\lambda(\hat{\boldsymbol{\theta}}_j | \boldsymbol{X})}{\partial\boldsymbol{\gamma}_{jk}\partial\boldsymbol{\gamma}'_{jk}}\right| + \log\left|-\frac{\partial^2 l_\lambda(\hat{\boldsymbol{\theta}}_j | \boldsymbol{X})}{\partial(\sigma_j^2)^2}\right|$$

Then, we define an information criterion, named BNRC, for the Bayesian network and nonparametric regression model in order to select a graph structure:

$$\mathrm{BNRC}(G) = \sum_{j=1}^{p} \mathrm{BNRC}_j$$

The optimal graph is chosen such that the criterion BNRC is minimal.

Algorithms

Learning Parameters

Consider the nonparametric regression model defined in (3). The estimate $\hat{\boldsymbol{\theta}}_j$ is a mode of $l_\lambda(\boldsymbol{\theta}_j \mid \boldsymbol{X})$ and depends on the hyperparameters. In fact, the hyperparameter plays an essential role in estimating the smoothed curve.

In our model, we construct the nonparametric regression model by 20 *B*-splines. We confirmed that the differences of the smoothed estimates against the various number of the basis functions cannot be found visually. Because when we use a somewhat large number of the basis functions, the hyperparameters control the smoothness of the fitted curves. The algorithm for obtaining $\hat{\boldsymbol{\theta}}_j$ and selection of β_{jk} is summarized as follows:

Step 1: Set j and initialize: $\boldsymbol{\gamma}_{jk} = \boldsymbol{0}(k = 1, ..., q_j)$.

Step 2: Set k and find the optimal β_{jk} by repeating the following steps:

A. Compute:

$$\boldsymbol{\gamma}_{jk} = (\boldsymbol{B}'_{jk} \boldsymbol{B}_{jk} + n\beta_{jk} \boldsymbol{K}_{jk})^{-1} \boldsymbol{B}'_{jk}(\boldsymbol{x}_{(j)} - \sum_{k'\neq k} \boldsymbol{B}_{jk'} \boldsymbol{\gamma}_{jk'}),$$

for fixed β_{jk}. Here $\boldsymbol{x}_{(j)} = (x_{1j}, ..., x_{nj})'$

B. Evaluate: Repeat Step 2-A against the candidate value of β_{jk}, and choose the optimal value of β_{jk}, which minimizes the BNRC_j.

Step 3: Convergence: Repeat Step 2 for $k = 1, ..., q_j, 1, ..., q_j, 1,...$ until a suitable convergence criterion is satisfied.

The mode $\hat{\sigma}_j^2$ is given by:

$$\hat{\sigma}_j^2 = \frac{1}{n}\left\| \boldsymbol{x}_{(j)} - \sum_{k=1}^{q_j} \boldsymbol{B}_{jk}\hat{\gamma}_{jk} \right\|^2$$

Learning Graph Structure

Finding optimal Bayesian networks is computationally hard. Potentially, we need to search the space of directed acyclic graphs of p vertices whose size c_p is approximately:

$$c_p = \frac{p! \times 2^{p(p-1)/2}}{r \times z^p}$$

where $r \sim 0.57436$ and $z \sim 1.17881$ (Robinson, 1973). From this formula we can see that there are roughly 2.34×10^{72} networks with 20 vertices and 2.71×10^{158} for 30 vertices. This complexity does not allow us any brute force approach even with a supercomputer system. Furthermore, without obtaining the optimal Bayesian networks, we cannot have any right insight that the Bayesian network model can really extract biologically meaningful regulatory information from microarray gene expression data. Thus, we are faced with two issues. The first issue is how to cope with this complexity and the second is the search for optimal Bayesian networks and their biological evaluation.

Greedy Heuristics for Searching Bayesian Networks

Some heuristic approaches have been employed for this search problem, for example, greedy algorithms (Friedman et al., 2000; Heckerman et al., 1995; Imoto et al., 2002), simulated annealing (Hartemink et al., 2002), and genetic algorithms (Somereren et al., 2002).

The greedy hill-climbing algorithm due to Heckerman et al. (1995) is shown below as a typical example, where n is the number of repeats. The greedy hill-climbing algorithm assumes a score function for solutions. It starts from some initial solution and successively improves the solution by selecting the modification from the space of possible modifications which yields the best score. When no improvement is found, then the algorithm terminates with the currently best solution. Some ideas should be employed for the choice of the initial solution and for the choice of the space of possible modifications. Biologically reasonable locally optimal Bayesian networks of several hundred genes are reported (Imoto et al., 2002; Imoto et al., 2003; Nariai, Kim, Imoto, & Miyano, 2004; Tamada, Kim, Bannai, Imoto, Tashiro, Kuhara, & Miyano, 2003). The following algorithm is a greedy hill-climbing algorithm provided by Heckerman et al. (1995):

Step 1: Initialize the network as the empty network.
Step 2: Randomly select a permutation π: $\{1, ..., |X|\} \rightarrow X$, where X is the set of vertices.
Step 3: For all $i = 1, ..., p$ do the following two steps:
- **A.** Compute the changes of the score when adding a new parent for $\pi(i)$ or removing or reversing the edge of a parent gene of $\pi(i)$.
- **B.** Select the modification among the modifications which improve the score most without violating the acyclicity condition.

Step 4: Repeat Step 3 until their score does not improve.
Step 5: Repeat Steps 1 through Step 4 for T times and return the best solution found in these iterations.

Search Algorithm for Optimal Bayesian Networks

BNRC score can be decomposed to an additive form $\mathrm{BNRC}(G) = \sum_{j=1}^{p} \mathrm{BNRC}_j$ as shown in Section 2.2. We will formulate this optimization problem in an abstract way: For a finite set X (of genes), we call a function s: $X \times 2^X \rightarrow R$ a score function for X. Then for a DAG $G = (X, E)$ we define the score of X by $\mathrm{score}(G) = \Sigma_{g \in X}\, s(g, Pa(g))$. The problem is to find the best network $G = (X, E)$ which attains the optimal score. In the case of the BNRC score, the problem is defined as a minimization problem. Furthermore, it is noted in Ott et al. (2004) that the case for the MDL score (Friedman & Goldszmidt, 1998) is also formulated as a minimization problem while the case for the BDe score (Cooper & Herskovits, 1992; Friedman & Goldszmidt, 1998; Heckerman et al., 1995) is defined as a maximization problem.

Ott et al. (2004), have devised an algorithm which can find optimal Bayesian networks of size up to 35 if a supercomputer such as SUN FIRE 15K with 96CPUs 900MHz each is used. The algorithm decomposes the search space into subspaces and employs the dynamic programming technique for finding the right subspace as well as for determining the optimal solution in the subspace. In order to describe the algorithm, several notations shall be introduced as follows: For a gene g in X and a subset $A \subseteq X$:

$$F(g, A) = \min_{B \subseteq A} s(g, B)$$

gives the optimal choice of parents for g if the parents are selected from A. An order on a subset $A \subseteq X$ is given as a permutation π: $\{1, ..., |A|\} \rightarrow A$. We denote by Π^A the set of all permutations on A. We denote the subnetwork of $G = (X,E)$, restricted to A by $G(A) = (A, E(A))$. For a permutation $\pi \in \Pi^A$, we say that G is *π-linear* if $\pi^{-1}(g) < \pi^{-1}(h)$ holds for all $(g, h) \in E(A)$. The idea of the algorithm is to decompose the set of all DAGs on A into subsets of π-linear DAGs for all $\pi \in \Pi^A$. Then, we divide the problem into (i) to find the subspace of the search space that contains the optimal network and (ii) to find the optimal network within the selected subspace. We denote:

$$Q^A(\pi) = \sum_{g \in A} F(g, \{h \in A \mid \pi^{-1}(h) < \pi^{-1}(g)\})$$

Then we find the best π-linear network for any given permutation by F and Q. The optimal network can be found by finding the optimal permutation which yields the global minimum, which is given by:

$$M(A) = \arg\max_{\pi \in \Pi^A} Q^A(\pi)$$

Then, the whole algorithm derived in Ott et al. (2004) is described as follows:

Step 1: Compute $F(g, \phi) = s(g, \phi)$ for all $g \in X$.
Step 2: For all $g \in X$ and all $A \subseteq X - \{g\}$ with $A \neq \phi$, compute $F(g, A)$ as

$$\min\{s(g, A), \min_{a \in A} F(g, A - \{a\})\}$$

Step 3: Set $M(\phi) = \phi$.
Step 4: For all $A \subseteq X$ with $A \neq \phi$, execute the following steps:
A. Compute

$$g^* = \arg\max_{g \in A}\{F(g, A - \{g\}) + Q^{A-\{g\}}(M(A - \{g\}))\}$$

B. For all $1 \leq i \leq |A|$, set $M(A)(i) \equiv M(A - \{g^*\})(i)$ and $M(A)(|A|) \equiv g^*$.
Step 5: Return $Q^G(M(G))$.

Here we have the following theorem:

Theorem (Ott et al., 2004)

Optimal Bayesian networks can be found using $(|X| / 2 + 1) \times 2^{|X|}$ *dynamic programming steps, where X is a set of genes.*

A rigorous proof is required to show the correctness of this algorithm that can be found in Ott et al. (2004). Furthermore, with some biologically reasonable constraints on the networks, we can obtain a much faster algorithm (Ott & Miyano, 2003). By computing optimal Bayesian networks of small size and evaluating them, it is reported that optimal Bayesian networks are not necessarily biologically optimal. However, by combining optimal to near optimal Bayesian networks thoroughly, we can extract biologically more accurate information from microarray gene expression data (Ott & Miyano, 2003; Ott et al., 2004; Ott et al., 2005).

Combining Biological Knowledge with Microarray Data

Discrete Information

The criterion BNRC(G), introduced in Section 2.2, contains two quantities: the prior probability $p(G)$ of the network, and the marginal likelihood of the data. The marginal likelihood shows the fitness of the model to the microarray data. The biological knowledge can then be added into the prior probability of the network $p(G)$. Let U_{ij} be the interaction energy of the edge from $gene_i$ to $gene_j$ and let U_{ij} be categorized into I values, H_1, ..., H_I, based on biological knowledge. For example, if we know a priori that $gene_i$ regulates $gene_j$, we set $U_{ij} = H_1$. However, if we do not know whether $gene_k$ regulates $gene_l$ or not, we set $U_{kl} = H_2$. We treat the prior information of each edge independently. Note that in $0 < H_1 < H_2$, it is more natural to choose the network with a large number of H_1 edges rather than H_2 edges in the sense of prior information. Our setting, $H_1 < H_2$, gives a higher prior probability to the graph with a lot of H_1 edges than to the graph with a lot of H_2 edges.

The total energy of the network G can then be defined as:

$$E(G) = \sum_{\{i,j\} \in G} U_{ij}$$

where the sum is taken over the existing edges in the network G. Under the Bayesian network framework, the total energy can be decomposed into the sum of the local energies:

$$E(G) = \sum_{j=1}^{p} \sum_{i \in L_j} U_{ij} = \sum_{j=1}^{p} E_j \tag{7}$$

where L_j is the index set of parents of gene$_j$ and $E_j = \Sigma_{i \in L_j} U_{ij}$ is the local energy defined by gene$_j$ and its parents.

The probability of a network G, $p(G)$, is modeled by the Gibbs distribution (Geman & Geman, 1984):

$$p(G) = Z^{-1}\exp\{-\zeta E(G)\}, \tag{8}$$

where ζ (> 0) is a hyperparameter and Z is a normalizing constant called the partition function

$$Z = \sum_{G} \exp\{-\zeta\, E(G)\}$$

Here the sum is taken over the set of possible networks. By replacing ζH_1, ..., ζH_I with ζ_1, ..., ζ_I, respectively, the normalizing constant Z is a function of ζ_1, ..., ζ_I. We call ζ_j an inverse normalized temperature. By substituting (7) into (8), we have:

$$p(G) = Z^{-1} \exp\{-\zeta\, E_j\} = Z^{-1} \prod_{j=1}^{p} \prod_{i \in L_j} \exp\{-\zeta_{\alpha(i,j)}\}$$

with α $(i, j) = k$ for $U_{ij} = H_k$. Hence, by adding biological knowledge into the prior probability of the network, BNRC can be rewritten as:

$$\mathrm{BNRC}(G, \zeta_1, ..., \zeta_I) = 2\log Z + \sum_{j=1}^{p} \left\{ 2 \sum_{i \in Lj} \zeta_{\alpha(i,j)} + \mathrm{BNRC}_j \right\} \tag{9}$$

We can choose an optimal network under the given ζ_1, ..., ζ_I. Also the optimal values of ζ_1, ..., ζ_I are obtained as the minimizer of (9). Therefore, we can represent an algorithm for estimating a gene network from microarray data and biological knowledge as follows:

Step 1: Set the values ζ_1, ..., ζ_I.

Step 2: Estimate a gene network by minimizing BNRC(G) under the given ζ_1, ..., ζ_I.

Step 3: Repeat Step 1 and Step 2 against the candidate values of ζ_1, ..., ζ_I.

Step 4: An optimal gene network is obtained from the candidate networks obtained in Step 3.

In Step 2, we use the greedy hill-climbing algorithm for learning networks.

The computation of normalizing constant, Z, is intractable even for moderately sized gene networks. To avoid this problem, we compute upper and lower bounds of the partial function and use them to choose the optimal value of inverse normalized temperature. An upper bound is obtained by directed graphs that are allowed to contain cyclic graphs. The number of graphs that has b_1, b_2, ..., b_I edges of ζ_1, ζ_2, ..., ζ_I out of a_1, a_2, ..., a_I edges, respectively, is obtained by:

$$S(b_1, ..., b_I) = \prod_{i=1}^{I} \frac{a_i\,!}{b_i\,!(a_i - b_i)!}$$

The upper bound of Z is then:

$$\sum_{b_1, ..., b_I} S(b_1, ..., b_I) \exp\left(-\sum_{i=I}^{I} b_i \zeta_i \right)$$

Thus, the true value of the partition function is not greater than the upper bound. A lower bound is computed by multilevel directed graphs with following assumptions: (A1) There is one top gene and (A2) Genes at the same level have a common direct parent gene. We also consider joined graphs of some multilevel directed graphs satisfying (A1) and (A2). Because the number of possible graphs is much larger than those included in the computation, the true value of the partition function should be greater than the lower bound. Because the optimization of the network structure for fixed ζ_1, ..., ζ_I does not depend on the value of the partition function, our method works well in practice. Of course, when the number of genes is small, we can perform an exhaustive search and compute the partition function completely. However, we think that the development of an effective algorithm to enumerate all possible networks or approximate the partition function is an important problem.

As a related work, Segal et al. (2003a) proposed an interesting method for combining protein-protein interaction data with microarray gene expression data. They modeled protein-protein interaction data based on Markov networks (Kindeman & Snell, 1980) and considered the joint probability of microarray data and protein-protein data for estimating molecular pathways. Although our model is different from their model, their model contains a hyperparameter, denoted by α, that plays a quite similar role of ζ_1 and ζ_2. Similar to our criterion, their joint probability contains the normalizing constant, which is a function of the hyperparameter, α. While we optimize the hyperparameters by our criterion, they did not compute the normalizing constant and chose the value of α heuristically.

Discrete and Continuous Information

Imoto, Higuchi, Goto, Tashiro, Kuhara, and Miyano (2004), proposed a general framework for combining biological knowledge with expression data aimed at estimating more accurate gene networks. In Imoto et al. (2004), the biological knowledge is represented as the binary values, for example, known or unknown, and is used for constructing $p(G)$. In reality, there are, however, various confidence in biological knowledge in practice. Bernard and Hartemink (2005) constructed $p(G)$ using the binding location data (Lee et al., 2002) that is a collection of p-values (continuous information). In this section, we construct $p(G)$ by using multisource information including continuous and discrete prior information.

Let $\boldsymbol{Z}_k$ be the matrix representation of k-th prior information, where (i, j)-th element $z_{ij}^{(k)}$ represents the information of $\text{gene}_i \rightarrow \text{gene}_j$. For example, (1) If we use a prior network G_{prior} for $\boldsymbol{Z}_k$, $z_{ij}^{(k)}$ takes 1 if $e(i, j) \in G_{\text{prior}}$ or 0 for otherwise. Here $e(i, j)$ denotes the direct edge from gene_i to gene_j. (2) By using the gene knock-down data for $\boldsymbol{Z}_k$, $z_{ij}^{(k)}$ represents the value that indicates how gene_j changes by knocking down gene_i. We can use the absolute value of the log-ratio of gene_j for gene_i knock-down data as $z_{ij}^{(k)}$. Using the adjacent matrix $E = (e_{ij})_{1 \le i, j \le p}$ of G, where $e_{ij} = 1$ for $e(i, j) \in G$ or 0 for otherwise, we assume the Bernoulli distribution on e_{ij} having probabilistic function:

$$p(e_{ij}) = \pi_{ij}^{e_{ij}}(1-\pi_{ij})^{e_{ij}}$$

where $\pi_{ij} = \Pr(e_{ij} = 1)$ For constructing π_{ij}, we use the logistic model with linear predictor

$$\eta_{ij} = \sum_{k=1}^{K} w_k (z_{ij}^{(k)} - c_k)$$

as $\pi_{ij} = \{1 + \exp(-\eta_{ij})\}^{-1}$, where w_k and c_k $(k = 1, ..., K)$ are weight and baseline parameters, respectively. We then define a prior probability of the graph based on prior information $\mathbf{Z}_k (k = 1, ..., K)$ by:

$$p(G) = \prod_i \prod_j p(e_{ij})$$

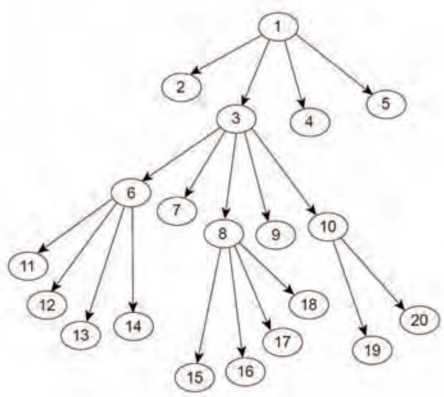
This prior probability of the graph assumes that edges $e(i, j)(i, j = 1, ..., p)$ are independent of each other. In reality, there are several dependencies among e_{ij}'s such as $p(e_{ij} = 1) < p(e_{ij} = 1 \mid e_{ki} = 1)$, and so forth, and we consider that adding such information into $p(G)$ is premature by the quality of such information.

COMPUTATIONAL EXPERIMENTS

Monte Carlo Simulations

Before analyzing real gene expression data, we perform Monte Carlo simulations to examine the properties of the proposed method. We assume an artificial network with 20 nodes shown in Figure 1 (a). The functional relationships between nodes are listed in Figure 1 (b). The data were generated from the artificial network of Figure 1 (a) with the functional structures between nodes shown in Figure 1 (b). Then, the observations of the child variable are generated after transforming the observations of the parent variables to mean 0 and variance 1. A network was rebuilt from simulated data consisting of 50 or 100 observations, which corresponds to 50 or 100 microarrays, because, recently more microarray data have become available and it is often the case that we can use more than 100 microarrays. While at the starting point of the analysis, we have over 6000 genes for yeast, after some pretreatments of the data or using some prior knowledge, the number of target genes is typically less than 50 or so. We consider such a case in this simulation. As for the biological knowledge, we tried the following situations: (**Case 1**) we know some gene regulations (100%, 75%, 50%, or 25% out of 19 edges shown in Figure 1 (a)) and (**Case 2**) we know some gene regulations, but some (1, 2, or 3) incorrect edges are kept in the database. We set {0.5, 1.0} and $\{\zeta_1$, 2.5, 5.0, 7.5, 10.0} as the candidate values of ζ_1 and ζ_2, respectively.

Figure 1. Artificial gene network and functional structures between nodes



(a)

$$g_1 = \varepsilon_1,\quad g_2 = .7g_1 + \varepsilon_2 \quad g_5 = .7g_1 + \varepsilon_5,\quad g_{10} = 1/\{1 + \exp(-4g_3)\} + \varepsilon_{10}$$

$$g_3 = \begin{cases} -1 + \varepsilon_3\ (g_1 \le -.5) \\ g_1 + \varepsilon_3\ (|g_1| < .5) \\ 1 + \varepsilon_3\ (g_1 \ge .5) \end{cases} \qquad g_6 = \begin{cases} .8g_3 + \varepsilon_6\ (g_3 \le -1) \\ (g_3 + 1)^{1.5} + \varepsilon_6\ (-1 < |g_3| < 0) \\ 1 + \varepsilon_6\ (g_3 \ge 1) \end{cases}$$

$$g_4 = \begin{cases} .4g_1 + 1 + \varepsilon_4\ (|g_1| \le .3) \\ (g_1 + 1)^2 + \varepsilon_4\ (|g_1| < .3) \end{cases} \qquad g_8 = \begin{cases} .2g_3 - 1 + \varepsilon_8\ (g_3 \le .2) \\ 1.4g_3 + \varepsilon_8\ (g_3 > .2) \end{cases}$$

$$g_{11} = .7g_6 + \varepsilon_{11},\ g_{14} = .7g_6 + \varepsilon_{14},\ g_{15} = 1/\{1 + \exp(-4g_8)\} + \varepsilon_{15}$$

$$g_9 = \begin{cases} .4g_3 + 1 + \varepsilon_9\ (|g_3| \le .3) \\ (g_3 + 1)^{1.2} + \varepsilon_9\ (|g_3| < .3) \end{cases} \qquad g_{13} = \begin{cases} .4g_6 + 1 + \varepsilon_{13}\ (|g_6| \le .3) \\ (g_6 + 1)^2 + \varepsilon_{13}\ (|g_6| < .3) \end{cases}$$

$$g_{12} = \begin{cases} -1 + \varepsilon_{12}\ (g_6 < -.5) \\ g_6 + \varepsilon_{12}\ (|g_6| \le .5) \\ 1 + \varepsilon_{12}\ (|g_6| > .5) \end{cases} \qquad \begin{array}{l} g_{16} = .8g_8 + \varepsilon_{16} \\ g_{19} = 1/\{1 + \exp(-4g_{10})\} + \varepsilon_{19} \\ g_{20} = 1.1g_{10} + \varepsilon_{20} \end{array}$$

$$g_{17} = \begin{cases} .2g_8 - 1 + \varepsilon_{17}\ (g_8 \le .2) \\ 1.4g_8 + \varepsilon_{17}\ (g_8 > .2) \end{cases} \qquad g_{18} = \begin{cases} .4g_8 + 1\ (|g_8| > .3 \\ (g_8 + 1)^{1.2}\ (g_8 \le .3) \end{cases}$$

(b)

Figure 2 shows two estimated networks: One is estimated by 100 observations (microarrays) alone. We use $\zeta_1 = \zeta_2 = 0.5$, that is, we did not use any prior knowledge (we denote this network by N_0 for convenience). The other is estimated by 100 observations and prior information of 75% gene regulations, that is, we know 14 correct relations out of the all 19 correct edges (we denote this network by N_1). Edges appearing in both networks are colored green, while edges appearing in N_0 or N_1 only are colored blue and red, respectively. By adding prior knowledge, it is clear that we succeeded in reducing the number of false positives. We also find four additional correct relationships. Figure 3 shows the behavior of BNRC, when $\zeta_1 = 0.5$. We find that the optimal value of ζ_2 is 5.0. From the Monte Carlo simulations, we observed that ζ_2 can be selected by using middle values (depicted by a blue line) of upper and lower bounds or upper bounds in practice. For the selection of ζ_1, we use the middle value of the upper and lower bounds of the score of our criterion.

Figure 4 shows the boxplots of the average squared errors (ASEs) that are defined by:

$$ASE = \sum_{i=1}^{100} \sum_{j=1}^{20} (x^*_{ij} - \hat{x}_{ij})^2$$

where x^*_{ij} is the true value of x_{ij}, that is, x^*_{i2} is given by $0.7x_{i1}$, and $\hat{x}_{ij}$ is the estimate of x_{ij} based on the estimated network. Because we repeated the Monte Carlo simulation 1000 times, each boxplot is obtained from 1000 ASEs. Smaller ASE means a more accurate estimated network. From Figure 4, it is clear that by adding prior information we succeeded in reducing the ASE. The distributions of the number of true positives and false positives of the estimate networks are shown in Figure 5. While the estimated networks without prior information contain many correct

Figure 2. An example of resulting networks based on 100 samples. We used $\zeta_1 = 0.5$ *and* $\zeta_2 = 5.0$ *that are selected by our criterion (see Figure 3).*

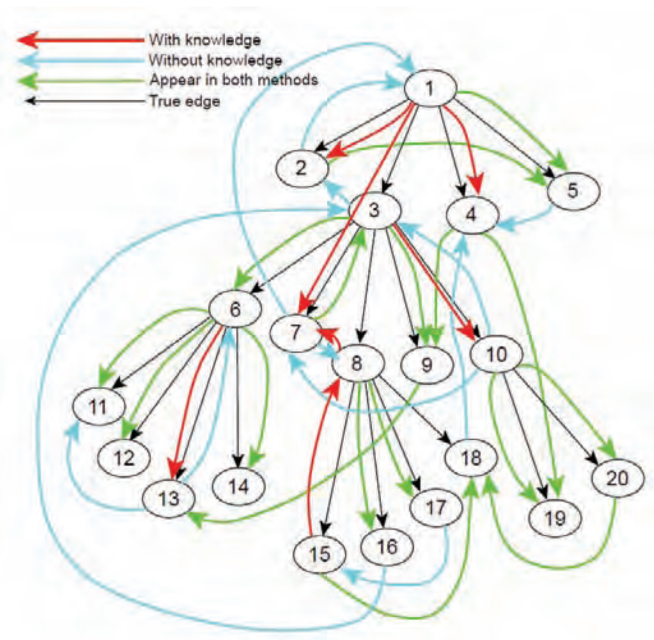


Figure 3. The behavior of BNRC when $\zeta_1 = 0.5$*. We can find out the optimal inverse normalized temperature* ζ_2 *is 5.0.*

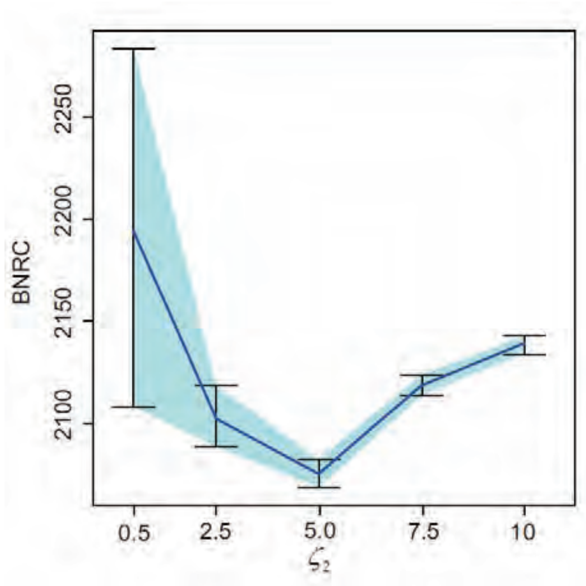
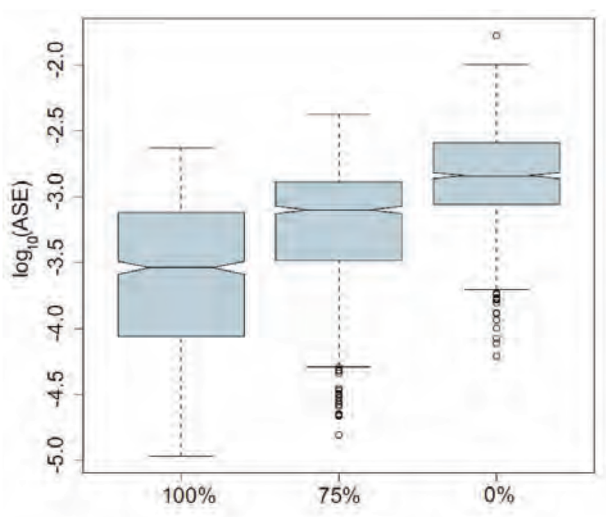


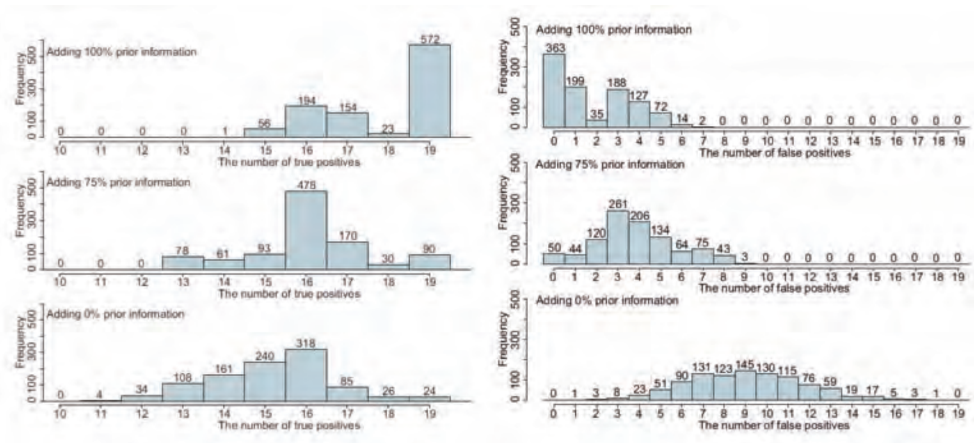
Figure 4. Boxplots of the average squared errors



edges, we observe that the proposed method could reduce the number of false positives even if we added only a part of the true relations.

The results of the Monte Carlo simulations are summarized as follows:

Figure 5. Distribution of the number of true positives and false positives of the estimated networks. The left three figures show the number of true positives when we use 100%, 75%, and 0% prior information, respectively. The right three figures show the number of false positives with respect to 100%, 75%, and 0%. Note that because there are 19 edges in the true network, the maximum number of true positives is 19.



In (**Case 1**), we obtained networks more accurately as long as we add correct knowledge. We observed that the number of false positives decreased drastically. We presume the reason is the nature of directed acyclic graphs. Because a Bayesian network model is a directed acyclic graph, one incorrect estimate may affect the relations in its neighborhood. However, by adding some correct knowledge, we can restrict the search space of the Bayesian network model learning effectively.

In (**Case 2**), the results depend on the type of incorrect knowledge:

1. If we use misdirected relations, for example, $gene_8 \rightarrow gene_3$, as prior knowledge, serious problems occur. Because microarray data to some degree support the misdirected relations, they tend to receive a better criterion score.
2. If we add indirect relations such as $gene_1 \rightarrow gene_8$, we observed that our method controlled the balance between this prior information and microarray data and could decide whether the prior relation is true.
3. If irrelevant relations such as $gene_{20} \rightarrow gene_5$ are added as prior information, our method could reject these prior information because the microarray data do not support these relations.

Example Using Experimental Data of Yeast

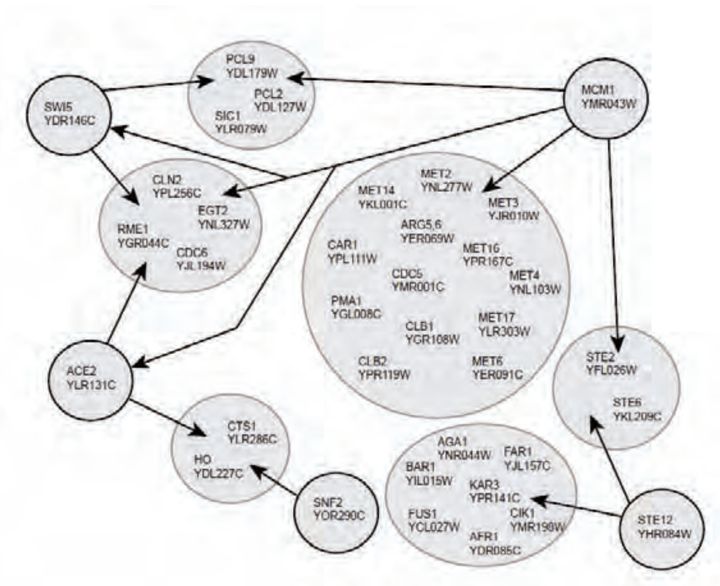
In this section, we demonstrate our method by analyzing *Saccharomyces cerevisiae* gene expression data obtained by disrupting 100 genes, which are almost all transcription factors. We used the BY4741 (*MATa, HIS3D1, LEU2D0, MET15D0, URA3D0*) as the wild type strain and purchased

Table 1. Five transcription factors and their regulating genes

MCM1 :	transcription factor of the MADS box family
	MET14, CDC6, MET2, CDC5, MET6, SIC1, STE6, CLN2, PCL2, STE2, ACE2, MET16,MET3, MET4, CAR1, SWI5, PCL9, CLB1, MET17, EGT2, ARG5,6, PMA1, RME1, CLB2
SWI5 :	transcription factor
	CDC6, SIC1, CLN2, PCL2, PCL9, EGT2, RME1, CTS1, HO
ACE2 :	metallothionein expression activator
	CLN2, EGT2, HO, CTS1, RME1
SNF2 :	component of SWI/SNF global transcription activator complex
	CTS1, HO
STE12 :	transcriptional activator
	STE6, FAR1, KAR3, SST2, FUS1, STE2, BAR1, AGA1, AFR1, CIK1

gene disruptions from Research Genetics, Inc. We focus on five genes, *MCM1*, *SWI5*, *ACE2*, *SNF2*, and *STE12* (see Table 1) and extract genes that are regulated by these 5 genes from the Yeast proteome database. Thus, we construct a prior network, shown in Figure 6, based on the database information. We include the prior network in our Bayesian network estimation method. That is, the purpose of this analysis is to estimate the gene network containing above 36 genes from microarray data together with the prior network. For constructing a Bayesian network with prior

Figure 6. Prior knowledge network. The genes that are in each shadowed circle are regulated by the parent genes.



knowledge, the simplest way is to fix the prior edges and learn the other parts of the network based on the observed data. However, we observed that the score of our criterion, BNRC, of the estimated network learned with fixed prior edges cannot decrease compared with the optimal one. Figure 7 shows the estimated gene network using microarray data only. There are many non-prior edges and many of them are probably false positives. In addition, we find three misdirected relations: "*SWI5* → *MCM1*," "*HO* → *ACE2*," and "*STE6* → *STE12*." By adding the prior network, we obtain the gene network shown in Figure 9. As for the inverse normalized temperatures ζ_1 and ζ_2, we set $\zeta_1 = 0.5$ and choose the optimal value of ζ_2. We also estimated a gene network based on $\zeta_1 = 1$ and found the results described below to be essentially unchanged.

Figure 8 shows the behavior of BNRC with respect to ζ_2. We find that the optimal value of ζ_2 is 2.5. Figure 9 shows the resulting network based on microarray data and the biological knowledge represented by the prior network in Figure 6. We show the edges that correspond to the prior knowledge in black. The edges between genes that are regulated by the same transcription factor in the prior network are shown in blue. The red edges do not correspond to the prior knowledge. In particular, we find that the relationships around *MCM1* improve drastically. The network based on microarray only shown in Figure 7 indicates that only *SIC1* and *ACE2* are regulated by *MCM1*. Note that the underlined genes correspond to the prior network information. After adding the prior knowledge and optimizing the inverse normalized temperatures, we find that 10 genes out of 24 genes that are listed as coregulated genes of *MCM1* in Table 1 are extracted. Also, the relationships around *STE12* become clearer. Before adding prior knowledge, the estimated network in Figure 7 suggests that *FUS1*, *AFR1*, *KAR3*, *BAR1*, *MET4*, *MET16*, and *MCM1* are

Figure 7. Resulting network based on microarray only

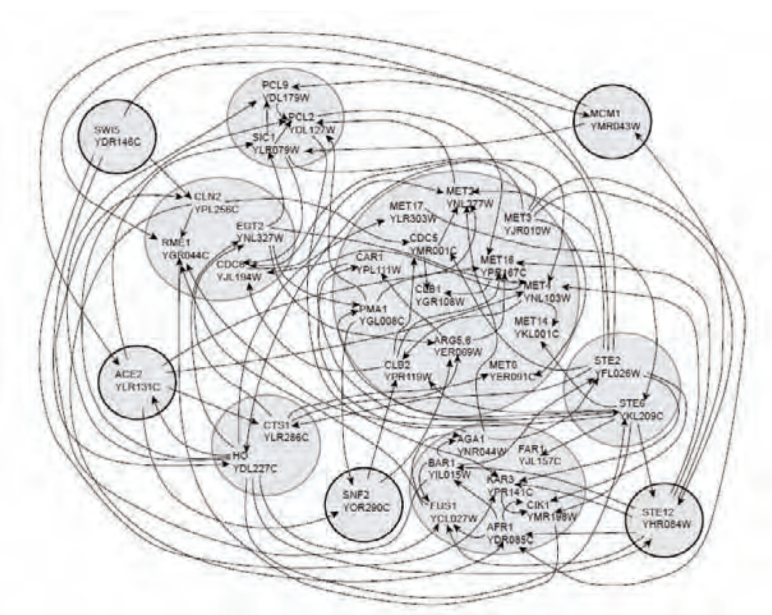


Figure 8. Optimization of ζ_2. We can find out that the optimal value of ζ_2 is 2.5.

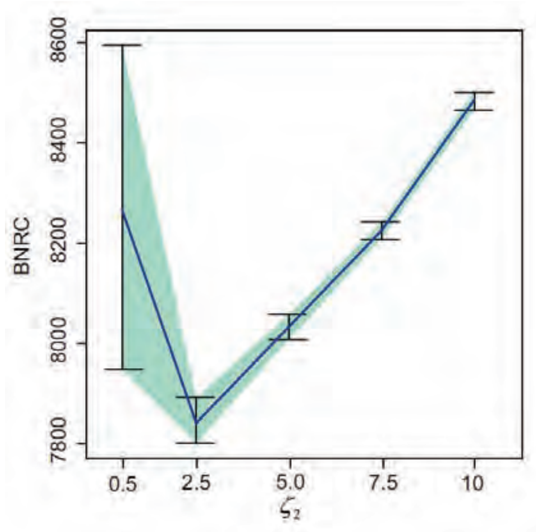
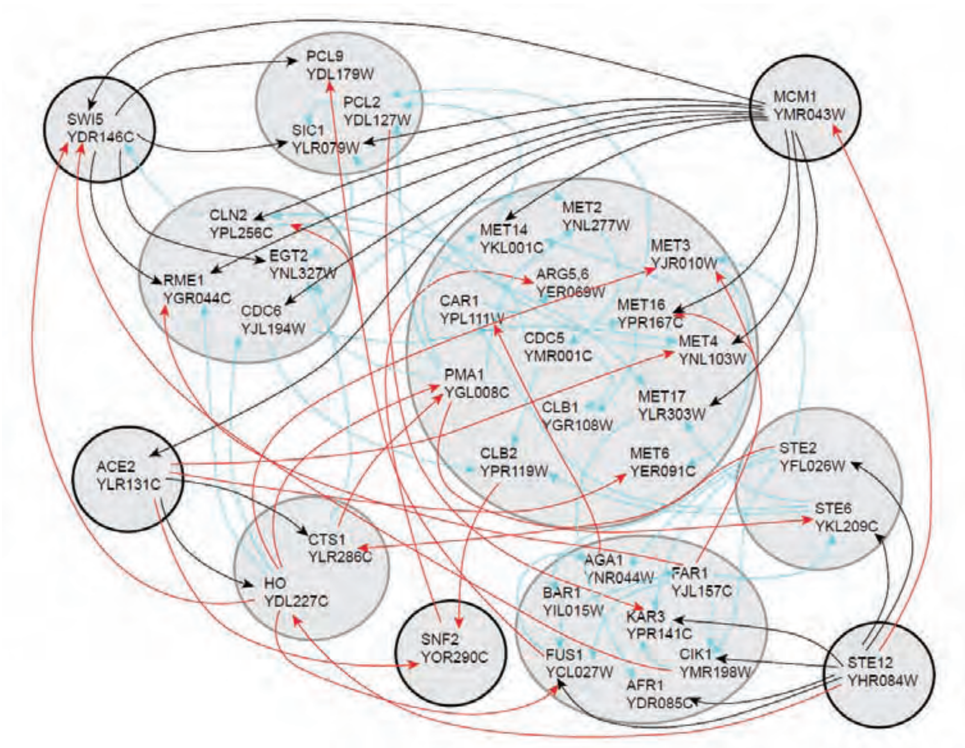


Figure 9. Resulting network based on microarray data and biological knowledge. The inverse normalized temperatures are selected by our criterion (ζ_1 = 0.5, ζ_2 = 2.5).



regulated by *STE12*, while *STE12* is controlled by *HO*, *STE6*, and *MET3*. On the other hand, the network in Figure 9 shows that *STE12* regulates *FUS1*, *AFR1*, *KAR3*, *CIK1*, *STE2*, *STE6*, *HO*, and *MCM1*. Note that the three misdirected relations described above are corrected in Figure 9. The difference between the inverse normalized temperatures $\zeta_1 = 0.5$ and $\zeta_2 = 2.5$ is small because the score of the criterion is added as $2\zeta_1$ or $2\zeta_2$, when we add an edge that is listed or not listed in the prior network, respectively. Therefore, microarray data contain this information and we succeeded in extracting this information with the slight help of the prior network.

We optimized the inverse normalized temperature ζ_2 based on the proposed criterion. From the network based on the optimal inverse normalized temperatures, we can find the difference between microarray data and biological knowledge. By comparing Figure 7 with Figure 9, we find that the microarray data reflect the relationship between seven genes (*CLN2*, *RME1*, *CDC6*, *EGT2*, *PCL2*, *PCL9*, and *SIC1*) and two transcription factors (*MCM1* and *SWI5*). On the other hand, we find that there are somewhat large differences between microarray data and the prior network for the relationship between *MCM1* and the 13 genes that are in the biggest circle.

APPLICATION TO HUMAN ENDOTHELIAL CELLS' GENE NETWORK

Fenofibrate Time-Course Data

We measure the time-responses of human endothelial cell genes to 25mM fenofibrate. The expression levels of 20,469 probes are measured by CodeLinkTM Human Uniset I 20K at six time-points (0, 2, 4, 6, 8, and 18 hours). Here, time 0 means the start point of this observation and just before exposure to the fenofibrate. In addition, we measure this time-course data as the duplicated data in order to confirm the quality of experiments.

Because our fenofibrate time-course data are duplicated data and contain six time-points, there are $2^6 = 64$ possible combinations to create a time-course dataset. We should fit the same regression function to a parent-child relationship in the 64 datasets. Under this constrain, we consider fitting nonparametric regression model to the connected data of 64 datasets. That is, if we consider gene i → gene j, we will fit the model $x_j^{(c)}(t) - m_j(x_i^{(c)}(t-1) + e_j^{(c)}(t)$, where $x_j^{(c)}(t)$ is the expression data of gene j at time t in the c-th dataset for $c = 1, ..., 64$. In the Bayesian networks, the reliability of estimated edges can be measured by using the bootstrap method. For time-course data, several modifications of the bootstrap method are proposed, such as block resampling, but it is difficult to apply these methods to the small number of data points generated by short time-courses. However, by using above time-course modeling, we can define a method based on the bootstrap as follows: Let $D = \{D(1), ..., D(64)\}$ be the combinatorial time-course data of all genes. We randomly resample $D(c)$ with replacement and define a bootstrap sample $D^* = \{D^*(1), ..., D^*(64)\}$. We then re-estimate a gene network based on D^*. We repeat 1000 times bootstrap replications and obtain $\hat{G}_T^{*1},...,\hat{G}_T^{*1000}$, where $\hat{G}_T^{*B}$ is the estimated graph based on the B-th bootstrap sample. The estimated reliability of edge can be used as the matrix representation of the first prior information $\boldsymbol{Z}_1$ as:

$$z_{ij}^{(1)} = \#\{B | e(i,j) \in \hat{G}_T^{*B}, B = 1,...,1000\}/1000$$

Gene Knock-Down Data by siRNA

For estimating gene networks, we newly created 270 gene knock-down data by using siRNA. We measure 20,469 probes by CodeLinkTM Human Uniset I 20K for each knock-down microarray after 24 hours of siRNA transfection. The knock-down genes are mainly transcription factors and signaling molecules. Let $\tilde{\boldsymbol{x}}_{D_i} = (\tilde{x}_{1|D_i}, ..., \tilde{x}_{p|D_i})'$ be the raw intensity vector of i-th knock-down microarray. For normalizing expression values of each microarray, we compute the median expression value vector $\boldsymbol{v} = (v_1, ..., v_p)'$ as the control data, where $vj = \text{median}_i(\tilde{x}_{j|D_i})$ We apply the loess normalization method to the MA transformed data and the normalized intensity $x_{j|D_i}$ is obtained by applying the inverse transformation to the normalized $\log(\tilde{x}_{j|D_i}/V_j)$ We refer to the normalized $\log(\tilde{x}_{j|D_i}/V_j)$ as the log-ratio.

In 270 gene knock-down microarray data, we know which gene is knocked-down for each microarray. Thus, when we knock-down gene D_i , genes that significantly change their expression levels can be considered as the direct regulatees of gene D_i. We measure this information by computing corrected log-ratio as follows: The fluctuations of the log-ratios depend on their sum of sample's and control's intensities. From the normalized MA transformed data, we can obtain the conditional variance $s_j = \text{Var}[\log(x_{j|D_i}/v_j) \mid \log(x_{j|D_i} \cdot v_j)]$ and the log-ratios can be corrected $z_{ij}^{(2)} = \log(x_{j|D_i}/v_j) / s_j$, satisfying $\text{Var}(z_{ij}^{(2)}) = 1$.

Results

For estimating fenofibrate-related gene networks from fenofibrate time-course data and 270 gene knock-down data, we first define the set of genes that are possibly related to fenofibrate as follows: First, we extract the set of genes whose variance-corrected log-ratios, $|\log(x_{j|D_i}/v_j)/s_j|$, are greater than 1.5 from each time point. We then find significant clusters of selected genes using GO Term Finder. Table 2 shows the significant clusters of genes at 18 hours. The first column indicates how expression values are changed, that is, right-up-arrow and right-down-arrow mean "overexpressed" and "suppressed," respectively. The GO annotations of clusters with right-down-arrow are mainly related to cell cycle, the genes in these clusters are expressed ubiquitously and this is a common biological function. On the other hand, the GO annotations of clusters with right-up-arrow are

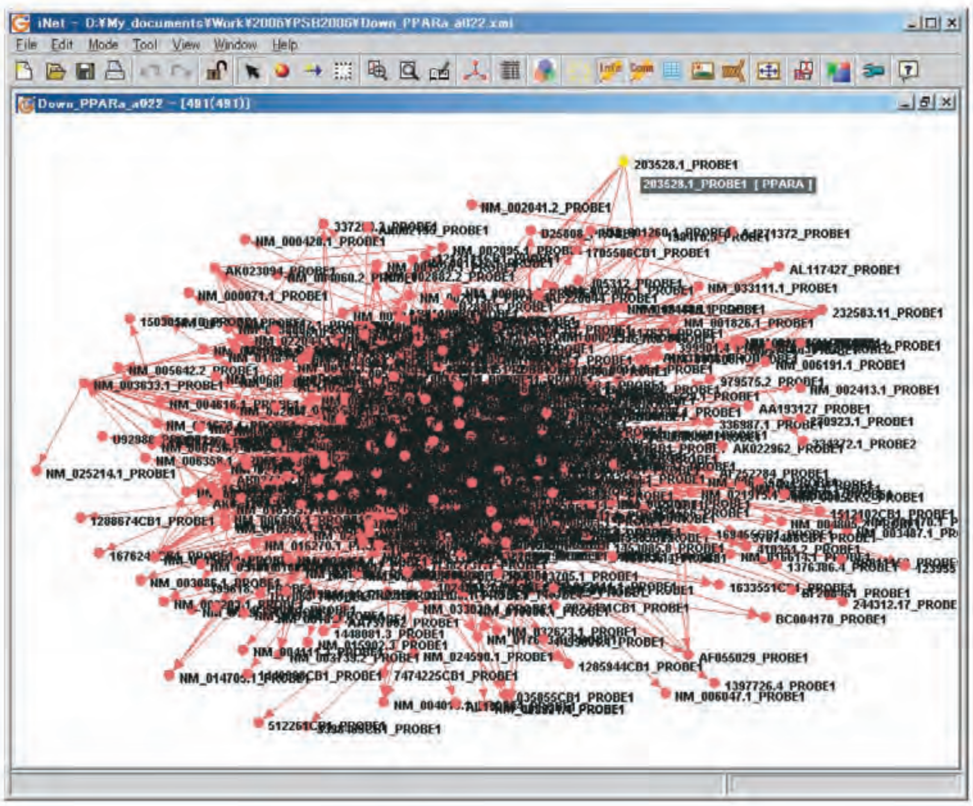
Table 2. Significant GO annotations of selected fenofibrate-related genes from 18 hours microarray

		GO Function	p-value	#genes
↘	GO:0007049	cell cycle	1.0E-08	35
↘	GO:0000278	mitotic cell cycle	3.7E-07	19
↘	GO:0000279	M phase	5.0E-06	17
↗	GO:0006629	lipid metabolism	1.3E-05	25
↘	GO:0007067	mitosis	1.3E-05	15
↘	GO:0000087	M phase of mitotic cell cycle	1.6E-05	15
↘	GO:0000074	regulation of cell cycle	2.7E-05	22
↗	GO:0044255	cellular lipid metabolism	4.4E-05	21
↗	GO:0016126	sterol biosynthesis	4.3E-04	6
↗	GO:0016125	sterol metabolism	4.5E-04	8
↗	GO:0008203	cholesterol metabolism	1.5E-03	7
↗	GO:0006695	cholesterol biosynthesis	2.4E-03	5
↗	GO:0008202	steroid metabolism	3.6E-03	10
↘	GO:0000375	RNA splicing, via transesterification reactions	4.1E-03	9
↘	GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	4.1E-03	9
↘	GO:0000398	nuclear mRNA splicing, via spliceosome	4.1E-03	9
↗	GO:0006694	steroid biosynthesis	6.0E-03	7
↘	GO:0016071	mRNA metabolism	6.3E-03	13

mainly related to lipid metabolism. In biology, it is reported that the fenofibrate acts around 12 hours after exposure (Goya, Sumitani, Xu, Kitamura, Yamamoto, Kurebayashi et al., 2004; Hayashida, Kume, Minami, Inui-Hayashida, Mukai, Toyohara, & Kita, 2004). Our first analysis for gene selection suggests that fenofibrate affects genes related to lipid metabolism, and this is consistent with biological facts. We also focus on the genes from the 8 hour time-point microarray. Unfortunately, no cluster with specific function could be found in the selected genes from the 8 hour time-point microarray However, there also exist some genes related to lipid metabolism. Therefore we use the genes from the 8 and 18 hour time-point microarrays. Finally, we add the 267 knock-down genes (three genes are not spotted on our chips) to the selected genes above, total 1192 genes are defined as possible fenofibrate-related genes and used for the next network analysis.

By converting the estimated dynamic network and knock-down gene information into the matrix representations of the first and second prior information $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$, respectively, we estimate the gene network $\hat{G}_k$ based on $\boldsymbol{Z}_1$, $\boldsymbol{Z}_2$ and the knock-down data matrix $\boldsymbol{X}_K$. For extracting biological information from the estimated gene network, we first focus on lipid metabolism-related genes, because the clusters related to this function are significantly changed at 18 hours microarray. In

Figure 10. Down-stream of PPAR-α

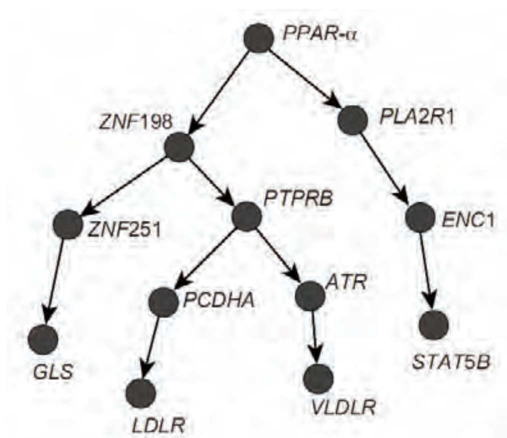


the estimated gene network, there are 42 lipid metabolism-related genes and *PPAR-α* (*Homo sapiens* peroxisome proliferative activated receptor, alpha) is the only transcriptionfactor among them. Actually, *PPAR-α* is a known target of fenofibrate. Therefore, we next focus on the node down-stream of *PPAR-α*. In Figure 10, the node down-stream of *PPAR-α* (491 genes). Here, we consider that genes in the four steps down-stream of *PPAR-α* are candidate regulatees of *PPAR-α*. Among the candidate regulatees of *PPAR-α*, there are 21 lipid metabolism-related genes and 11 molecules previously identified experimentally to be related to *PPAR-α*. Actually, *PPAR-α* is known to be activated by fenofibrate. We show one subnetwork having *PPAR-α* as a root node in Figure 11. One of the drug efficacies of fenofibrate whose target is *PPAR-α* is to reduce LDL cholesterol. LDLR and VLDLR mainly contribute the transporting of cholesterol and they are children of *PPAR-α*, namely candidate regulatees of *PPAR-α*, in our estimated network. As for LDLR, its relationship with *PPAR-α* has been reported (Islam, Knight, Frayn, Patel, & Gibbons, 2005). Moreover, several genes related to cholesterol metabolism are children of *PPAR-α* in our network. We also could extract STAT5B and GLS that are children of *PPAR-α* and have been reported their regulation-relationships with *PPAR-α* (Kersten, Mandard, Escher, Gonzalez, Tafuri, Desvergne, & Wahli, 2001; Shipley & Waxman, 2003). Therefore, it is not surprising that our network shows that many direct and indirect relationships involving known *PPAR-α* regulates are triggered in endothelial cells by fenofibrate treatment. In the node up-stream of *PPAR-α*, *PPAR-α*, and RXR-α, which form a heterodimer, share a parent. We could extract fenofibrate-related gene network and estimate that *PPAR-α* is the one of the key molecules of fenofibrate regulations without previous biological knowledge.

DISCUSSION

In this chapter, we illustrate a statistical method for estimating gene networks from microarray gene expression data. The use of Bayesian networks and nonparametric regression is a key of our method to capture nonlinear relationships

Figure 11. A sub-network related to PPAR-α



between genes. An information criterion we call BNRC, derived from the Bayesian point of view in Section 2, is powerful to evaluate candidate networks and to choose one as the best one. Also, the framework we used to derive BNRC is flexible to combine other biological knowledge. We have conceptually shown the capability of our framework to use protein-protein interaction, binding site information, and other types of information to estimate gene networks. In fact, research has been done using this framework. Tamada, et al. (2003) combined the promoter element detection method with Bayesian networks to find binding sites of transcription factors and to estimate gene networks simultaneously. Nariai et al. (2004) used protein-protein interaction data observed by yeast two hybrid and TAP experiments together with microarray data to estimate gene networks. More recently, Tamada et al. (2005) used protein sequence similarity as evolutionary information of two distinct organisms and estimate gene networks of two organisms, simultaneously. As we described above, combination of multiple types of genomic data is a key to estimate accurate gene networks.

The next step of gene network research is to understand mode-of-action of a chemical compound when we do it to cells by using estimated gene networks. In reality, the affected genes by a chemical compound, or drug, form subgraphs in the whole gene network. Therefore, it is possible to consider estimating subnetworks affected by a drug as a next challenge. This network enables us to know not only genes that are directly affected by the drug (Imoto et al., 2003), but also the drug affected pathways (Tamada et al., 2005). Imoto, Tamada, Araki, Yasuda, Print, Charnock-Jones et al. (2006) provides a concept druggable gene networks to understand mode-of-action of a chemical compound. The druggable gene networks contain known drug target gene and novel drug target candidate genes. They show that several known drug target genes are hubs in the druggable gene network, computationally.

Statistical estimation approach is a way to know gene networks in bioinformatics. This has a background that we do not know or partially know the true gene network and estimate unknown parts by using observational data like microarray gene expression data. However, there is another approach to understand networks in cells. This is based on the known facts and builds a simulation model like Petri net (Matsuno, Tanaka, Aoshima, Doi, Matsui, & Miyano, 2003; Nagasaki, Doi, Matsuno, & Miyano, 2004a, 2004b). We consider that the combination of these two approaches will become a key technology to understanding cellular networks.

ACKNOWLEDGMENT

We wish to thank our colleagues and collaborators Hideo Bannai, Yoshinori Tamada, Ryo Yoshida, Takao Goto, Sunyong Kim, Naoki Nariai, Sascha Ott, Tomoyuki Higuchi, Sachiyo Aburatani, Kousuke Tashiro, Satoru Kuhara, Ben Dunmore, Sally Humphries, Muna Affara, Kaori Yasuda, Hiromitu Araki, Cristpher J. Savoie, Stephen D. Charnock-Jones, Yuki Tomiyasu, Deborah Sanders, and Cristin G. Print. Computation time was provided by the Super Computer System, Human Genome Center, Institute of Medical Science, University of Tokyo.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle (pp. 267-281). In B.N. Petrov & F. Csaki (Eds.). Akademiai Kiad´o, Budapest.

Bader, G.D., Donaldson, I., Wolting, C., Ouellette, B.F.F., Pawson, T., & Hogue, C.W.V. (2001). BIND: The biomolecular interaction network database. *Nucleic Acids Research*, *29*, 242-245.

Bannai, H., Inenaga, S., Shinohara, A., Takeda, M., & Miyano, S. (2002). A string pattern regression algorithm and its application to pattern discovery in long introns. *Genome Informatics, 13*, 3-1.

Berger, J.O. (1985). *Statistical decision theory and Bayesian analysis*. New York: Springer-Verlag.

Bernard, A., & Hartemink, A. (2005). Informative structure priors: Joint learning of dynamic regulatory networks from multiple types of data. In *Proceedings of the Pacific Symposium on Biocomputing* (vol. 10, pp. 459-470).

Burnham, K.P., & Anderson, D.R. (1998). *Model selection and inference: A practical information-theoretical approach*. New York: Springer-Verlag.

Bussemaker, H.J., Li, H., & Siggia, E.D. (2001). Regulatory element detection using correlation with expression. *Nature Genetics*, *27*, 167-171.

Cooper, G.F., & Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning, 9*, 309-347.

De Boor, C. (1978). *A practical guide to splines*. Berlin: Springer-Verlag.

Friedman, N., & Goldszmidt, M. (1998). Learning Bayesian networks with local structure. In M.I. Jordan (Ed.). Kluwer Academic Publisher.

Friedman, N., Linial, M., Nachman, I., & Pe'er, D. (2000). Using Bayesian network to analyze expression data. *Journal of Computational Biology*, *7*, 601-620.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restorations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721-741.

Goya, K., Sumitani, S., Xu, X., Kitamura, T., Yamamoto, H., Kurebayashi, S. et al. (2004). Peroxisome proliferator-activated receptor α agonists increase nitric oxide synthase expression in vascular endothelial cells. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *24*, 658-663.

Green, P.J., & Silverman, B.W. (1994). *Nonparametric regression and generalized linear models*. Chapman & Hall.

Hartemink, A.J., Gifford, D.K., Jaakkola, T.S., & Young, R.A. (2002). Combining location and expression data for principled discovery of genetic regulatory network models. In *Proceedings of the Pacific Symposium on Biocomputing* (vol. 7, pp. 437-449).

Hastie, T., & Tibshirani, R. (1990). *Generalized additive models*. Chapman & Hall.

Hayashida, K., Kume, N., Minami, M., Inui-Hayashida, A., Mukai, E., Toyohara, M., & Kita, T. (2004). Peroxisome proliferator-activated receptor α ligands activate transcription of lectin-like oxidized low density lipoprotein receptor-1 gene through GC box motif. *Biochemical and Biophysical Research Communications*, *323*, 1116-1123.

Heckerman, D., & Geiger, D. (1995). Learning Bayesian networks: A unification for discrete and Gaussian domains. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (pp. 274-284).

Heckerman, D., Geiger, D., & Chickering, D.M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, *20*, 197-243.

Ideker, T., Ozier, O., Schwikowski, B., & Siegel, A.F. (2002). Discovering regulatory and signaling circuits in molecular interaction networks. *Bioinformatics*, *18*, S233-S240.

Imoto, S., Goto, T., & Miyano, S. (2002). Estimation of genetic networks and functional structures between genes by using Bayesian networks and nonparametric regression. In *Proceedings of the Pacific Symposium on Biocomputing* (vol. 7, pp. 175-186).

Imoto, S., Higuchi, T., Goto, T., Tashiro, K., Kuhara, S., & Miyano, S. (2004). Combining microarrays and biological knowledge for estimating gene networks via Bayesian networks. *Journal of Bioinformatics and Computational Biology*, *2*, 77-98.

Imoto, S., Higuchi, T., Kim, S., Jeong, E., & Miyano, S. (2005). Residual bootstrapping and median filtering for robust estimation of gene networks from microarray data. In *Proceedings of the 2nd Computational Methods in Systems Biology* (LNIB 3082, pp. 149-160). Springer-Verlag.

Imoto, S., Kim, S.-Y., Goto, T., Aburatani, S., Tashiro, K., Kuhara, S., & Miyano, S. (2003). Bayesian network and nonparametric heteroscedastic regression for nonlinear modeling of genetic network. *Journal of Bioinformatics and Computational Biology,* 1, 231-252.

Imoto, S., & Konishi, S. (2003). Selection of smoothing parameters in B-spline nonparametric regression models using information criteria. *Annals of the Institute of Statistical Mathematics*, 55, 671-687.

Imoto, S., Savoie, C.J., Aburatani, S., Kim, S., Tashiro, K., Kuhara, S., & Miyano, S. (2003). Use of gene networks for identifying and validating drug targets. *Journal of Bioinformatics and Computational Biology, 1,* 459-474.

Imoto, S., Tamada, Y., Araki, H., Yasuda, K., Print, C.G., Charnock-Jones, S.D. et al. (2006). Computational strategy for discovering druggable gene networks from genome-wide RNA expression profiles. In *Proceedings of the Pacific Symposium on Biocomputing* (vol. 11, pp. 559-571).

Islam, K.K., Knight, B.L., Frayn, K.N., Patel, D.D., & Gibbons, G.F. (2005). Deficiency of PPARα disturbs the response of lipogenic flux and of lipogenic and cholesterogenic gene expression to dietary cholesterol in mouse white adipose tissue. *Biochimica et Biophysica Acta,* 1734, 259-268.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., & Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Science, 97,* pp. 4569-4574).

Kersten, S., Mandard, S., Escher, P., Gonzalez, F.J., Tafuri, S., Desvergne, B., & Wahli, W. (2001). The peroxisome proliferator-activated receptor α regulates amino acid metabolism. *Federation of American Societies for Experimental Biology Journal*, *15*, 1971-1978.

Kindeman, R., & Snell, J. (1980). *Markov random fields and their applications*. Providence, RI: American Mathematical Society.

Konishi, S. (1996). Statistical model evaluation and information criteria. In S. Ghosh (Ed.). Marcel Dekker.

Konishi, S., & Kitagawa, G. (1996). Generalised information criteria in model selection. *Biometrika*, *83*, 875-890.

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K. et al. (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science*, *298,* 799-804.

Masys, D.R. (2001). Linking microarray data to the literature. *Nature Genetics*, *28,* 9-10.

Matsuno, H., Tanaka, Y., Aoshima, H., Doi, A., Matsui, M., & Miyano, S. (2003). Biopathways representation and simulation on hybrid functional Petri net. *In Silico Biology*, *3*, 389-404.

Nagasaki, M., Doi, A., Matsuno, H., & Miyano, S. (2004a). Integrating biopathway databases for largescale modeling and simulation. In *Proceedings of the Second Asia-Pacific Bioinformatics Conference* (vol. 29, pp. 43-52). Australian Computer Society.

Nagasaki, M., Doi, A., Matsuno, H., & Miyano, S. (2004b). A versatile petri net based architecture for modeling and simulation of complex biological processes. *Genome Informatics, 15,* 180-197.

Nariai, N., Kim, S.-Y., Imoto, S., & Miyano, S. (2004). Using protein-protein interactions for refining gene networks estimated from microarray data by Bayesian networks. In *Proceedings of the Pacific Symposium on Biocomputing* (vol. 9, pp. 336-347).

Ong, I.M., Glasner, J.D., & Page, D. (2002). Modelling regulatory pathways in E. coli from time series expression profiles. *Bioinformatics*, *18*, S241-S248.

Ott, S., Hansen, A., Kim, S.-Y., & Miyano, S. (2005). Superiority of network motifs over optimal networks and an application to the revelation of gene network evolution. *Bioinformatics*, *21*, 227-238.

Ott, S., Imoto, S., & Miyano, S. (2004). Finding optimal models for small gene networks. In *Proceedings of the Pacific Symposium on Biocomputing*, *9*, 557-567.

Ott, S., & Miyano, S. (2003). Finding optimal gene networks using biological constraints. *Genome Informatics*, *14*, 124-133.

Pe'er, D, Regev, A., Elidan, G., & Friedman, N. (2001). Inferring subnetworks from perturbed expression profiles. *Bioinformatics*, *17*, S215-S224.

Pilpel, Y., Sudarsanam, P., & Church, G.M. (2001). Identifying regulatory networks by combinatorial analysis of promoter elements. *Nature Genetics*, *29*, 153-159.

Robinson, R.W. (1973). Counting labeled acyclic diagraphs. In F. Harary (Ed.), *New directions in the theory of graphs* (pp. 239-273). New York: Academic Press.

Segal, E., Barash, Y., Simon, I., Friedman, N., & Koller, D. (2002). From promoter sequence to expression: A probabilistic framework. In *Proceedings of the 6th Annual International Conference on Research in Computational Molecular Biology* (pp. 263-272).

Segal, E., Wang, H., & Koller, D. (2003). Discovering molecular pathways from protein interaction and gene expression data. *Bioinformatics*, *19*, i264-i272.

Segal, E., Yelensky, R., & Koller, D. (2003). Genome-wide discovery of transcriptional modules from DNA sequence and gene expression. *Bioinformatics*, *19*, i273-i282.

Shipley, J.M., & Waxman, D.J. (2003). Down-regulation of STAT5b transcriptional activity by ligand-activated peroxisome proliferator-activated receptor (PPAR) α and PPAR γ. *Mol. Pharmacol.*, *64*, 355-364.

Tamada, Y., Bannai, H., Imoto, S., Katayama, T., Kanehisa, M., & Miyano, S. (2005). Utilizing evolutionary information and gene expression data for estimating gene regulations with Bayesian network models. *Journal of Bioinformatics and Computational Biology*, *3*, 1295-1313.

Tamada, Y., Imoto, S., Tashiro, K., Kuhara, S., & Miyano, S. (2005). Identifying drug active pathways from gene networks estimated by gene expression data. *Genome Informatics*, *16*, 182-191.

Tamada, Y., Kim, S.-Y., Bannai, H., Imoto, S., Tashiro, K., Kuhara, S., & Miyano, S. (2003). Estimating gene networks from gene expression data by combining Bayesian network model with promoter element detection. *Bioinformatics*, *19*, ii227-ii236.

Van Someren, E.P., Wessels, L.F.A., Backer, E., & Reinders, M.J.T. (2002). Genetic network modeling. *Pharmacogenomics*, *3*, 507-525.

This work was previously published in Bayesian Network Technologies: Applications and Graphical Models, edited by A. Mittal & A. Kassim, pp. 269-299, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 7.32
Fuzzy Logic in Medicine

Michelle LaBrunda
Cabrini Medical Center, USA

Andrew LaBrunda
University of Guam, USA

ABSTRACT

This article explores the use of fuzzy logic in the medical field. While giving a comparison of classic and fuzzy logic we present the various uses of the applications made possible by fuzzy logic, focusing on diagnosis and treatment. The ever evolving technology making the line between medicine and technology thinner every year, is helping to make the treatment of disease and the mending of injury easier for medical professionals. We also propose several questions that arise from, and may by answered by, fuzzy logic and its applications.

INTRODUCTION

In order to understand the intricacies of fuzzy logic one must build from a thorough understanding of classical logic. A basic component of classic logic is the proposition that a statement can be characterized as either true or false. An example of a proposition is "The country of France contains the Eiffel tower" or "The Eiffel tower is closed on Sundays." In classical logic, propositions are said to be either true or false. Propositions are typically connected using AND, OR, and NOT. On occasion, one might use other connections, but they can all be derived from a combination of these three. The notation used to describe classic logic is called propositional calculus. In most computer programming languages it is common to assign numerical values to the correctness of a proposition where 1 = true and 0 = false. A proposition can either be true or it can be false. It cannot be both at the same time, nor can it simultaneously be neither. Collections of propositions can be transformed to prove truths that might not necessarily be evident on their own. The basic rules of mathematics have been transcribed into propositional calculus and as a result computers are now able to transform a series of propositions into mathematical proofs. Computers are now able to solve proofs in ways never previously conceived.

Fuzzy logic is similar to classical logic in the search for truthfulness of a proposition. Sometimes truth is subjective ill defined. As an example, it is difficult to assign a true or false value to the proposition “Andy is tall” or “Shell is old.” How tall does one have to be before being categorized as “tall”? Likewise, how old does one have to be before being considered “old”? Most would agree that 100 years is old for a person but young for a planet. Like many real-world propositions, the concept of age is relative to its usage. To solve these problems there was a need to develop a more robust system of logic. Rather than assigning a proposition as either 0 or 1 the idea of variable truth was added. The variable is measured over the interval of [0, 1]. Fuzzy logic rose from this concept. One major focus of this discipline is in the development of computational models, which can accurately assign fractional values to the level of truthfulness.

Contrary to the name, the goal of fuzzy logic is to create computer programs that can accept input and provide the user a clear answer. The system is defined as “fuzzy” because it is not always evident, given the input parameters, what logic path the system will take to derive a solution. Fuzzy logic systems are frequently used as expert systems. This type of system attempts to emulate a field expert of a specific discipline. Ideally this software-based expert would be able to accept input, process the information, and output clear concise responses. Unfortunately, in emulating the thought processes of such an expert, the expert system must emulate human thought. Human thought is fuzzy in nature complete with uncertainties, ambiguities, and contradictions. Two experts might not place the same level of importance on the same piece of information. Additionally, they might not look at the same information the same way. Should a glass filled 50% be classified as half full or half empty?

Many techniques have been used to create fuzzy logic programs that function as an expert. The earliest systems used conditional statements with tolerance thresholds using if-then-else rules (Jackson, 1999). This approach, while seemingly simplistic, has been used successfully in a wide variety of medical applications including diagnostics and psychological bias (Shortliffe, 1976). Other less known approaches of fuzzy logic systems are association nets and frames, which have proven difficult to implement with only marginal results. The two most common implementations of fuzzy logic are rule-based and neural networks. Both fuzzy implementations have a diverse range of applications including medicine, avionics, security, and machine learning.

Unlike rule-based fuzzy logic, neural nets do not require thinking patterns to be explicitly specified. Typically two data sets are created to program a neural network. The first data set is the trainer. This set of input is passed into the neural network and processed. The processing phase consists of sorting the input values among an array of memory structures call nodes. Each node retains some information and sorts the remaining information between neighboring nodes. Once all the information has been processed it is evaluated and stored as the template for which all other datasets will be compared.

This technique can loosely be compared to the Japanese game Pachinko, also seen in the game show Price is Right as Plinko. The input values are represented by silver balls that are dropped into an arrangement of pins held by a board. Before the balls reach their final stop at the bottom of the board they make contact with many pins, which change the balls direction and velocity. This makes it almost impossible to predict where a ball will end up when dropped. In the real world, chaos prevents the same input from yielding a consistent output. In a computer model, the same input will always produce the same output. So as the computer receives input, these data are feed into a virtual pachinko machine. The first batch of input is called the trainer input and represents where the balls should be landing under ideal conditions. Once the machine has been trained, new input can be fed into the same virtual pachinko machine. Once processing is complete, the landing spots for the input are compared against a trainer dataset. The greater the discrepancy between the trainer set and the actual set, the greater the error in the system. To complete the analogy, computer scientists put a great deal of effort into determining the optimal location and elasticity of the pins to make sure that good input and bad input produce different results. In the case of expert systems, neural networks do not explicitly extract the rules of an expert. Rather the neural network has learned only to recognize patterns. Therefore, while the results may be good, the neural network will not know why an answer is correct other than to say it looks like other successful inputs. The strength of this approach is that it is not necessary to program each business rule. The weakness of this approach is that substantial training may be required.

Rule-based systems explicitly collect the expert's knowledge. These rules and thinking patterns are then programmed into the system. An important phase of this approach is knowledge acquisition. During this phase a team of programmers develop a rigorous and complete model of the domain rules to be implemented. Rule-based systems do not require a large training set like that of neural network solutions. This is because the domain expert has clearly specified the rules and parameters, whereas a neural network has no knowledge of domain rules, only patterns that seem to have been enforced in past successes. During testing there is usually a tuning phase, where parameters are modified based upon the results of the test data. In addition to smaller datasets, another advantage of a rule-based fuzzy logic system is that it is easier for the system to rationalize its behavior to users. Rule-based fuzzy logic system behavior is determined by rules or parameters and changes to these parameters represent the incentives for the system to take action. This is much more easily communicated to an expert and programmer using rule-based fuzzy logic than a neural network.

FUZZY LOGIC IN THE MEDICAL FIELD

While computer scientists are refining and advancing fuzzy logic, applications within the medical field have already started to emerge. Some of these applications are rudimentary, but with time will prove invaluable. Fuzzy logic has made its way into general medicine, basic science, as well as diagnosis and treatment.

General Medicine

General medicine encompasses an enormous body of knowledge drawing from almost every field imaginable. As technology advances, the lines between medicine and technology are blurring. Fuzzy technology is a relatively new concept in computer science and few articles have been written applying fuzzy logic to medical topics. The application of fuzzy logic in general medicine has been used in addiction modeling, mapping of bruising after being shot while wearing body armor, and in the analysis and evaluation of chronic disease.

The data that are available focus heavily on the mathematics and computer science involved in applying basic fuzzy concepts to medicine. For example, an article by Nieto and Torres (2003) describes a model examining the risk of illicit drug use based on alcohol consumption and smoking habits of adolescents. Instead of labeling people as smokers or non-smokers, the degree of smoking is given a value between 0 (non-smoking) and 1 (smokes like a chimney). The same is true for alcohol consumption. Kosko's hypercube was used and a fuzzy set with a point in a unit hypercube was identified. A hypercube is an "n" dimensional figure analogous to a square. For example, if n=2 the hypercube is a square, if n=3 the hypercube is a cube. They performed a similar successful analysis looking at stroke risk factors. The article largely focuses on the engineering and mathematics behind fuzzy systems. As the utilization of fuzzy systems becomes more intuitive it is likely that articles focusing on the medical application of these systems will become available.

Lee, Kosko, and Anderson (2005) also examine a fuzzy system application to medicine. This article models gunshot bruises in soft body armor with a fuzzy system. When someone wearing soft body armor is shot, the bullet is stopped, but a bruise is produced. Analysis of the bruise and armor deformation can give information on the bullet's mass and momentum. Detailed analysis of information provided from studies such as this may allow engineers to design new types of armor better able to protect the body while minimizing armor weight. While the direct applications to medicine may be somewhat distant, the thought process involved in developing this article further combine medicine and technology bringing the real application closer to existence.

Many facets of general medicine deal with chronic disease. Chronic diseases are illnesses that last for a long time. Examples of chronic disease that will be discussed in this article are chronic kidney disease, cancer, and HIV.

The Diatelic project is a prospective randomized research project designed to test the ability of a fuzzy system to monitor chronic kidney disease patients and alert a physician of disease progression (Jeanpierre & Charpillet, 2004). In this project, patients provided daily Internet updates. These updates were analyzed by a computer using partially observable Markov decision process (POMPD) logic. POMPD is a mathematical framework from which decision-making modeling systems can be made when some of the factors are random and others are controlled. It assumes that the current state of the modeled object cannot be completely known. In the aforementioned study, a nephrologist (kidney specialist) was notified if a patient's condition appeared to be progressing. This system utilizing POMPD requires relatively low computing complexity. As the number of physicians per person continues to decline, systems of this sort will play an important roll in monitoring patient's health, disease progress, and physician alerting as problems develop.

Recent research has been done looking at fuzzy logic in diagnosing breast cancer (Polat, Şahan, Kodaz, & Güneş, 2007). The artificial immune recognition system (AIRS) has been around since 2001. It is a learning algorithm modeled after the functioning of the immune system. Polat et al. applied performance evaluation with fuzzy resource allocation mechanisms to the AIRS creating a new system, Fuzzy-AIRS. Fuzzy-AIRS was used to analyze data from a Wisconsin breast cancer dataset and predicted which samples were most likely to be cancerous. The Fuzzy-AIRS results were compared with histological diagnosis and found to be 99% accurate. Methods such as Fuzzy-AIRS can greatly reduce the time required to analyze samples and allow for more prompt diagnosis and treatment.

HIV is no longer the death sentence it once was. HIV-infected individuals in developed nations are now living nearly normal life spans thanks to advances in medication therapy and diagnosis. Many questions face infectious disease physicians when deciding how to treat HIV. These questions include when to start therapy, what medications to give, and when to change the medications. Medication is generally changed when the HIV mutates and becomes resistant to the medications that the patient is receiving. Recent research by Sloot, Boukhanovsky, Keulen, Tirado-Ramos, and Boucher (2005) utilizes multivariate analyses combined with rule-based fuzzy logic to produces a physician advice system. This system integrates a number of factors affecting the outcome of therapy and suggests a clinical course of action. As the understanding of disease, disease process, and the availability and complexity of medications increases, physicians will become more and more dependent on advice systems such as this in medical decision making. This system would also allow less skilled physicians to provide a basic level of care. This would be of enormous value in developing countries where access to physicians and therapeutic intervention is limited.

Basic Science

Taking a scientific concept and eventually developing a medical application for that concept is a long process. Basic biomedical science is the study of components of biological systems. The goals of this research do not often have direct medical applications but rather provide the building blocks for advanced biomedical diagnostic and therapeutic regimes. All medical advances have their roots in the basic sciences. Basic science includes fields such as biology, microbiology, physics, chemistry, and bioengineering. Fuzzy logic has allowed researchers to make advances that would be difficult and time consuming through other modalities, especially in the arena of modeling complex systems.

One example of the application of fuzzy logic in basic science has been done in the field of gene expression. Almost every cell in the human body contains DNA. DNA is the blueprint directing cellular functions to make each person into a human being and unique individual. On each strand of DNA are regions called genes. Genes can be thought of as the active part of DNA. Each gene contains specific directions on how to make one of the bodies many proteins. Some genes are always active. Others are only active before birth, while others are activated only under special circumstances such as pregnancy or adolescence. Genes that are activated are called expressed genes because their corresponding proteins are produced. The control of gene expression is extremely complex involving numerous chemical messengers that are not completely understood. Disregulation of gene expression has been associated with numerous diseases, most notably cancer.

A study by Du, Gong, Wurtele, and Dickerson (2005) models gene expression utilizing fuzzy logic. Because there are numerous layers of interconnecting factors regulating expression, it is amenable to fuzzy logic modeling. Cluster analysis assumes that genes with similar functions have similar mechanisms of control and produce models based on this assumption. By analyzing these repeated patterns new mechanisms of control have been elucidated.

Another related basic science topic is protein engineering. Proteins are typically built by cellular machinery that is difficult to manipulate. Engineered proteins have applications in the pharmaceutical, agricultural, and synthetic organic chemistry fields, but pure proteins must be produced. The screening of proteins to determine which synthetic proteins have desired traits is a labor intensive and expensive process involving the processing of huge amounts of data. Kato et al. (2005) developed a strategy utilizing a fuzzy neural network to screen proteins constructing a method for more quickly and efficiently screening large sets of proteins.

Diagnosis and Treatment

Once basic science has been well understood, the medical applications of this technology can be explored. Neural network and fuzzy logic have been slowly making their way into medical diagnosis and treatment as well as basic sciences. Fuzzy logic is primarily used as a modeling tool and can help to manage large sets of data while identifying individuals who follow a typical pattern of disease or disease progression. An automated system is much more efficient in processing large amounts from a number of simultaneous sources of data than the manual systems currently employed. Efforts are currently being made to determine optimal modeling systems that will allow data to be processed in an accurate and clinically meaningful manner.

A number of studies are being undertaken evaluating the role of fuzzy logic modeling in medicine. Many diseases have overlapping symptoms and the diagnosis is not always clear. John and Innocent (2005) recently published an article utilizing a fuzzy logic system to differentiate between diseases with similar symptoms. While not yet a perfect model, systems such as this may one day be of great benefit to physicians in trying to distinguish between diseases with similar symptoms. A recent review article by Grossi and Buscema (2006) looks at the application of artificial intelligence and fuzzy systems to outcomes research. As technology advances, fuzzy logic modeling systems may prove invaluable to outcomes research allowing complex data from multiple sources to be integrated into one conclusion that is accessible to patients and their healthcare providers.

Artificial intelligence and fuzzy systems have provided the opportunity to study what has traditionally fallen into the realms of philosophy and psychiatry. Sripada, Jobe, and Helgason (2005) have devised a fuzzy logic model of empathy. The ability to model complex systems may one day suggest treatments for diseases that currently have no therapies.

CONCLUSION

Computer-generated models may be highly accurate, but they will never be perfect. As many questions as answers are generated by high-powered fuzzy modeling systems. How precise must a fuzzy modeling system be before it can be implemented? If a modeling system fails and a patient is not properly diagnosed, who is at fault, the programmer, the physician, or the patient?

With these new modeling systems comes a plethora of ethical questions that will no longer be limited to science fiction films. Will we some day be able to model a person's behavior throughout their life based on their DNA? What would happen if modeling suggests that a baby is likely to become a murderer? If we can model classically human traits such as fear, love, and happiness, where will the limits between human and machine fall? Will we one day select our children based on computer modeling systems predicting their behavior, physical appearance, and aptitudes?

While these questions may seem somewhat premature, they are within the scope of reality. Fuzzy systems and neural networking provide a powerful modeling tool, the breadth of which we are only beginning to comprehend.

REFERENCES

Buchanan, B., & Shortliffe, E. (Eds.). (1984). *Rule-based expert systems*. Reading, MA: Addison-Wesley.

Du, P., Gong, J., Wurtele, E., & Dickerson, J. (2005). Modeling gene expression networks using fuzzy logic. *IEEE Transactions on Systems, Man and Cybernetics—Part B: Cybernetics, 35*(6), 1351-1359.

Grossi, E., & Buscema, M. (2006). Artificial intelligence and outcome research. *Drug Development Research, 67*, 227-244.

Jackson, P. (1999). *Introduction to expert systems*. Reading, MA: Addison-Wesley.

Jeanpierre, L., & Charpillet, F. (2004). Automated medical diagnosis with fuzzy stochastic models: Monitoring chronic diseases. *Acta Biotheoretica, 52*, 291-311.

John, R., & Innocent, P. (2005). Modeling uncertainty in clinical diagnosis using fuzzy logic. *IEEE Transactions on Systems, Man and Cybernetics—Part B: Cybernetics, 35*(6), 1340-1350.

Kato, R., Nakano, H., Konishi, H., Kato, K., Koga, Y., Yamane, T., et al. (2005). Novel strategy for protein exploration: High-throughput screening assisted with fuzzy neural network. *Journal of Molecular Biology, 351,* 683-692.

Lee, I., Kosko, B., & Anderson, F. (2005). Modeling gunshot bruises in soft body armor with an adaptive fuzzy system. *IEEE Transactions on Systems, Man and Cybernetics—Part B: Cybernetics, 35*(6), 1374-1390.

Nieto, J., & Torres, A. (2003). Midpoints for fuzzy sets and their application in medicine. *Artificial Intelligence in Medicine, 27,* 81-101.

Polat, K., , S. (2007). Breast cancer and liver disorders classification using artificial immune recognition system (AIRS) with performance evaluation by fuzzy resource allocation mechanism. *Expert Systems with Applications, 32*, 172-183.

Shortliffe, E. (1976). *Computer-based medical consultations: MYCIN*. New York: Elsevier.

Sloot, P., Boukhanovsky, A., Keulen, W., Tirado-Ramos, A., & Boucher, C. (2005). A grid-based HIV expert system. *Journal of Clinical Monitoring and Computing, 19*, 263-278.

Sripada, B., Jobe, T., & Helgason, C. (2005). From fuzzy logic toward plurimonism: The science of active and empathic observation. *IEEE Transactions on Systems, Man and Cybernetics—Part B: Cybernetics, 35*(6), 1328-1339.

This work was previously published in the Journal of Information Technology Research, edited by Mehdi Khosrow-Pour, Volume 1, Issue 1, pp. 27-33, copyright 2008 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Section VIII
Emerging Trends

This section highlights research potential within the field of medical informatics while exploring uncharted areas of study for the advancement of the discipline. Chapters within this section highlight evolutions in evidence-based medicine, the future of e-health, and web portals for medical data. These contributions, which conclude this exhaustive, multi-volume set, provide emerging trends and suggestions for future research within this rapidly expanding discipline.

Chapter 8.1
E-Health Systems:
Their Use and Visions for the Future

Pirkko Nykänen
Tampere University, Finland

ABSTRACT

E-health refers to use of information and communication technologies to improve or enable health and healthcare. E-health broadens the scope of healthcare delivery; citizens are in the center of services and services are offered by information systems often via the Internet. In this chapter e-health systems are classified on the basis of their use and their functionality and the use is discussed from the viewpoints of citizens and health professionals. Citizens are increasingly using Internet and e-health systems to search for medicine or health-related information, and they become better informed and may take more responsibility of their own health. Health professionals are more reluctant to use the Internet and e-health systems in physician-patient communication due to the power and responsibility problems of decisions. In the future the sociotechnical nature of e-health should be considered and future systems developed for real use and user environment with user acceptable technology.

INTRODUCTION

In the information society it is important to develop and apply technologies in such a way that we empower citizens to play a full role. An essential part of the information society, healthcare services are needed by citizens and should be provided efficiently and made accessible to all (Haglund, 2002).

With the information society a new concept, e-health, has been introduced to refer to the use of emerging information technology to improve or enable health and healthcare. Silber (2003) defined e-health as "application of information and communication technologies (ICT) across the whole range of functions that affect health" (p. 3). Eysenbach (2001) gave a broader definition for e-health: An emerging field in the intersection of medical informatics, public health and business, referring to health services and information delivered or enhanced thorough the Internet and related technologies. Alvarez (2002) emphasised the consumer-viewpoint when he defined e-health as a consumer-centered model of healthcare where stakeholders collaborate, utilising ICT and Internet technologies to manage health; arrange, deliver, and account for care; and manage the healthcare system. All these definitions support the conception that e-health means application of information technologies to promote health, and to support healthcare services delivery and use. E-health covers all health strategies: Prevention, treatment, and rehabilitation. It is essential that e-health applications meet the needs of citizens, patients, healthcare professionals, and policy makers. Therefore, evaluation studies are needed to assess the benefits, effects, and impacts of e-health on citizens, professionals, healthcare systems, and healthcare outcomes.

E-health conceptualization broadens the scope of healthcare delivery; citizens are placed at the centre of services, services are in many situations offered to be used through the Internet (e.g., at home) and citizens can have interaction with health professionals who look after their health needs (Silber, 2003; Wilson, Leitner, & Moussalli, 2004). E-health is expected to contribute to development of new ways of delivering health services and to impact on the organisation and structure of the healthcare delivery system. E-health is not only a technological improvement, it is a reengineering of healthcare processes and is of consideration of the sociotechnical aspects of design and development of applications.

E-HEALTH SYSTEMS

E-health applications should make citizens better informed. All citizens should have access to services, use of services should be economically affordable, and citizens should benefit from the use of services. On the other hand, e-health services should improve the quality, availability and effectiveness of healthcare (Grimson et al., 2000; Silber, 2003; Wilson, 2002).

Types of E-Health Systems

Traditionally, three broad categories of e-health applications can be identified: Delivery of care to patients by healthcare professionals, education and dissemination of health-related information and knowledge, and trading health products (Ruotsalainen et al., 2003).

The first category covers systems for *delivery of care to patients by healthcare professionals*, including a wide range of applications from pure administrative to those for care delivery:

- **Hospital systems:** Scheduling systems, logistics systems, management information systems, hospital and patient administration systems, laboratory information systems, radiology information systems, pharmacy systems, nursing systems, and networked services such as electronic messaging between the hospital and other healthcare actors for communication of clinical information and administrative data, including telemedical services such as telepathology and teleconsultation for remote areas.
- **Primary care systems:** Information systems for general practitioners, pharmacists, and dentists for patient management, medical records, electronic prescribing, and information exchange.
- **Home care systems:** Systems that are used to deliver care services via telecommunication or wireless to the patient at home. Examples of such systems are remote vital-signs monitoring systems that enable the patient to receive targeted treatment and medication without the need to visit an outpatient clinic or occupy a hospital bed. These kinds of systems are particularly well developed in diabetes medicine, hypertension management, asthma monitoring, and home dialysis.

The second category covers *systems for education and dissemination of health related information and knowledge* including web-portals and specific health-related Web sites, virtual hospitals, and Internet-based consultation services. These systems may be targeted for:

- **Medical consultation:** Search for the second opinion, search for health-, disease-, or treatment-related information,
- **Medical education and dissemination** of medical publications, preventive materials, and public health related information.

These dissemination systems may help citizens to become informed and empowered through information and knowledge they are able to retrieve and access themselves from the Internet sources. However, the quality and validity of information can be questioned with many Internet information sources (Wilson, 2002). To promote these kinds of Internet information sources and to ensure people that they can confidently and with full understanding of known risks access and use information from the Internet, initiatives like Health on the Net, HI Ethics Program, and Health Online Actions have been established to develop guidelines and quality criteria and to promote codes of ethics for health-related Web sites (Spink et al., 2004; Wilson, 2002).

The third category covers *systems that are developed to trade health related products.* e-commerce or e-trading of medical products, health-related goods, pharmaceuticals, and medical devices is a growing e-health area, and current procedures enable citizens to enter Internet shopping in an easy and secure way.

E-health applications can also be classified according to their functionality and, following this principle, we may currently find the following groups of applications (Ruotsalainen et al., 2003):

- **Regional health information networks** that deploy advanced healthcare services at various levels of healthcare delivery system, including primary care, prehospital health emergency management, and hospital care. These systems are networked and implemented using various technologies. A typical feature for these systems is the integration of existing legacy systems, imaging systems, departmental systems, and administrative systems into one network and development of new, innovative interfaces and applications to provide comprehensive services regionally.
- **Hospital systems, clinical systems, diagnostic systems, hospital management systems** that cover hospital information systems (HIS), various departmental systems like radiology information systems (RIS), pathology information systems, laboratory information systems (LIS), diagnostic systems like decision support, and knowledge-based systems and hospital management systems like accounting, resource management and booking systems.
- **Telemedicine or teleconsultation systems** are used to access an expert opinion, a second opinion, or to monitor remote patient at home or in another healthcare organisation. These systems are especially often planned to support delivery of medical expertise for rural areas.
- **Insurance, cards, or systems that present the payer's view on health services:** These applications are developed to support the use of cards as means to access health services, to get information on the health insurance status, and to register users that are entitled to special services or reimbursements based on their health, age, or employment status.
- **Citizen-centred systems, patient-centred systems and health information portals:** These applications provide health-related

information for patients and health professionals, and additionally they may provide possibilities for consultation services or for buying pharmaceuticals or other health-related products.

- **Home care systems and health-related fitness systems:** Home care systems are meant to monitor chronic diseases at home, to monitor elderly patients, or to teleconsult professionals from home. These systems are often based on wireless technology, such as mobile phones or handheld computers. Home care systems are also used to help in the management of care, in preparing care plans and in coordination of actions and tasks taken by members of care teams. Health-related fitness systems are those meant for healthy people that want to monitor their well-being and fitness.

We see from these typologies that enablers for most e-health systems are the electronic patient records and Internet-based technologies. Electronic patient records make it possible to share medical information between the care providers across the health strategies and medical disciplines and facilitate consultation between care providers on a given patient. The electronic patient records also give possibilities for further networked applications such as electronic prescribing and integrated regional health information networks (Ruotsalainen et al., 2003). However, current electronic patient records still lack a uniform infrastructure for data exchange and systems are heterogeneous and lack agreed and shared vocabularies (Safran & Goldberg, 2000). This restricts sharing of information. Today many e-health systems are building more and more on e-commerce and e-government strategies and experiences on how to use Internet technologies to redesign operation of public services.

E-Health Systems Use by Citizens and Health Professionals

The use of the Internet by citizens and also by health professionals is increasing worldwide (Budtz & Witt, 2002; Fallis & Fricke, 2002; Gruen, 1999; Holliday & Tam, 2004; Rodrigues & Risk 2003; Silber, 2003; Wilson et al., 2004). The Internet is used to access health-related information by people who are ill but also by people who are healthy who look for advice on healthy lifestyles, diets, habits, and health-related products (Jones et al., 2001). Accessing health information is one of the most common reasons for using the Internet: 50% to 75% of Internet users have used it to search for health-related information (Powell & Clarke, 2002). In 2003 approximately 62% of American Internet users searched the Internet for health-related information (Spink et al., 2004). In some cases citizens do not use tailored health-related Web sites but general Internet search engines to access health information. In 2003 40% of Europeans used the Internet to access healthcare services and information (Wilson et al., 2004). Budtz and Witt (2002) showed that of 93 patients in a study, 20% had used the Internet to get health information. In another study, Larner (2002) followed patients for 6 months and found that more than 50% of them had Internet access, and 82% of them were interested in accessing Web sites with relevant medical information. People using the Internet for health information searches reported that they value the convenience, anonymity, and volume of online information (Powell & Clarke, 2002). The patients who search for specific information on their diseases, diagnoses, and treatments report that it is beneficial to have information, advice, and social support from the Internet (Potts & Wyatt, 2002).

In 2002, on average 78% of European general practitioners were online (E-Health, 2004) and 48% of them used electronic patient record systems and, to some extent, other information systems to receive laboratory results and to transfer patient data to other healthcare organisations. Even 36% of general practitioners used telemedicine systems for home monitoring via the Internet or e-mail (Wilson et al., 2004).

When analysing the use of e-health services and health information sources through the Internet, the key findings are that citizens want to have more information and they want better information (Wilson et al., 2004). The use of the Internet for health purposes is rising and citizens would like to have guidance from the health professionals regarding quality Web sites (Silber, 2003). The major reason for the citizens to use Internet information resources is to know more, to be able to ask more precise questions from health professionals, and to understand better health and well-being. Internet health resources support citizens to become better informed and knowledgeable, and through this they also may take more responsibility of their health and well-being.

However, the Internet and other new technologies will interfere with the communication between health professionals and between them and the patient. Impacts of the Internet use do not only come from the communication but from the operational availability of information resources. Holliday and Tam (2004) found that more than 90% of patients had expressed their wish to communicate with their physician via e-mail, but physicians were reluctant to do so. Physicians explained that the physician-patient confidentiality, time concerns, and increased exposure for malpractice were the major reasons for their reluctance. Physicians are somewhat opposing the use of the Internet, especially when it interferes with their decision-making activity (Gerber & Eiser, 2001; Kleiner, Akers, Burke, & Werner, 2002). There seems to be two reasons: A question of power and a question of danger. A question of power implies that physicians want to keep the control on the medical activities, and the question of danger implies that they want to keep control on actions and decisions that are their responsibility. Some physicians even fear that they become technical executors of decisions taken from third parties (e.g., Internet information sources).

Expected Benefits by E-Health Systems

Many studies (Alvarez, 2002; Ruotsalainen et al., 2003; Iliakovidis, Wison, & Healy, 2004) identify the promises of e-health to build on the advances in ICT to support the development of the healthcare infrastructure. Healthcare services are expected to be better accessible and data available any place and any time, independent from where data are stored or created.

Healthcare professionals expect that e-health systems have remarkable impacts on healthcare routines and practices (Eng 2001; Grimson et al., 2000; Ruotsalainen et al., 2003; Wilson et al., 2004). Reliable and accurate information would be available easily and rapidly at the time and place where it would be needed. It would be possible to view information on the prior history of the patient and on diagnostic investigations to avoid redundant testing. Communication between the healthcare providers would be independent of their physical location, and this results in time savings and increased accuracy of diagnosis and effectiveness of care. Health professionals would be able to update their knowledge and expertise though online training sessions and they would be able to consult international colleagues in dealing with particularly complex cases (Ruotsalainen et al., 2003).

Patients and citizens would benefit through better quality health information and services available to them (Ruotsalainen et al., 2003; Silber, 2003; Wilson et al., 2004). The possibility for home monitoring and treatment or follow-up would reduce the need for hospitalisation or travel in order to receive professional care. Patients would regulate with the consent the provision of their data to various healthcare providers. The possibility to control one's own health data would facilitate and increase the mobility of citizens and patients, particularly those suffering of chronic conditions.

E-health systems are expected to harmonise healthcare systems, thus enabling provision of seamless and continuous care (Grimson et al., 2000; Ruotsalainen et al., 2003; Wilson et al., 2004). Interorganisational cooperation would allow sharing of data and information whenever necessary. E-health systems are also expected to result in improved quality and cost-effectiveness of healthcare systems. The fact that healthcare processes would be better documented and thus accountable would provide more possibilities for evaluation and quality assessment, and information could be made available to support decision making and interventions at the public health level.

Eysenbach (2001) listed the 10 *e*'s for e-health systems:

1. Improved *e*fficiency in health care and decrease of costs
2. *E*nhanced quality of care (e.g., by comparisons of care providers)
3. *E*vidence-based in the sense that the effectiveness and efficiency of interventions are proved by scientific evaluations
4. *E*mpowerment of consumers and patients
5. *E*ncouragement of a new relationship between the patient and the physician
6. *E*ducation of physicians through online resources
7. *E*nabling information exchange and communication
8. *E*xtending the scope of health
9. *E*thics as new forms of patient-physician interaction become possible
10. *E*quity to have health care more accessible to all

There is at the same time a threat that those people who do not have money, skills, computers, and Internet connections cannot access e-health services.

Despite of the availability of e-health systems and their potential benefits, e-health systems are not yet widely used in healthcare practices (E-Health, 2004). Iliakovidis et al. (2004) reported some examples of e-health systems which are now entering the European market. However, many promises of e-health research and development have not been fulfilled yet (Ruotsalainen et al., 2003). The major barriers (E-Health, 2004) include lack of commitment by healthcare authorities, missing interoperability of health information systems (which is mostly due to lacking conceptual models), ontologies, and sharable vocabularies. Additionally, the e-health developments by far have been rather technology-oriented and thus the developed systems are not easy to use, technology is expensive and the systems are vulnerable to changes in the environment. Maybe the most important reason for failures is that the socio-technical nature of e-health systems have not been fully considered. Development of e-health systems requires thorough understanding of the health care work practices where systems are to be installed. The needs of the users and the contextual aspects of the systems use should be the starting point for the development (Berg, 1999; Nykänen & Karimaa, 2004).

E-HEALTH VISIONS FOR THE FUTURE

The challenges for e-health are technical, social, economic and political. Eng (2001) pointed out that there is need for strong leadership from health policy makers and need for practical solutions for interoperable systems that are secure, respect confidentiality, and promote the best possible access to health care for all citizens.

In our survey (Ruotsalainen et al., 2003), several of the respondents looked for a comprehensive e-enabled healthcare system, standardised and interoperable, which meets the necessary legal and functional requirements for optimal, seamless, cross-border delivery of services. Implementation of such a system is expected to lead to more equal, accessible, holistic and human-centred healthcare.

In the future we suppose that several Internet-related and other trends will have influence on the design, content, functionality, dissemination, and use of future e-health systems. With globalisation, increasing number of e-health resources will be developed overseas and for global audience (Eng, 2001). Thus, standardisation and cross-cultural factors will become increasingly important.

The digital generation will most likely demand immediate access to information and will rely on online resources to inform health and other decisions (Eng, 2001). Wireless technologies may contribute to the growth of mobile e-health applications for both providers and consumers. Digital television may serve in the future as a cheap and easy platform to offer e-health systems and information for large audiences. Personalisation and tailoring of applications to specific users will put emphasis on privacy and data security issues.

There are many important questions still unanswered with e-health systems: Ethical and legal issues? Who will pay for the use? Who has access to e-health systems? What are the standards, guidelines and good practices for development and use of these systems and technologies?

The basic requirements for the future e-health applications can be derived from the current health informatics and e-health situation. First, e-health applications should serve the users needs, the needs of citizens, patients, healthcare professionals, and healthcare organisations. Second, the applied technology should be user acceptable, cheap and not too cumbersome to use, and vulnerable to changes. Third, e-health applications should be based on the health care information systems infrastructure, they should be integrated with the environment and preferably also be interoperable. Fourth, standards should be applied on design and development. Fifth, security and safety issues have to be solved within the practice and legal frameworks. And finally, all e-health systems and applications should be evaluated to assess the effects and impacts and to find reasons for success or failure.

REFERENCES

Alvarez, R. C. (2002). The promise of e-health: A Canadian perspective. *eHealth International, 1*(4), 1-8.

Berg, M. (1999). Patient care information systems and health care work: A sociotechnical approach. *International Journal of Medical Informatics, 55*, 87-101.

Budtz, S., & Witt, K. (2002). Consulting the Internet before visit to general practice. Patients' use of the Internet and other sources of health information. *Scan. Journal of Primary Health Care, 20*(3), 174-176.

E-Health. (2004). *Making better for European citizens: An action plan for a European e-health area. Commission of the European Communities.* COM (2004), 356, Brussels, Belgium.

Eng, T. R. (2001). *The eHealth Landscape: A terrain map of emerging information and communication technologies in health and health care.* Princeton, NJ: The Robert Wood Johnson Foundation.

Eysenbach, G. (2001). What is e-health? [editorial]. *J Med Internet Res, 1*(2), e20.

Fallis, D., & Fricke, M. (2002). Indicators of accuracy of consumer health information on the Internet: A study of indicators relating to information for managing fever in children at home. *JAMIA, 9*(1), 73-79.

Gerber, B. S., & Eiser, A. R. (2001). The patient-physician relationship in the Internet age: Future prospects and the research agenda. *Journal of Med. Internet Res., 3*(2), e15.

Grimson, J., Grimson, W., Flahive, M., Foley, C., O'Moore, R., Nolan, J., et al. (2000). Multimedia approach to raising awareness of information and communications technology amongst health care professionals. *International Journal of Medical Informatics, 58/59,* 297-305.

Gruen, J. (1999). The physician and the Internet: Observer or participant? *MD Comput, 16*(6), 46-48.

Haglund, H. (2002). *The significance of welfare services and their electronic applications for the business activities in the future*. Retrieved February 11, 2003, from www.etampere.fi

Holliday, I., & Tam, W. (2004). eHealth in the East Asian tigers. *International Journal of Medical Informatics.*

Jones, R., Balfour, F., Gillies, M., Stobo, D., Cawsey, A. J. & Donaldson, K. (2001), The accessibility of computer-based health information for patients: Kiosks and the Web. *Medical Information, 10,* 1469-73.

Iliakovidis, I., Wilson, P., & Healy, J. C. (2004). eHealth. *Current situation and examples of implemented and beneficial e-health applications.* Amsterdam: IOS Press.

Kleiner, K. D., Akers, R., Burke, B. L., & Werner, E. J. (2002). Parent and physician attitudes regarding electronic communication in pediatric practices. *Pediatrics, 109*(5), 740-744

Larner, A. J. (2002). Use of the Internet medical Websites and NHS Direct by neurology outpatients before consultation. *International Journal Clin. Pract., 56*(3), 219-221.

Nykänen, P., & Karimaa, E. (2004). Success and failure factors during development of a regional health information system: Results from a constructive evaluation study. *Methods of Information in Medicine* (accepted for publication).

Potts, H. W. W., & Wyatt, J. C. (2002). Survey of doctors' experience of patients using the Internet. *J Med Internet Res, 4*(1), e5.

Powell, J., & Clarke, A. (2002). The WWW of the World Wide Web: Who, what and why. *Journal of Med. Internet Res., 18, 4*(1), e4.

Rodrigues, R. J., & Risk, A. (2003). eHealth in Latin America and the Caribbean: Development and policy issues. *Journal of Med. Internet Res., 5*(1), e4.

Ruotsalainen, P., Nykänen, P., Doupi, P., Cheshire, P., Pohjonen, H., Kinnunen, J., et al. (2003), *The state of eHealth in Europe* [Rep. No. D1]. Stakes, Helsinki: MEDITRAV-project, EU IST 1999-11490,

Safran, C., & Goldberg, H. (2000). Electronic patient records and the impact of the Internet. *International Journal of Medical Informatics, 60,* 77-83.

Silber, D. (2003). *The case for ehealth*. European Commission, Information Society, eHealth Conference, Atlanta, Belgium.

Spink, A., Yang, Y., Jansen, J., Nykänen, P., Lorence, D., Ozmutlu, S., et al. (2004), A study of medical and health queries to Web search engines. *Health Information and Libraries Journal, 21,* 44-51

Wilson, P. (2002). How to find the good and avoid the bad or ugly: A short guide to tools for rating quality and health information on the Internet. *BMJ*, *9*(324), 598-602

Wilson, P., Leitner, C., & Moussalli, A. (2004). Mapping the potential of eHealth. Empowering the citizen through eHealth tools and services. *eHealth Conference 2004.* European Institute for Public Administration. Maastricht, The Netherlands.

This work was previously published in E-Health Systems Diffusion and Use: The Innovation, the User and the USE IT Model, edited by T. A. M. Spil & R. W. Schuring, pp. 281-293, copyright 2006 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 8.2
Healthcare Organizations and the Internet's Virtual Space:
Changes in Action

Stefano Baraldi
Catholic University, Italy

Massimo Memmola
Catholic University, Italy

ABSTRACT

For some years now, the opportunity of innovating business models has basically been linked to continual progress in ICT. Healthcare is no exception; information and communication technologies are generally considered the most effective driver for changing organizations, improving quality, optimizing resources, and so forth, at least in theory. In practice, it is not clear which and how many of these opportunities are really exploited by organizations operating in healthcare. This chapter presents the results of a research project aimed at understanding to what degree and how Italian healthcare organizations make use of the virtual space made available to them by the Internet.

INTRODUCTION

"Who doesn't know the Internet?!"

This is certainly one of the most common answers when hospital patients are asked about their knowledge of the Internet phenomenon.

In fact, it is likely that the word Internet is one of the most common and widely used in Western societies. Whether walking in the streets of New York, Paris, London, or Rome, if passersby are asked the simple question, "What is the Internet?" the answer is always the same: The Internet is something that enables us to communicate, read, learn, play, purchase goods and services, make transactions, and more.

The answer does not vary. Often, in fact, people do not know the technology behind the Internet, but they know what it can be used for. In other words, people know how to use technology even if they do not know how it works (which can be said for many new technologies).

Certainly, new technological developments and their awareness can only increase further thanks to the spread in recent years (or rather in recent months) of mobile communication devices. Equally, it is easy to predict the same happening in developing countries: Because of its low cost, the Internet is often preferred to the telephone for communication with these countries.

The Web, therefore, has changed our lifestyle, our habits, and our way of working, interacting with our acquaintances, and, in few words, dealing with so many aspects of our daily life. We no longer have to physically go to the bank but can transfer money and check our account balances by logging on to our bank's Web site. In the same way, booking our holiday is easier when we can click on the site of a tour operator. The list goes on and on: We can shop online instead of at our local supermarket, or buy books and DVDs from a company abroad.

In the world of healthcare, this revolution has yet to be realized, at least fully. In recent years, literature has often indicated how the digitalization of clinical, organizational, and management processes of health structures brings undeniable benefits both for the efficiency and effectiveness of the company, as well as improving the quality of service offered to patients (Coile, 2002; Goldstein, 2000; Nicholson, 1999). More particularly, the Internet considerably affects the entire process of the creation of the value of a healthcare organization (HCO) if we consider the following:

- It gives the organization a global presence
- It means it can offer more services that are more readily available to more users
- It allows the collection and elaboration of a greater amount of information
- It provides a new channel, the World Wide Web, through which it is possible to transfer information, provide services, make transactions, and create a privileged area of interaction between physicians and patients

The possible applications in a healthcare organization, limited only by the state of the art of technological development, can be schematized according to their functional characteristics.

THE INTERNET AS A TOOL FOR REDUCING GEOGRAPHIC DISTANCES AND AIDING THE DISTRIBUTION OF SERVICES OF A HEALTHCARE ORGANIZATION

The opportunity to provide services in inaccessible areas is one of the aspects of the Internet that has multiple applications for healthcare:

- **Home monitoring:** This service allows the patient to be cared for at home, or in the most convenient place, either through prearranged or automatic transmission of signals and vital parameters, or through the sending of alarms activated by predefined emergency situations. Healthcare personnel can also call back and give instructions to the patient or to semiautomatic equipment.
- **Telemedicine:** This is the provision of a treatment service or remote assistance in the form of tele-consultation, tele-presence, second opinions, Electronic Medical Record, (EMR), and so forth. Generally, these services are offered via the Web and may involve videoconference connection between the patient and the doctor or between medical personnel.

The Internet as a Tool for Sharing Information

The Internet is a special way to exchange information:

- The Internet can be used to share clinical information, both for applications developed to be shared totally or to be shared partially and securely. This can include information contained in clinical records (electronic health records, EHRs) to be shared with healthcare personnel belonging to different organizations who treat or have treated the same patient using secure electronic post services (encoded and authenticated).
- It can be used to research and develop new diagnostic-therapeutic processes or protocols aimed at providing welfare assistance through the use of applications similar to those already developed.
- It can also be used for epidemiologic observation through the development of an information flow for the rapid collection and distribution of epidemiologic information for healthcare companies and organizations.

The Internet as a Support Tool for Management and Training of Human Resources

The technologies of information and communication provide an effective tool the following:

- Running and managing the organization (business intelligence, data warehouse systems, etc.).
- Distance training of operators, providing considerable possible savings in costs and significant benefits for the levels of services offered.
- Managing the organization's information system and integrating, through the Internet, departments and organizational units.

The Internet as a Tool for Reducing Informative Asymmetry

Informative asymmetry is characteristic of any service-providing organization. The transmission of clinical information through the Web allows patients to be guided and informed about their health, and provided with expert advice about illnesses, prevention, and treatments available.

In this way, patients become more aware of the progress of their treatment and the traditional physician-patient relationship is transformed. However, there is still the problem of appointing an authority to guarantee the quality of the information provided.

The Internet as a Tool for Reducing the Cost of Transactions

The most popular service offered is that for making appointments online for diagnostic examinations or specialist treatments through an SBC (single booking centre) set up for a single healthcare organization or for a larger catchment area such as a city or even a region. Other possible applications include the possibility to access the results of diagnostic investigations online, to search for a medical specialist for the treatment of a pathology, to download required documentation for requests and reports, and so forth.

The Internet as a Multipurpose Tool with Virtually Infinite Capacity

The opportunity to provide services through the Internet to several users simultaneously without the physical presence of the doctor has meant that many other operations in the chain of value of a healthcare organization have been revised; this has been shown in many of the solutions described previously (telemedicine, home treatment, online booking, clinical information services, etc.).

In light of this, it is clear that the Internet offers the world of healthcare enormous potential, which can be exploited to improve both efficiency (helping reduce some of a country's highest public costs) and effectiveness (with a consequent improvement in the level of the quality of services offered). There are, however, factors that may limit the spread of Internet solutions or, at worst, lead to the failure of projects under way.

For patients, a highly negative factor is the still-limited access to the Web for many potential users (even in economically advanced countries) and what can be defined as electronic illiteracy, particularly for low-income families. The difference in terms of the needs of those who have access to the Web and those who do not, which is destined to increase, will make it even more difficult for healthcare organizations to decide what to offer. Furthermore, some patients are reluctant to use the Web in so far as they are unable to perceive the qualitative level of the information and services available or simply because they prefer to maintain a more personal relationship with their own physicians.

It is physicians themselves who can be the pioneers of the new technological standards of healthcare when they accept the Internet and ICT solutions in their work and consequently redefine their own competencies. It is not impossible to foresee that in the future the teaching of medicine will involve computers and Web tools for distance learning.

Finally, there are still healthcare organizations and managers who run them who are unable to understand how or why to use available technology, or, more importantly, how to evaluate what the returns for investments could be. In particular, whoever is in charge of an organization (production or healthcare) should be aware that introducing the technology is never an end in itself, but should be regarded as the result of a careful strategic plan through which a thorough evaluation of the real, physical, and virtual consequences of some activities of the chain of value is produced. The benefits, in cost reductions, that may come from the introduction of virtual processes in an organization might, in fact, be cancelled out by the greater costs that other units in the organization accrue.

Moreover, it is also relatively easy to find examples of failure, often caused by a lack of strategic vision or knowledge of how the Internet (and more generally ICT technologies) can change the traditional basic rules of treatment services (Given et al., 2001a; De Luca & Enmark, 2000; Fattah, 2000; Flory, 1999; Glaser, 2002; Minard, 2001, 2002).

On the other hand, several factors (including health demand, medical developments, reduced resources, and increasing competitive pressure) lead us to consider ICT as one of the fundamental means for solving the paradox of doing (ever) better with less. In healthcare, too, the key question seems to be not whether to deploy Internet technologies, but how to deploy them (Given et al., 2000; Porter, 2001).

Starting from this point of view, we carried out a research project (named Health.Net) in order to understand to what extent and how Italian HCOs actually exploit and use the virtual space offered by the Internet.

OBJECTIVES AND RESEARCH METHODOLOGY

In our research, we have tried to focus on the real ability of HCOs to use the Internet to support their organizational, clinical, and management processes (Given et al., 2001b; Alemi, 2000; De Luca & Enmark, 2000; Solovy & Serb, 1999), and to create value (Glaser, 2000) not only for their own users, the patients, but also for all their stakeholders (physicians, staff, students, other HCOs, etc.).

In particular, we tried to shed some light about these issues:

- To what extent and how do Italian HCOs inhabit the Internet?
- What kind of strategy is guiding Italian HCOs in creating their own visibility on the Web?
- What are the results achieved? How much of the opportunities offered by the Internet have been actually exploited?

- What progress is being made by Italian HCOs in the exploitation of Internet potential?
- What are the latest trends and the scope of change?
- Is there any difference or delay in regard to other national healthcare systems?

The project has been carried out in two separate steps: the first in 2002 and the second in 2005. Both in 2002 and in 2005 we tried to accomplish the following:

1. Position the Italian healthcare system in an international context.
2. Analyze the public presence of Italian HCOs on the Internet.

In each year, the first stage of the project (aimed to position the Italian healthcare system in an international context) has been carried out by comparing Italian HCOs with the following:

- A sample of 121 companies in six foreign countries (United States, Canada, United Kingdom, France, Germany, Spain); we paid attention to the biggest structures (i.e., those having over 200 beds) in the capital city or in a similar city in terms of healthcare facilities (Boston, Toronto, London, Paris, Frankfurt, and Milan). HCOs were identified thanks to data displayed on the Internet sites of national ministries.
- Nine HCOs acknowledged as benchmarks in the international literature; benchmarks were identified according to the conclusions drawn in a study carried out in the United States (Goldsmith, 2000). The benchmark Web sites achieved the best performance in terms of contribution to business growth and patients' satisfaction.

During the second stage, we analyzed the presence on the Web of all the Italian HCOs (both public and private) that are officially registered in the database of the Italian Ministry of Health.

In order to give greater significance to the information gathered, it was decided, nevertheless, to consider only autonomous structures as much from a legal point of view as an economic one, in other words, only those organizations that have registered (or can register) a domain with their own name. Thus, for organizations dependent on a local health authority, we considered only the site of the latter; equally, the same approach was used for private structures that are subsidiaries to larger companies.

1,002 organizations were identified in 2002, and 990 in 2005. The reduction in the number of organizations usable for research depended on the following:

- The mergence of some hospital structures by local health authorities decided by some Italian regions.
- The setting up of a process of concentration in large national private healthcare organizations.

Each organization under study was classified as one of five main types: local health unit (LHU, 20%), independent hospital (IH, 10%), scientific institute for research and care (SIRC, 4%), university polyclinic (UP, 1%), and private structure (PS, 65%). The percentage of the total for each type did not change between 2002 and 2005.

For each of these organizations, a preliminary investigation accomplished the following:

a. Identified the HCOs' Web domains by means of the most common (national and international) search engines; in case of broken (no) links, companies were contacted by phone.
b. Took note of Web site status (active, not active, under construction, partially under construction).

At the end of this preliminary phase, 473 sites were identified (47% of the total) in 2002 and 574 sites (58% of the total) in 2005.

In the first year, successive elaborations of the research concerned all the organizations present on the Web. In 2005, we decided to take a sample of 136 sites of the 574 originally identified. The statistical technique used to describe the research sample is that of stratification (Frosini, Montinaro, & Nicolini, 1999).

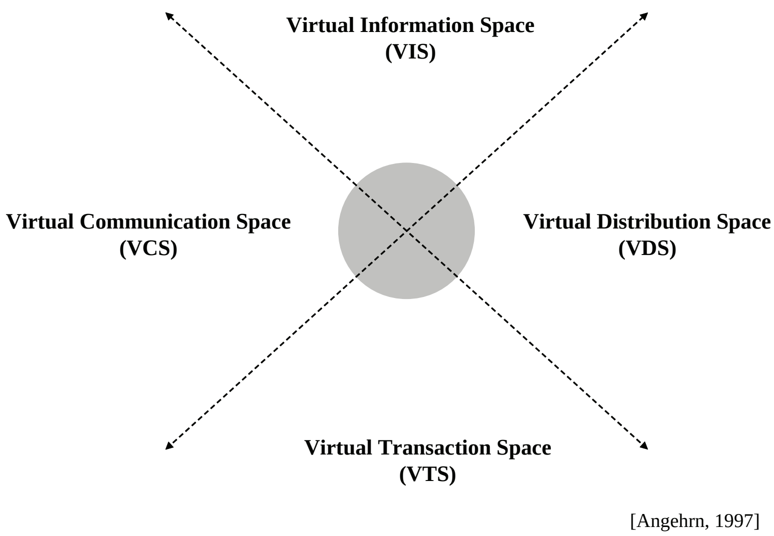
Using this technique, it is possible to obtain an increase in the efficiency of a sampling plan as it reduces the sampling's order of greatness of error without increasing the number of samples.

The population was separated into strata internally as homogeneous and heterogeneous as possible. In other words, the population was divided into *n* underpopulations. From each stratum or underpopulation, a single random sample was taken; there were therefore as many samples as strata. The numerousness of each stratum was decided according to the variance that the stratum presented with reference to the data from the first year, for which a higher number of samples of strata with a greater internal variability were considered. The total number of the samples is obviously equal to the sum of the numbers of each single stratum.

To evaluate not only how many HCOs are on the Web but, above all, the features of their Net presence, we used the ICDT model (Information, Communication, Distribution, Transaction; Angehrn, 1997). Angehrn states that companies should look at the Internet not as a simple communication tool by which information can be exchanged (Nicholson, 1999), but rather as a space in itself, virtual of course, which they can "colonize" in different ways and to different levels, following a strategic vision or not. In this sense, the ICDT framework (Figure 1) effectively works to evaluate which kind of strategic approach is actually guiding HCOs in the conquest of their space on the Web, and, to sum up, how and to what extent, and how successfully, they exploit the opportunities offered by the Internet.

The model takes its name from the segmentation of Internet virtual space into four areas: a virtual information space (VIS), a virtual communication space (VCS), a virtual distribution space (VDS), and a virtual transaction space (VTS). This segmentation emphasizes that the "Internet has extended the traditional market

Figure 1. The ICDT model



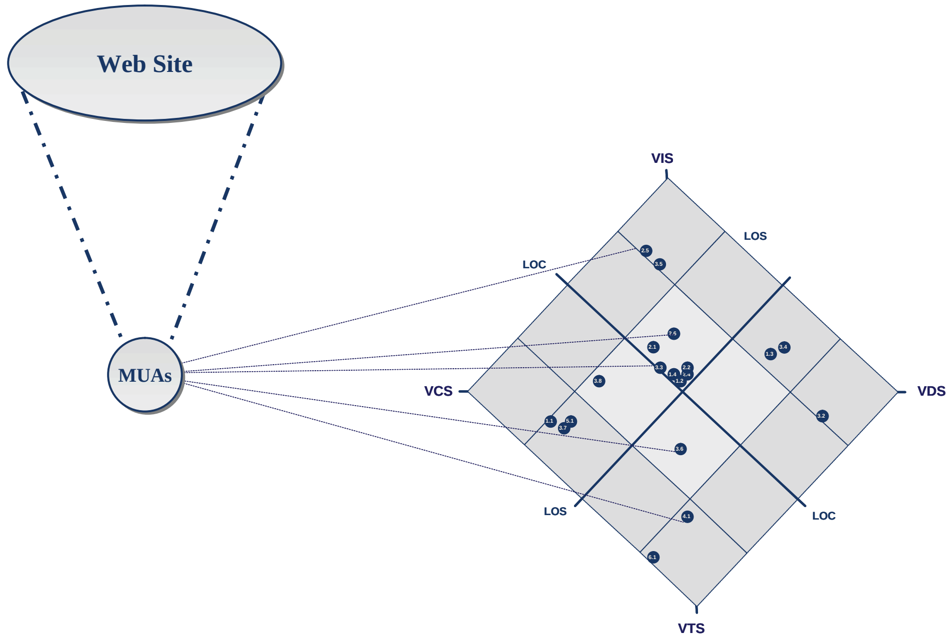
space by providing new spaces in which economic agents can interact by exchanging information, communicating, distributing different types of products and services and initiating formal business transaction" (Angehrn, 1997, p. 362). Some particularities are as follows:

- The VIS consists of the new Internet-based channels through which economic agents can display information about themselves, and the products and service they offer. A hospital can provide patients and any other user with different kinds of information, such as its structure and history, information about how to reach the hospital, available services and clinical specialties, access modes, research projects, and so forth.
- The VTS consists of the new Internet-based channels through which agents can exchange formal business transactions such as orders, invoices, and payments. An HCO can refer a patient to remote transactions such as checkup or hospital-admission online booking, diagnostic-reports consultation, service payments, or medicine purchase.
- The VDS represents a new distribution channel suitable for a variety of products and services. It is a Web space by which companies can provide e-services or e-products. Using such space effectively definitely calls for modifications in the traditional physician-patient relationship as it hosts a wide range of services in telemedicine (teleconsultations, tele-diagnoses, remote access to health-improvement or health-monitoring programs), in prevention, in education, and in information on medicines.
- The VCS is the extension of traditional spaces in which economic agents meet to exchange ideas and experiences, influence opinions, negotiate potential collaboration, and so on. It is an opportunity to create a virtual community in which users can exchange ideas and opinions. Forums, chat lines, and newsgroups can promote debate and discussion among patients and/or between physicians and patients about their own experience, pathologies, and related issues.

We mapped the HCOs' presence on the Web according to their approaches and modes to actually use these different spaces (Figure 2). In particular, we paid attention to the following variables:

- **Complexity**, that is, how far HCOs turn to the Internet to support their institutional activity; an intensive and extensive use of such technology is likely to result in a rich and highly informative Web site that can provide users with corporate information (VIS), services (VDS), interaction and communication (VCS), and transactions (VTS).
- **Typology**, that is, the more or less balanced way in which HCOs exploit the Internet virtual space, achieving a presence in some rather than in all the different kind of spaces (VIS, VTS, VDS, VCS) of the IDCT model.
- **Quality**, measured by considering the level of sophistication (LOS) and the level of customization (LOC) of their Web sites. These two indicators were chosen in order to evaluate the concept of quality objectively and concretely rather than subjectively and vaguely. LOS refers to the technological level (indicating any multimedia elements), while LOC determines the degree of customization of the site content in accordance with the user's specific needs.
- **Effectiveness**, in terms of HCOs' Web site visibility (hits obtained using the most common search engines) and success (number of visitors).

Figure 2. Mapping the Internet presence



Thus, the presence of HCOs on the Internet has been evaluated by the study of the MUAs retrieved from their Web sites (see Table 1). An MUA is a Web site area featuring homogeneity in content and thus embodying a precise territory in the Internet virtual space. Each MUA may spread over more than a single Web page or share the same Web page as other MUAs. Each MUA has been positioned with regard to the following:

- Within its own virtual space of competence (VIS, VDS, VCS, VTS)
- The closer to the margins, the higher the LOS and LOC indexes, which appear on the axes in the graph.

This method has enabled us to evaluate to what degree and how the HCOs that are officially online are really enhancing their presence on the Web. Figure 3 gives as an example the maps relative to the Internet presence of HCOs. Each HCO exhibits one of the following characteristics:

- Although starting with the use of the Internet as a simple, cheap means of communication, it is beginning to expand and significantly improve its presence on the Web, "attacking" other types of virtual space (e.g., X).
- It seems to be interested only in some of the opportunities offered by the Internet in various virtual spaces (e.g., Y).
- Although present on the Internet, it is unable to fully exploit the Web's potential (e.g., Z).

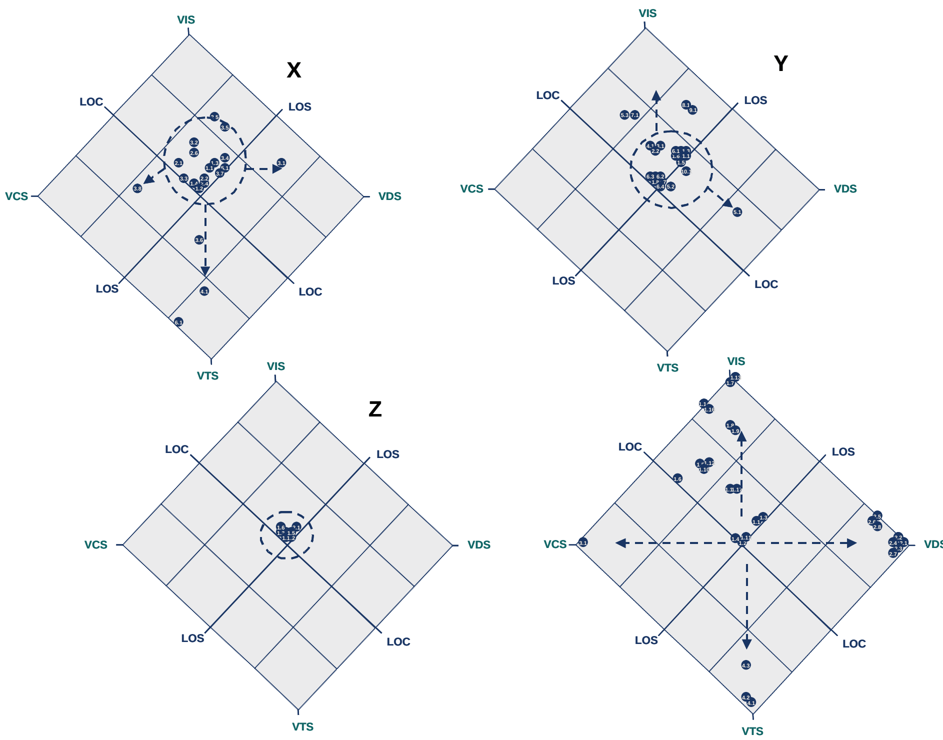
The different types of behavior shown by HCOs in occupying the Internet's virtual space have been grouped into three main clusters:

- The **pioneers**, whose results are better than the national average. These are the HCOs that are more convinced of the potential offered by the Internet and therefore exploit it to a greater extent.
- The **followers**, whose results are in line with the national average. These are the HCOs

Table 1. Methodology for analyzing the Internet presence

Step 1: Preliminary research	• identify Web-site domain by search engine or contacting the HCO by phone • verify Web-site status (active, not active, under construction, partially under construction)
Step 2: Internet presence complexity	• examine Web-site contents and classify them in minimum units of analysis (MUAs) • assess the existence of other structural features (foreign-language version, native-language search engine, Web-site map, extranet)
Step 3: Internet presence typology	• position retrieved MUAs in corresponding virtual space areas (VIS, VDS, VCS, VTS) • verify Web-site use of Internet potential
Step 4: Internet presence quality	• assign the LOS of each MUA • assign the LOC of each MUA
Step 5 : Internet presence effectiveness	• verify Web-site visibility by means of the most common search engines • measure average Web-site traffic

Figure 3. The Internet strategy of an HCO



that, compared with the other members of the sample, are starting to expand and improve their presence on the Web to a greater extent.

- The **latecomers**, whose results are below the national average. These are the companies that, although present on the Web, only exploit its potential to a very limited extent.

MAIN FINDINGS: INTERNATIONAL SURVEY

The Complexity of Internet Presence in the International Sample

There is a considerable distance between the benchmark and the other organizations that make up the international sample. Organizations in the United States and Canada perform best both for the quantity of contents offered on their sites and for the index of complexity. From 2002 to 2005 there was a substantial increase of values assumed from the variables investigated (particularly in the number of MUAs), with a reduction of the gap between European healthcare organizations when compared with those in the United States and Canada.

Through this analysis, we wanted to examine the consistency of the Internet presence of healthcare organizations, not simply according to the number of pages of content on the site, but rather for the value of the content offered in order to achieve the following:

- Give information in the VIS area.
- Deliver services in the VDS area.
- Provide a virtual communication space in the VCS area.
- Accomplish transactions in the VTS area.

Evidence shows that richer content in the different areas is limited to a higher Web site complexity as measured by two indicators: (a) the number of MUAs on the Web site and (b) the structural complexity index (SCI).

The number of retrieved Web site MUAs measures the richness of the content of HCOs' Web sites and ranks their complexity.

SCI is related to a series of structural features intended to improve Web site user friendliness. It is indeed likely that richer content requires greater skill of the user (i.e., patients, physicians, students, or any other stakeholder) in searching for desired information: In such cases, a native-language search engine and a Web site map are, no doubt, helpful. The SCI, then, results from the ratio between actual Web site structural features and research-relevant ones, including the following:

- Native-language search engine
- Web site map
- Foreign-language version of the Web site
- Extranet access

An extranet network can be described as a nonpublic presence on the Internet of an organization. Unlike an intranet, which presupposes a totally private exploitation of the network inside the organization and is reserved for the people who work there, an extranet enables information and content to be offered securely outside the organization. In fact, it can be used to share treatment protocols or, more generally, clinical information between hospitals or physicians, or to provide suppliers with information. Therefore, although not being a tool for improving the readability of a site's contents, an extranet does directly affect the degree of complexity of the Internet presence.

In general, what clearly emerges from examination of the complexity of Internet presence is the considerable distance that separates the group of the benchmark organizations from the others making up the international sample. Moreover, this gap, which is mostly consistent in other areas of research analysis, further increased in 2005.

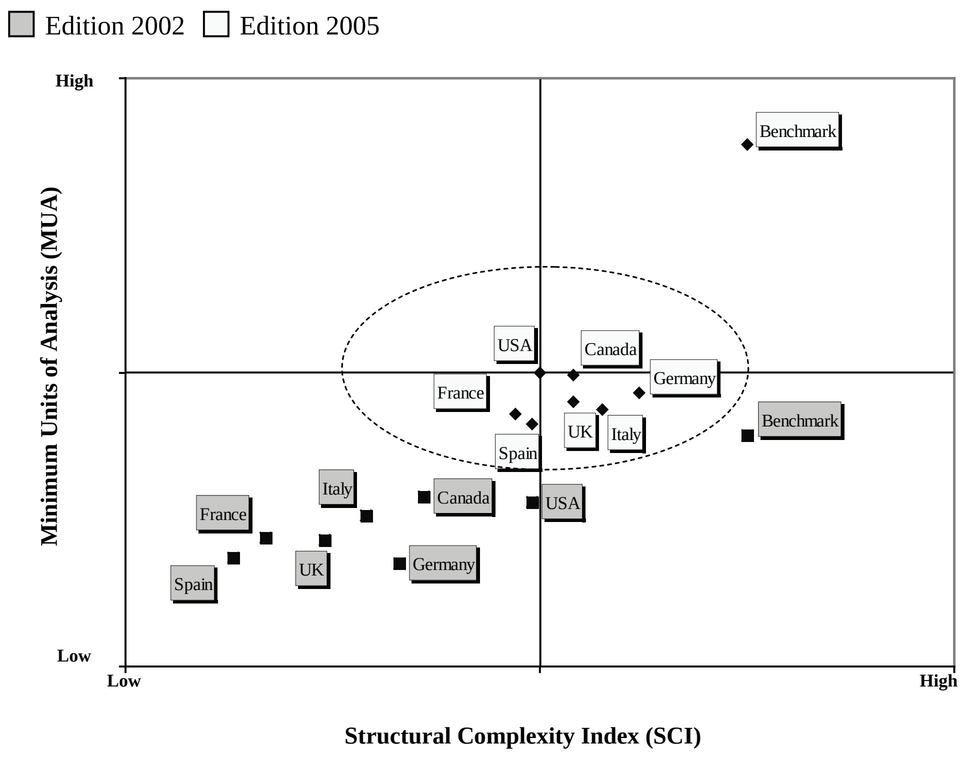
Regarding the quantity of informative contents or services offered by the site, in the year, only the organizations in the United States and Canada performed in any way comparable with those of the benchmark organizations: The average number of MUAs identified was 16.79 for the United States and 17.35 for Canada compared with an average for the benchmark organizations of 23.56. The European average was considerably lower (12.71 MUAs) with the United Kingdom (12.96), Italy (15.47), and Spain (13.2) being more advanced than France (11.17) and Germany (10.75).

Analysis of data from 2005 shows that there was considerable growth in the quantity of content available on sites of healthcare organizations in both Europe and America, but not enough, however, to bridge the gap with respect to the benchmark organizations, which continued their steady growth. Once more, it was the U.S. and Canadian organizations, with an average number of MUAs of around 30 units that were nearer the benchmark 53.22. European organizations, nevertheless, markedly improved their position with an average number of MUAs of 26.32; in contrast to 2002, there was greater homogeneity with no major differences between European countries (Figure 4).

Analysis of the SCI, apart from the theme of the gap between the benchmark and the rest of the international sample (particularly as regards the European organizations), showed for 2002 a particularly high result for U.S. organizations (49%) compared to Europe (23%) and Canada (36%). In Europe, the highest values were in Italy (29%), Germany (33%), and the United Kingdom (24%). Scores for France and Spain were considerably lower at 13% and 17%, respectively.

Here, too, in 2005, there was greater homogeneity of values with an average European figure of 54%, perfectly in line with U.S. organizations at

Figure 4. The complexity of Internet presence of the international sample



50% and those in Canada at 54%; still, however, it is a long way from the benchmark of 75%.

Further interesting points emerge if we examine individual organizational features (foreign-language versions, site maps, internal search engines, access to an extranet).

We now find that the Web sites most likely to have an international slant are those in Italy, Germany, France, and, to a lesser degree, Spain: In these countries, in fact, a foreign-language version of the site is present in 35% of the sites analyzed in 2005. It is also the case that, because of the international importance of the English language (the official Internet language), organizations in the United Kingdom do not much concern themselves with presenting the content of their sites in another language. This factor is probably mitigated in the U.S. and Canadian organizations by the presence of a multiracial society: 29% of U.S. organizations and 15% of those in Canada offer foreign-language versions.

The site map and internal search engine are the two factors that most influence the user friendliness of a site, and nowadays are almost universally present on organizations' Web sites in the international sample; the high number of MUAs activated in sites inevitably creates difficulties in searching for information or services.

The extranet, finally, continues to have a low presence, even at an international level. It is present in little over half of the benchmark organizations (56%), in about a third of those in Canada (33%), a quarter of those in the United States (24%), and a lowly 13% of European ones.

To indicate the complexity of Internet presence, Figure 4 shows three distinct clusters:

- Pioneering organizations that constitute the benchmark and are clearly more advanced than those in the international sample.
- U.S. and Canadian organizations that, despite European organizations making progress, continue to lead, but are still some way behind the benchmark organizations.
- European healthcare organizations, the late arrivals, whose reference values were more homogenous in 2005 and are approaching U.S. and Canadian values.

The Typology of Internet Presence in the International Sample

Only the organizations that make up the benchmark have a widespread and homogenous presence on the Internet, fully exploiting its potential. Despite the progress made in 2005 compared with 2002, the organizations that make up the international sample only partially exploit the Internet, limiting themselves, in the case of European organizations, to a Web strategy with a low to medium profile.

Corporate Web presence enables companies to take advantage of virtual space, whose different areas (VIS, VDS, VCS, VTS) can be variously exploited.

The typology of Internet presence (Step 3) maps HCOs' strategies in colonizing such space in terms of the number of colonized virtual spaces and the actual exploitation of the different kinds of colonized virtual spaces.

The indicators employed in measuring such elements are, respectively, as follows:

- The colonization index, obtained by listing colonized areas (VIS, VCS, VDS, VTS).
- The coverage index, that is, Internet virtual space coverage, both globally and by single area.

The colonization index (how many virtual spaces are actually used by HCOs) is a rather simple, but very useful, indicator for measuring the evolutionary stage reached by an HCO's Internet strategy. However, as areas can be colonized even by a single MUA, this indicator does undoubtedly provide only an approximate value, which has to be completed with the analysis of the coverage index.

While the colonization index measures corporate colonization of single Internet virtual areas, the coverage index ascertains how far companies benefit from it. More specifically, the coverage index by single area is given by the ratio between (a) the number of MUAs in each area of the Web site and (b) the total number of research-retrieved MUAs for each area. The global coverage index is thus the arithmetic average value of all single areas; it measures the overall, though approximate, dimension of how far companies are taking advantage of the Internet's potential.

Regarding the colonization of different sectors of the Internet, it is clear that only the benchmark organizations have a widespread and homogenous presence, fully exploiting the many opportunities that virtual space provides (Figure 5).

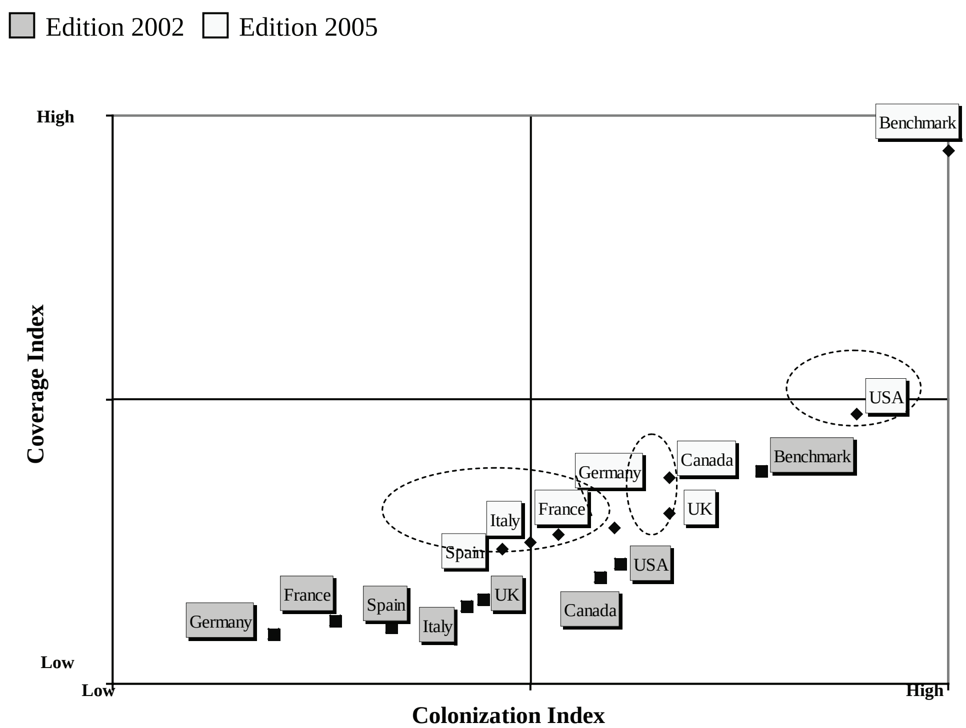
The USA and Canada, as usual, are at a higher level than the European average, in both 2002 and 2005. Organizations in these countries, in fact, occupy three or more of the four areas (VIS, VDS, VCS, VTS) that the Web offers: The average numbers of areas colonized in 2005 were 3.67 and 3.00 respectively. European organizations, on the other hand, adopt a more traditional or low-profile strategy to exploit the Web; the colonization of Internet space remains on average limited to less than three areas (2.66).

Organizations in the United Kingdom (average colonized areas was 3) have a more pervasive Internet presence than the rest of Europe (about 2.5 areas).

Therefore, as often happens, numbers only are not enough to fully describe a phenomenon. Even if, in several cases, development of the presence occurs rather casually depending on contingent circumstances, the logical sequence of colonization of the different spaces, as happens in other sectors, is in this case too, tied to the need to resolve particularly serious organizational crises.

A typical path of development for Internet presence can therefore be outlined as follows:

Figure 5. The typology of Internet presence of international sample



- Entry onto the Internet is by means of the VIS area, which in strategic terms is that which is first attacked by organizations in the initial phase of their presence on the Internet, generally without clear rethinking of their marketing policy, and featuring a traditional approach as regards the media (the traditional "shop window" site); none of the international organizations studied were of this type.
- Later developments involve the VTS area, with the aim of facilitating interaction with patients (online booking of appointments, distance medical reports, complaints and suggestions), suppliers, staff, and so forth; 32% of HCOs have colonized the virtual space made available by the Internet in terms of VIS and VTS.
- Only later, and so far to a limited degree, do HCOs try to provide some of their services through the Internet (VDS area), especially LHUs; 44% of HCOs have managed to increase their presence on the Internet beyond VIS and VTS into a virtual distribution space.
- The VCS area is clearly that which is considered least important, partly because it does not correspond to any of the traditional organizational roles, nor does it provide immediate advantages in terms of optimization of existing organizational processes; 24% of HCOs are currently present in all four types of virtual space provided by the Internet.

The information that emerges from analysis of the coverage index average (Figure 5) shows that, with the exception of the benchmark organizations, the Internet still offers the potential for enormous developments. This is supported by the fact that even with the marked improvements seen in 2005 with respect to 2002 (with percentage variations equal to or over 200% for the United States and Canada, and Europe too), the coverage index remains overall rather low.

There is an abyss between the benchmark organizations and the rest of the world: The ratio of the exploitation of the potential of the Internet is 1:2 for organizations in the United States and Canada (the coverage index in 2005 was 38% and 29% with respect to 75% for the benchmark organizations) and a low 1:3.3 for European organizations (which had an average coverage index of 21%). In Europe, the figures for 2005 were again more homogenous than those for 2002.

An examination of the typology of the Internet presence for organizations in the international sample puts them into three groups (Figure 5):

- The pioneers, comprising the benchmark organizations.
- The chasers, representing organizations in the USA and Canada.
- The latecomers, the other organizations in Europe.

The Quality of Internet Presence in the International Sample

If the benchmark organizations are excluded, once more it is the U.S. and Canadian organizations that perform best for the quality of Internet presence. However, for the user friendliness of informative content and services available on the Web sites, the period of 2002 to 2005 showed a partial closing of the gap by European organizations.

In evaluating the qualitative dimension of Web site content (Step 4), we paid attention to both technical and user-related aspects, such as user-friendliness in information retrieval. Among the several options at hand, research ascertained the quality of Internet presence using the following criteria:

- Level of sophistication.
- Level of customization.

Both are calculated in average terms on Web site-identified MUAs.

The LOS varies according to the technology (in its widest meaning) employed on Web site content editing; its index ranges from 1 to 5 in order to account for intermediate situations between the following:

- MUAs in static Web pages, edited without Internet-specific tools (using, for instance, MS Word, Powerpoint, and similar editors).
- MUAs in highly interactive Web pages, edited using the most recent multimedia tools (such as Flash View animations, audio and video files, pictures, etc.).

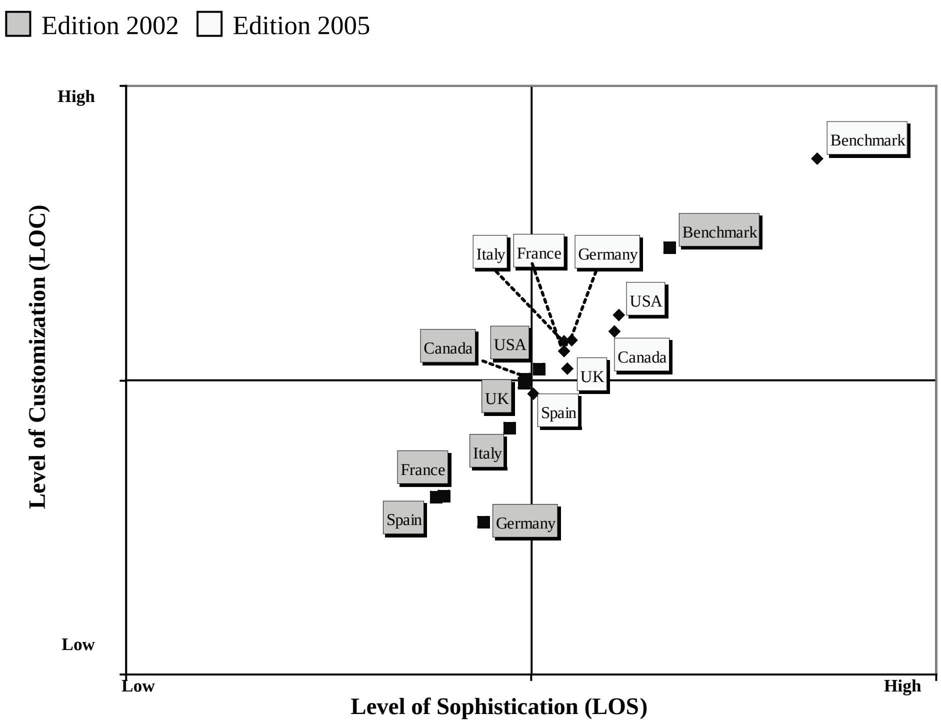
The LOC calculates how far users judge Web site content to fit their needs. In this case, too, the index ranges from 1 to 5 in order to account for intermediate situations between MUAs, as follows:

- MUAs in static Web pages that enable no customization of content and display no additional services.
- MUAs in highly customizable Web pages offering user-profile-related additional services such as e-mail, forums, chat lines, newsgroups, and so forth.

Overall, from the point of view of the level of technological sophistication, there emerges a substantial difference between the organizations that constitute the benchmark and the other healthcare organizations of the international sample, which, as has already been seen in other areas of research, grew further in the period from 2002 to 2005 (Figure 6).

In 2002, in fact, the LOS average for the benchmark organizations was 3.68 against an average value for the United States, Canada, and Europe of 3.04, 2.97, and 2.74 respectively. In Europe, sites of the healthcare organizations in

Figure 6. The quality of Internet presence of international sample



the United Kingdom had values of 2.96, more in line with those of the United States and Canada than with those of the rest of Europe.

In 2005 there was, as usual, a general improvement in values, but it was more consistent for the benchmark organizations (average improvement of 20%) compared with the rest of the sample (average improvement of 15%). More specifically, in the second year of the project, the benchmark organizations had an LOS average of 4.41, which was considerably higher than the average values for the USA (3.43), Canada (3.41), and Europe (3.14). These values indicate that the distance between America and Europe remains largely unchanged. In Europe, the most consistent improvements were in organizations in France (from 2.57 to 3.16), Germany (from 2.76 to 3.20), and Spain (from 2.53 to 3.01).

The same can be said for user friendliness (LOC; Figure 6). Here, from 2002 to 2005 there was greater homogeneity in average values in the USA and Canada compared to Europe. In fact, the average improvement percentage of American and Canadian organizations was 11.5% compared with 30% for Europe. More particularly, the LOC for benchmark organizations rose from 3.90 in 2002 to 4.50 in 2005, for USA organizations from 3.08 to 3.44, for Canadian organizations from 3.01 to 3.33, and for those in Europe from 2.43 to 3.14. Analysis in 2005 did not indicate major differences between European organizations.

Neither were there major variations for the quality of Internet presence (Figure 6):

- Benchmark organizations had an established position.
- Organizations in the United States and Canada were classified as chasers.
- The remainder of organizations in the international sample, mostly European, lags considerably behind the first two groups.

The Effectiveness of Internet Presence in the International Sample

The visibility of the sites of the foreign sample by the main search engines is very high. Generally, the organizations with the best performances for traffic recorded are those that best exploit the potential the Internet offers and consequently their results are higher in all areas of analysis.

HCOs are likely to turn to the implementation of a Web site for basically two main different reasons:

- To maximize their visibility on the Web, that is, rendering themselves available to the highest number of potential users.
- To maximize Web site use by their stakeholders.

As a consequence, their Internet presence effectiveness (and success) depends on reaching such goals. Thus, we decided to measure HCOs Internet effectiveness (Step 5) by jointly considering the following:

- A Web site visibility index, obtained using the most common national and international search engines.
- A Web site traffic ranking as measured by Alexa.com, which offers an international list of sites ranked according to average Web site traffic.

The first indicator measures how simple it is to find Web sites using the most common search engines. All HCOs were searched by name in Italian and international search engines (Virgilio, Lycos, Excite, Iltrovatore, Google, Yahoo, Search.msn, Altavista). Indexes were the ratios between positive responses and search engines' total numbers.

Alexa.com is a site ranking Web sites all over the world according to their traffic. Lists are updated quarterly, measuring both Web site and

Web-page access: The former quantifies Web site visitors per day (displaying its home page), and the latter is calculated considering the number of Web site URLs (uniform resource locators) accessed by the same user. The site hosting the highest Web site and Web-page traffic is ranked first.

It should be pointed out that Alexa.com can only provide an evaluation for the principal domain (e.g., http://www.domain.org or http://www.domain.com). For pages inside the domain (e.g., http://www.domain.org/page.html) or for sub-domains (e.g., http://www.hco.domain.edu), the value given is that of the principal domain. A zero value is given, however, in cases where traffic is so light compared to the average that it is not detectable (usually below position number 6,000,000 in the list).

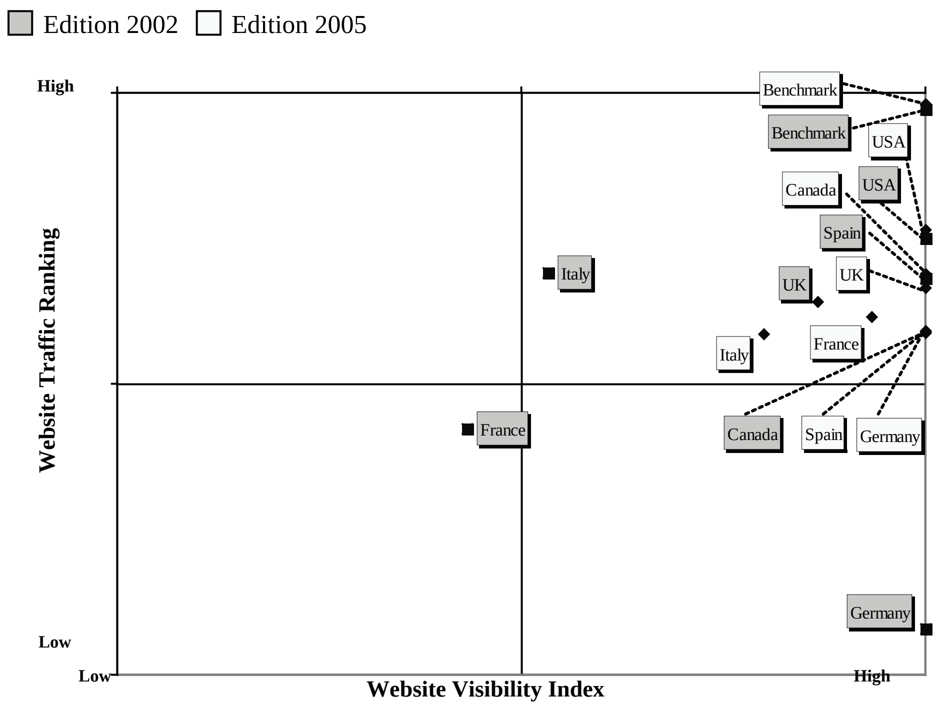
Finally, Alexa.com's evaluation includes all sites on the Web. A process to classify Italian healthcare organizations appropriately has therefore been undertaken.

By evaluating the visibility of organizations in the international sample with the selected search engines, the factor that emerges most clearly is the high traceability both for benchmark organizations and the rest of the sample (Figure 7).

The organizations studied have an almost total visibility in that they are reachable using any of the search engines considered. This, in light of the results of 2005, is the case for the benchmark organizations, for those in the United States and Canada, and almost always for those in Europe. In the latter case, exceptions are Italy and France, although visibility is still high at 94% and 98% respectively.

As regards the ranking of site traffic, generally, those organizations that best interpret the role and potential of the Internet get greater success in return: Figures from Alexa.com for the benchmark organizations are, in fact, considerably higher than for the rest of the organizations in the international sample, both in 2002 and 2005 (Figure 7).

Figure 7. The effectiveness of Internet presence of international sample



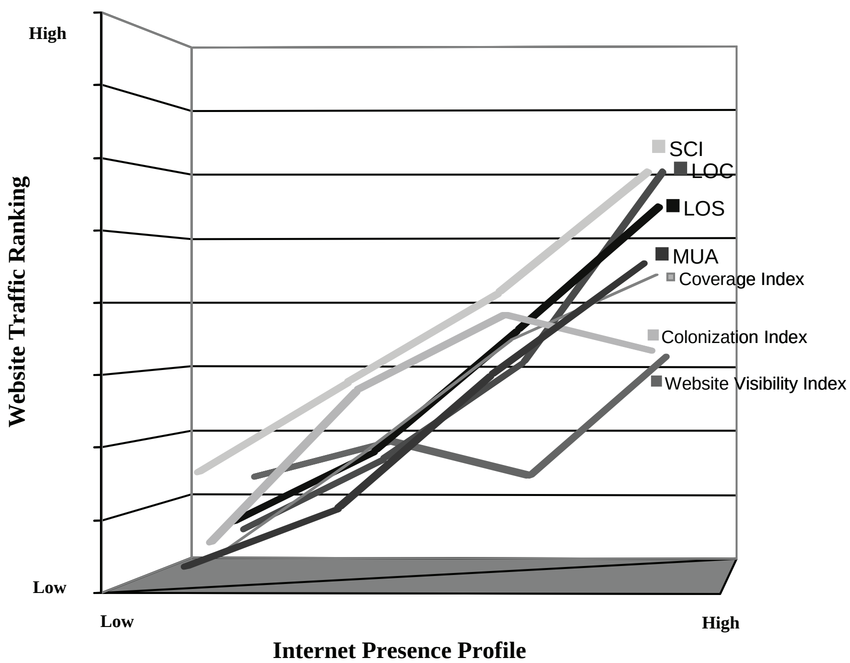
This general trend does not seem to be applicable in results recorded in 2002 for organizations in some countries, as, for example, Canada, Spain, and Italy. Ranking for site traffic, rather low for Canada and on the contrary high for Spain and Italy, are not in line with results obtained in other areas of analysis. In 2002, we attributed this to the following:

- The low population density for Canadian organizations.
- The fact that some sites of Spanish and Italian organizations were guests in the domain of a university or regional organization, and HCO site constituted only a section of the host site; in these cases, Alexa.com's evaluations were influenced by the volume of traffic recorded by the principal domain.

The 2005 research does seem to confirm these suppositions. Unlike in 2002, in fact, sites of organizations in Italy and Spain more often have their own domain that does not affect Alexa.com's evaluations. Recorded performances are therefore substantially in line with those obtained in other areas of analysis with the three types of HCOs:

- The pioneers, or rather the group of benchmark organizations that obtain the best results both in terms of visibility with the search engines and ranking for traffic recorded for their site.
- The chasers, made up of the organizations in the USA and Canada, to which can be added, different from in other areas of research, the organizations of the United Kingdom, which have a markedly better performance for traffic recorded than the average of other European countries.
- A group of latecomers, basically the remainders of the European organizations (France, Germany, Italy, Spain) that are different from the best particularly for traffic recorded for their Web sites rather than for their visibility with search engines.

Figure 8. The ranking variation for Alexa.com vs. the Internet presence profile



Finally, in light of major differences existing between the benchmark organizations and the rest of the foreign sample, it is useful to evaluate whether there is a relationship between the extent to which a site is used and the way in which the organization develops its Internet presence. In this respect, Figure 8 shows the ranking variations for Alexa.com vs. the indicators previously used to evaluate the Internet presence.

As a general, probably predictable, indication, it can be stated that the organizations that best make use of the potential of the Internet are rewarded. This is particularly evident as regards the number of MUAs, the SCI, and the possibility of personalizing the contents of the site.

MAIN FINDINGS: NATIONAL SURVEY

Italian HCOs and their Presence on the Web

The Internet is not yet a phenomenon that is present throughout Italian healthcare: Only 58% of all the organizations have a Web site.

The second part of the project presents a picture that is somehow contradictory. In fact, if the number of structures that are on the Web has not significantly increased (the rise over 3 years was only 11%, from 47% in the first year to 58% in the second), there is now a more homogeneous geographical distribution.

One of the striking facts about the first part of the study was that in the north, 63% of HCOs were on the Web, clearly distancing the centre (43%) and the south (33%). Today, there is still a difference, but the gap has narrowed: The north remains a benchmark with 68% of HCOs online, but the centre (53%) and the south (51%) are catching up.

This kind of heterogeneity between regions is not surprising and is confirmed by other studies concerning the spreading of a managerial culture in Italian HCOs (Anessi Pessina & Cantù, 2003; Baraldi & Monolo, 2004).

This general information has more relevance if analysis is undertaken by company type rather than by geographical area. For public organizations (local healthcare organizations and hospitals), SIRCs, and UPs, the full exploitation of the potential of the Internet is now a reality, regardless of geographical situation (almost everywhere penetration is around 90%); this is not the case for private organizations. These, therefore, seem to hold in much minor consideration the importance of a Web strategy as an essential component of their institutional strategy (the number of organizations online increased from 30% in the first year to 41% in the second).

Probably, public structures have a wider awareness of the potential offered by the Web than private ones because of (a) the recommendations provided by regional healthcare plans, (b) the need to provide information to the general public (e.g., emergency services, waiting lists for specific services), and (c) the attempt to simplify administrative procedures (e.g., booking appointments and giving medical exam results, such as blood tests, online). It should also be borne in mind that the private Italian HCOs are generally small (often with less than 300 beds) and have a limited budget for implementing a high-profile Web strategy.

The Complexity of Internet Presence

There is an increasing depth of presence on the Internet by Italian HCOs: the information available and services activated are increasing as are the tools that allow their content to be best exploited.

If we had to assess the tendency of the phenomenon under discussion based only on this aspect of analysis, we would have to affirm that the Internet is becoming increasingly important for Italian HCOs. The growth of the average amount of information or services offered through the

Internet by Italian HCOs has gone from 10.22 to 17.80, which is a considerable rise of 74%.

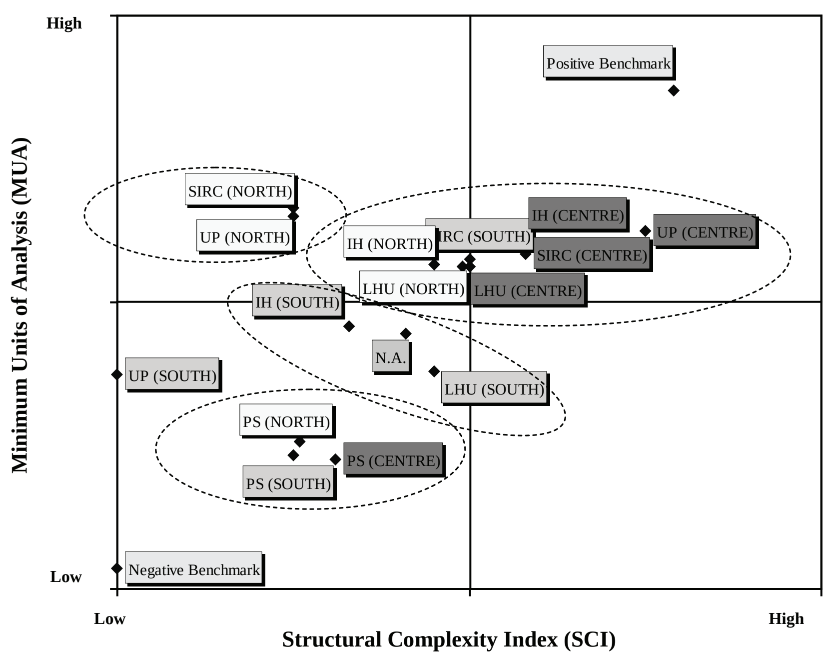
As always, however, more information can be obtained by further analysis. Geographically, despite an overall improvement, there remains a substantial difference between organizations in the north and centre of the country (an average MUA value of 19.2 and 17.33) and those in the south (where the average MUA value is 13.32).

Analysis by type of organization is even more interesting as it makes it possible to outline, in some way, a taxonomy of approaches to the Web strategy according to the individual characteristics of each structure. It affirms that LHUs, IHs, and SIRCs have a presence on the Internet that can be described as established and definitely consistent (the average number of MUAs is respectively 21.76, 23.32, and 25.36, with variation with respect to 2002 of 95%, 72%, and 97%).

The situation is different for UPs and PSs. Regarding the former, of particular interest in the first year was the rather weak, improvised, or inadequate role played by these organizations in the overall picture. The MUA figure of 8.8 was closer to that of private organizations than any other type previously seen. In light of the results of the second year for this and the rest of the analyses, it is confirmed that the presence of UPs on the Net has been transformed: The average MUA value in this sector has increased by 148% reaching 22.00, much more in line with the other main organizations.

For PSs, the trend that emerged in the preliminary analyses already reported in previous pages is confirmed. In brief, considering the results obtained from other aspects of the analysis, private organizations are 3 years behind the other HCOs: The situation shown in the second year of the project (MUA value of 9.70 on average with an increase of 26%) is generally in line with the average figures for the first year. Similar observations to those expressed in earlier lines can be made here regarding the index of structural complexity. Here, too, the average increase

Figure 9. The complexity of Internet presence of Italian HCOs



is consistent (a variation in percentage terms of 128%) by reason of a general improvement in those devices, in particular, site maps and search engines, which allow faster and easier access to the content available.

Considering the combination of the two variables (minimum unit of analysis and index of structural complexity), it is possible to position Italian healthcare organizations regarding the complexity of their presence on the Internet (Figure 9). Included in the HCOs are the following groups:

- A pioneer group of organizations (which includes the LHUs and IHs of the north and the centre, and the SIRCs and UPs of the centre) that invested early in the Internet and that hold that a Web strategy is an integral part of their business strategy and therefore make a substantial quantity of information available, with averages considerably higher than the national average.
- A chasing group of LHUs and IHs in the south that have average figures mainly in line with the national average.
- An anomalous group of organizations comprising the SIRCs and UPs of the north, which analysis reveals has an unbalanced position: The contents of their sites are substantial (above the national average), but they are not backed up by an adequate index of structural complexity.
- A group of latecomers, mainly private HCOs, that have a rather light presence on the Internet, with sites offering basic and essential content, mainly of an informative nature, easily accessible even when lacking navigational aids (measured by the index of structural complexity).
- The negative and positive benchmarks representing the average of 5% of HCOs, presenting, respectively, the best and the worst scores.

The Typology of Internet Presence

The Web strategies of Italian HCOs are still mostly traditional, tending to provide information rather than services: Of the four areas in which the Internet's virtual space can be articulated, in over half of cases only two are used. So, inevitably, utilization of the Internet remains low. Despite steady improvements in recent years, Italian HCOs utilize less than a fifth of the potential that the web offers.

If we consider the results of the second year of the project for this sector, it emerges that Italian HCOs would appear to have some difficulties in getting over this hurdle. In other words, the go-to Internet has undoubtedly occurred and in some instances progress made since 2002 has been consistent (also because in several cases, the starting objectives were on a small scale). In most cases, nevertheless, a clear strategic design is still lacking that sees the Internet not only as a sophisticated instrument of communication, but also a lever through which to carry out company strategy, involving the re-planning of, if necessary, operational methods of healthcare services.

So, progress recorded in the level of colonization of the Internet's virtual space (with an increase of 23%) is undoubtedly a positive, but equally, it should be noted that around 60% of Italian organizations still have a traditional Internet presence based on the distribution or completion of information (VIS).

Investigating the analysis for each type of organization in more depth reveals that there has been consistent progress for LHUs, IHs, and UPs (growth percentages of 33%, 29%, and 43% respectively), rather less for the SIRCs (11%), and absolutely no progress for PSs.

The consistent increase of available content, even if this occurs as previously shown in a rather traditional way, has resulted in a clear improvement in the coverage index with an average increase of 88%.

The percentage variations are really surprising: 133% for LHUs, 90% for IHs, 111% for the SIRCs, a striking 183% for Ups, and, finally, a more modest 40% for PSs.

A glance at the absolute values, though, enables us to put the phenomenon into proportion. The starting values were in fact very low: a national coverage index of 8%, varying between 9% and 10% for LHUs, IHs, and the SIRCs; 6% for UPs; and 5% for PSs.

So, if following the progress made (the national coverage average is 15%), it is easy to see how much more improvement remains to be made.

The overall evaluation of Internet presence typology is thus reached by considering both the colonization index and the global coverage index. Figure 10 shows the following:

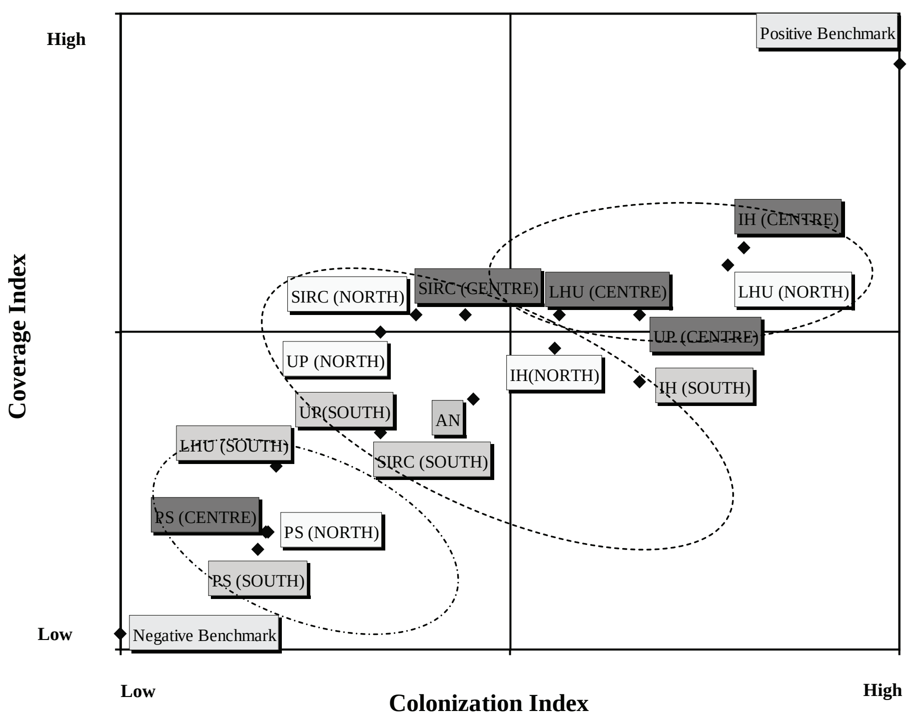
- The pioneers, in this case the IHs, LHUs, and the UPs of the centre to which can be added the LHUs in the north, exploit the Internet in a more homogeneous and broader way when compared with the national average.
- The SIRCs, IHs, and UPs in the north and the south perform mostly in line with the national average and can be classified together with the chasing group.
- The latecomers group comprises LHUs of the south and PSs, which have a less balanced presence (on average in fewer than two areas) and a much lower level of Internet use (particularly in the areas offering greater services such as the VDS, the VCS, the VTS) than the national average.

The Quality of Internet Presence

The quality of Internet presence of Italian HCOs is mainly good regarding both the degree of technological sophistication of the content available and the level of user satisfaction.

In 2002, we recorded for this dimension of analysis good values of reference with differences

Figure 10. The typology of Internet presence of Italian HCOs



both at a geographical level and for the type of organization, basically due to budgets and the consequent possibility or not of utilizing specialized consultants. The second year shows the inevitable improvements following a general pattern common to the other dimensions of analysis.

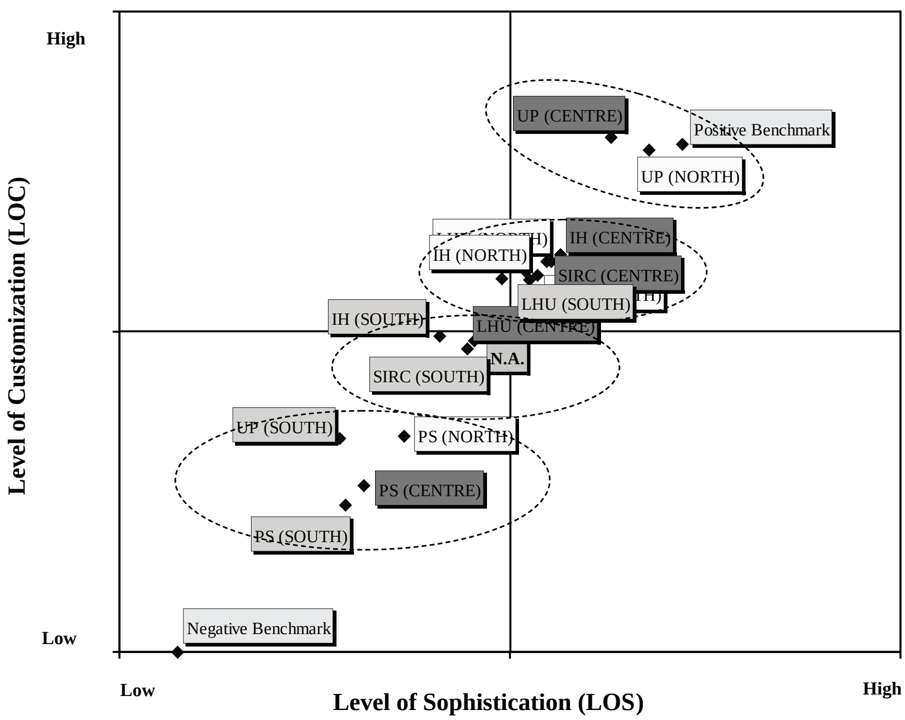
The degree of technological sophistication (LOS) increased by 16% nationally (from 2.44 to 2.82), with the best performance coming from the UPs (a rise of 45%, from 2.15 to 3.12) and the worst from the PSs (no change).

The improvements for LOC are more consistent because the starting values were more modest. Progress on a national scale was 49%, but performances varied in a range from the highest of the UPs (85%, from 1.92 to 3.56) to the lowest of the SIRCs (23%, from 2.71 to 3.33). PSs did have an increase of 33% but recorded the worst overall performance with a LOC of 2.16.

Figure 11 positions Italian HCOs according to the quality of their Internet presence. Overall values do align with results from the other research areas, showing the following:

- The UPs of the north and the centre are very close to the positive benchmark.
- The LHUs, SIRCs, and IHs of the north and the centre take the lead for national averages regarding quality of Internet presence.
- The chasers comprise the IHs and the SIRCs of the south, with a quality of Internet presence that, even if well below that of the leaders, is still in line with average national values.
- A final group of latecomers (mainly the UPs of the south and PSs), regarding this dimension of analysis, is in a generally weak position.

Figure 11. The quality of Internet presence of Italian HCOs



The Effectiveness of Internet Presence

Italian HCOs' Web sites have quite a good visibility on the Net: They can be reached by the most common search engines around 8 times out of 10. Their Internet effectiveness, however, is still limited in terms of overall Web site traffic.

Regarding the Web site visibility index, good results were already seen in 2002. In the national average, improvement was 27%, with a visibility index for search engines rising from 60% to 76%. Regarding the other dimensions of analysis, the best performance was from the UPs (with an increase of 72% and visibility with all the search engines used). The rise for the SIRCs was only 2%, although the visibility index was one of the highest (82%). PSs, despite a positive increase of 19%, are in line with the average values of the first year (visibility index of 63%).

Regarding the Web site traffic ranking, the second year of our research produced results that could be described as somewhat curious in that they are not in line with the results of previous analysis. There being generally consistent improvements in the complexity, typology, and quality of Internet presence of Italian HCOs, we would also have expected an improvement in the Alexa.com ranking. In other words, we thought that the real endeavors of the Italian HCOs to develop their presence on the Internet would have been translated into a greater success in terms of visitor numbers and traffic on their Web sites.

In fact, there is not a great difference between the results of the first and the second years. There has been, even if only slight, a reduction in the average national ranking (-3%) and increases only for the SIRCs and UPs (1% and 11% respectively).

Closer analysis leads to the following conclusions:

- Alexa.com evaluates traffic to the site, providing a global classification of Web sites. Performances are inevitably influenced by the rankings of other sites, be they healthcare related or not, or Italian or foreign.
- The world of the Internet is continually changing and evolving. Italian HCOs have undoubtedly progressed compared to the situation in 2002, but comparison with the international scene shows that the others have not been standing still and that Italian organizations, in several areas, are still behind the USA, Canada, and the United Kingdom (not to mention the abyss that distances Italy from international benchmarks).

Figure 12 positions Italian HCOs according to the effectiveness of their presence on the Internet. It shows the following:

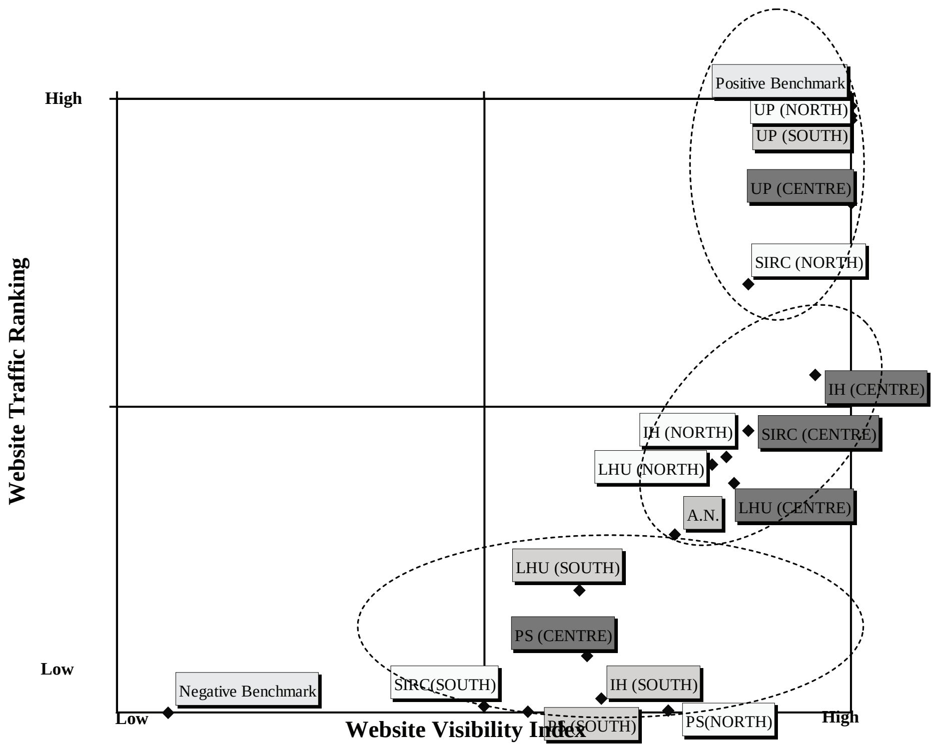
- A first group of organizations, comprising UPs and the SIRCs of the north, in a leadership position with results in line with or very close to national benchmark references.
- A second group of companies, representing IHs and LHUs in the north and centre and the SIRCs of the centre in a chasing position, with results for effectiveness in line with national average values.
- A third group of organizations comprising the remaining Italian HCOs, which, at present, are in a relatively weak position as regards the effectiveness of their Internet presence.

Limitations and Further Research Areas

While aiming to provide a complete summary of the presence of Italian HCOs on the Web, our study nevertheless inevitably suffers from some structural limitations, which can be summarized as follows:

- The research project was carried out to evaluate, through external analysis, how

Figure 12. The effectiveness of Internet presence of Italian HCOs



and to what degree healthcare organizations currently exploit the potential of the Internet through their Web sites. The aim was to understand the strategy of colonization of Web space through analysis of the informative content or services offered. In this phase, we did not consider the internal aspects of the organization in the definition of the Web strategy (definition of the strategy, evaluation of costs, evaluation of performance), which does constitute an essential part of the research we are undertaking.

- The Internet is constantly evolving. Through our research, we attempted to take a picture of the state of the art of healthcare online, nationally and internationally, at two different times, first in 2002, then in 2005. This enabled us to identify the changes that took place over this period, but it was not our aim to define trends, find a way forward, or predict future developments.
- The aim of the research project is that of providing as objective an evaluation as possible of Internet presence. Although accurate scales have been elaborated for the evaluation of LOS and LOC, there remain in this phase some inevitable and unavoidable sources of subjectivity.
- The evaluation of traffic recorded on the Web site was carried out using a site classifier Alexa.com. When sites are hosted by a main domain site (the sites of some LHUs on the regional site, or some PUs on the relevant university site), the figure refers to the main domain.
- An international comparison has been developed through the analysis of HCOs in the capital or similar city of each nation, similar in terms of the structure of the health service. As such, it inevitably provides only a general outline of the state of the art of online healthcare in each nation.

- The results obtained from the international sample, as for those from the survey regarding Italian structures, can be compared and confirm hypotheses only indirectly. This is because at the moment, and at least at a European level, we did not find similar studies about the actual level of exploitation of the Internet by healthcare providers. In this sense, it can be said that the lag shown by Italian structures compared with international benchmarks is basically in line with the limited attitude revealed by Italian companies to invest in technology (in 2003, only 2% of GDP [gross domestic product]) compared with the European average (3% of GDP) and the United States (3.8% of GDP).

Furthermore, the research activity carried out so far represents the first phase of a wider project that is planned for the carrying out of an in-depth analysis of a sample of organizations that will permit the analysis of the path followed in the definition of a Web strategy; its relative impact on organizational, clinical, and management processes; and the results in terms of cost reduction, quality improvement, customer satisfaction, and so forth.

CONCLUSION

"Who doesn't know the Internet?!"

The research undertaken leads to this conclusion: Nowadays, everyone, even in healthcare, knows the Internet; few, however, use it to its full potential.

In a few short years, the Internet has become one of the most widely used technologies in everyday life. In healthcare, however, despite the progress that has been made between the first and second years of our research project, it seems that the full potential of this technology is often not realized.

About the international survey, we can assert the following:

- There is a considerable distance between the benchmark and the other organizations that make up the international sample. HCOs in the United States and Canada perform best, but the European HCOs have reduced the gap in 2005.
- Only the organizations that make up the benchmark have a widespread and homogenous presence on the Internet, fully exploiting its potential. The HCOs that make up the international sample only partially exploit the potential of the Internet.
- If the benchmark organizations are excluded, once more it is the U.S. and Canadian organizations that perform best for the quality of Internet presence.
- Generally, the organizations with the best performances for traffic recorded are those that best exploit the potential the Internet offers and consequently their results are higher in all areas of analysis.

In Italy, many organizations are still unable to exploit this potential, not having activated a Web site. This is particularly true for PSs and those in the south. Public organizations, the SIRCs, and UPs have gradually become more aware of the Internet. The reasons for this change can be found in national health policies, in the increasing simplification of bureaucratic procedures and consequent reduction in costs, and, finally, in the growing awareness of the Internet's importance for the common good (improvement of relations with patients for healthcare organizations, and the facilitation of research and didactics for the SIRCs and UPs).

This has all led to a continual increase in the quantity of information and services offered by Web sites and greater focus on the tools (site maps, search engines, foreign-language versions) that facilitate access to their content. The gap in respect to international reference benchmarks and overseas organizations remains, however, substantially unchanged; this, moreover, is the same for all dimensions of analysis.

The approach to the Internet is rarely strategic, but often improvised, utilizing it simply as a means of communication. There rarely seems to be an awareness that investment in this technology should be in line with the institutional strategy of the organization (Minard, 2001), leading to a rethinking of the main ways of operating and assisting in healthcare.

This inevitably means that the possibilities that the Internet offers are not exploited as fully as they could be.

The quality of the contents available, in terms of technological sophistication and availability of access are, however, of a good level.

Like in the international survey, it is undeniably true that the best performance in terms of traffic recorded on the sites was achieved by the organizations that have invested more effort in this technology through the development of a high-profile Web strategy.

REFERENCES

Alemi, F. (2000). Management matters: Technology succeeds when management innovates. *Frontiers of Health Service Management, 17*(1), 17-30.

Anessi Pessina, E., & Cantù, E. (2003). *L'aziendalizzazione della sanità in Italia: Rapporto Oasi 2003*. Milan: Egea.

Angehrn, A. (1997). Designing mature Internet business strategies: The ICDT model. *European Management Journal, 15*(4), 361-369.

Baraldi, S., & Monolo, G. (2004, July). *Performance measurement in Italian hospitals: The role of the balanced scorecard* [Working paper]. Proceedings of the 2004 Performance Measurement Association Conference, Edinburgh, Scotland.

Coile, R. C. (2002). *The paperless hospital: Healthcare in a digital age*. Chicago: Health Administration Press.

De Luca, J. M., & Enmark, R. (2000). E-health: The changing model of healthcare. *Frontiers of Health Service Management, 17*(1), 3-15.

Fattah, H. (2000). Failing health. *Media Week, 10*(17), 100-103.

Flory, J. (1999). Healthcare communications approaches for an online world. *Marketing Health Services* (pp. 25-30).

Frosini, B. V., Montinaro, M., & Nicolini, G. (1999). *Il campionamento da popolazioni finite.* Torino, Italy: UTET.

Given, R., et al. (2000). *Winning the loyalty of the e-health consumer.* Deloitte Research. Retrieved from www.dc.com/research

Given, R., et al. (2001b). *Understanding the NHS market for eHealth*. Deloitte Research. Retrieved from www.dc.com/research

Given, R., et al. (2001a). *Strategy and e-health*. Deloitte Research. Retrieved from www.dc.com/research

Glaser, J. P. (2000). Management response to the e-health revolution. *Frontiers of Health Service Management, 17*(1), 45-50.

Glaser, J. P. (2002). *The strategic application of information technology in healthcare organizations, (The Jossey Bass Health Series).* San Fancisco: John Wiley & Sons.

Goldsmith, J. (2000). How will the Internet change our health system? *Health Affairs, 19*(1), 148-157.

Goldstein, D. E. (2000). *E-healthcare: Harness the power of the Internet e-commerce & e-care*. MD: Aspen Publication.

Malcolm, C. (2001). Making a healthcare Web site a sound investment. *Healthcare Financial Management* (pp. 74-79).

Malec, B., & Friday, A. (2001). *The Internet and the physician productivity: UK & USA perspectives.* Working paper, EHMA Congress. Granada: EHMA Congress.

Minard, B. (2001). CIO longevity: IT project selection and initiation in the healthcare industry. *IT Healthcare Strategist*, *3*(12), 1-12.

Minard, B. (2002). CIO longevity and IT project leadership. *IT Healthcare Strategist*, *4*(1), 3-7.

Nicholson, N. (1999). *The Internet and healthcare*. Chicago: Health Administration Press.

Porter, M. E. (2001). Strategy and the Internet. *Harvard Business Review, 79*(3), 63-78.

Rodgers, J., et al. (2000). *HealthCast 2010: Smaller world, bigger expectations.* PricewaterhouseCoopers. Retrieved from http://www.pwcglobal.com/healthcare

Solovy, A., & Serb, C. (1999). Healthcare's 100 most wired. *Hospitals and Health Network, 73*(2), 43-51.

This work was previously published in Web Mobile-Based Applications for Healthcare Management, edited by L. Al-Hakim, pp. 62-99, copyright 2007 by IRM Press (an imprint of IGI Global).

Chapter 8.3
Web Portal for Genomic and Epidemiologic Medical Data

Mónica Miguélez Rico
University of Coruña, Spain

Julián Dorado de la Calle
University of Coruña, Spain

Nieves Pedreira Souto
University of Coruña, Spain

Alejandro Pazos Sierra
University of Coruña, Spain

Fernando Martín Sánchez
University of Coruña, Spain

INTRODUCTION

Medical data and digital imaging for medical diagnosis currently represent a very important research area in computer science. The generation of medical information is continuously increasing. More specifically, genomic (molecular and histological) data and images have become key points for diagnosis. The specific processing these data require is more and more requested.

This article describes a Web portal based on the most common current standards. This platform is not only able to integrate the medical information available at several sources, but also to provide tools for the analysis of the integrated data, to use them for the study of any pathology. It will provide a common access point to share data and analysis techniques (or applications) between different groups that are currently working in several fields of health area.

BACKGROUND

Nowadays, several studies are being carried on with regard to the different levels of information about health (population, disease, patient, organ, tissue, molecule, and gene) but none of them integrates the information. The biomedical computer science must play an important role at the integration of these viewpoints and their data.

From a classical viewpoint, computer science in public health has been able to confront and solve problems at different population levels; has effectively managed levels of diseases and patients and lastly; has developed tools for image management and analysis to be used in non-invasive techniques for tissue or organ study. The source of knowledge regarding molecular and genetic levels is greater every day. One of the fields were developing new applications is Genomic Epidemiology, which performs population studies about the impact of genetic human variability on health and disease. Another field, Pharmagenetic, considers the differential genetic aspects among people (e.g., SNPs profiles) when developing new medicines and analyzing its influence after the administration of a medicine.

HUGE NET (from Office of Genetics and Disease Prevention, USA) is an example of this kind of application. Briefly, it is a communication network that allows sharing epidemiological information about Human Genome.

PharmGKB program (from Stanford University) is used in nine universities and medical centers, which investigate pharmacogenetics. The program makes a knowledge base possible with genomic data, laboratory fenotypes, clinic informations, etc.

However, these examples solve just partial aspects of the aim, but not the complete problem. Nowadays, there are not examples of integrated information systems to cover this kind of study completely. The development of such a system will facilitate the studies about complex diseases.

Digital imaging for medical diagnosis is currently one of the most relevant research areas. Since the discovery of the x-ray in 1895, the techniques for acquisition of medical images have evolved to images in digital format.

Every manufacturer used to design its own image storage format, therefore the development of applications should be specific for every device. Therefore, it makes it impossible to transfer information between different machines. A standard named DICOM (Digital Imaging and Communications in Medicine) was published (Bidgood & Horii, 1996; Clunie, 2005; Nema, 2005) as a solution for these problems. DICOM unifies imaging storage criteria for their transmission among heterogeneous equipment by a common procedure, which is open and public.

Another problem related to medical imaging is its accurate management, mainly due to the great volume to store. This way, the picture archiving and communication system (PACS) (Huang, 2004) makes the achievement of an imaging service that might integrate images and clinical information without films or paper documents possible.

The PACS DICOM duet, combined with Web technology provides the specialist with the possibility of gaining access to images and their related information from place, using the legally required security mechanisms (BOE, 15/1999, BOE, 994/1999, Garfinkel & Spafford, 2001).

The existing health databases and Web portals are heterogeneous and physically dispersed. These DBs may be relational, as PACS DICOM, public, as NCBI (NCBI, 2005), or HapMap (HapMap, 2005), etc. Therefore, there also exists a great variety of software for data processing. There are some development platforms for Windows and Linux in different programming languages as Java o C, several commercial tools for image management like Quantity One from Bio-Rad Laboratories (Bio-Rad, 2005), LabImage (LabImage, 2005), Phoretix 1D developed by Nonlinear Dynamics (Nonlinear, 2005), or Label Cell Counter Software create in the Image Management Laboratory, Otolaringology Department, Rochester University Medical Center.

These are potent tools, which cover the requirements of this kind of image, although not always in an automatic way. Besides, they are a commercial software, so it is not possible to add new functionalities and, in most cases, they can deal just with a specific type of images.

The previous circumstances disturb not only the access to information, but also the processing of data and image. For instance, to perform the study of any disease, the first step will consist on locating the different DBs containing the desirable information. Secondly, it is necessary to generate the appropriate queries to the DBs in a specific language and with a specific structure. Finally, obtained data must be adapted to every program wanted for the analysis. The process is, consequently, a tough task.

It would be desirable to have systems able to store, relate, manage, and visualize all the data and the information coming from several studies, and process it as a homogeneous dataset instead of multiple and separate sources.

These systems should be developed ready for their integration in a Web environment; which would facilitate independence of place and time and user personalization. In addition, the easy use of this environment decreases the learning time.

PROPOSED SOLUTION: WEB PORTAL

The proposed solution lies in a Web portal for managing and accessing heterogeneous information stored at several repositories. It also provides the different users with a tool repository for data processing.

The system can work in two different ways. First, it is able to generate specific applications for a given pathology. These applications are Web interfaces for retrieving information from several data repositories. Thus, the user can visualize relevant information and images by means of processing algorithms adjusted to his or her needs.

Besides, the system has a services layer where advanced users can include data he or she wants to analyze. This layer also provides the user with Web services (Colin, 2005, Sun, 2005) for accessing information from any application.

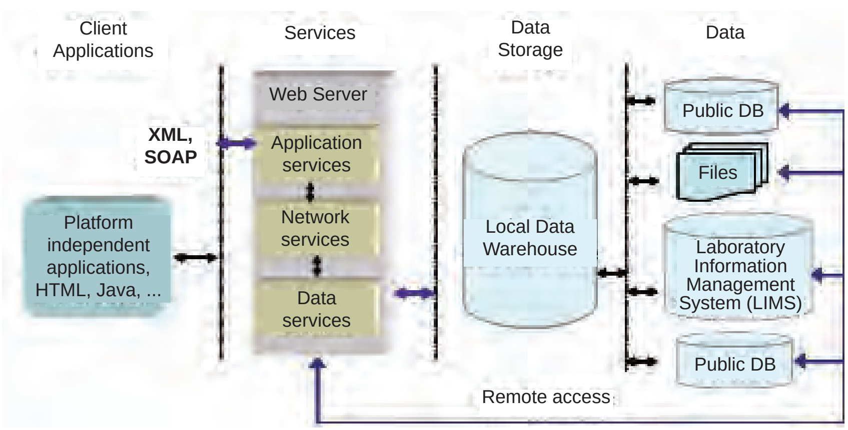
The developed system, which fulfills the previous requirements, is a four level platform: user applications, services, data storage, and data source levels are represented in Figure 1.

Data Management

The data sources level consists of data repositories (public DBs, files, LIMS, etc.) accessible for the portal users. It is possible to add new repositories by "federating" them to the platform.

The data storage level establishes the support for data warehouse and data mining techniques.

Figure 1. Platform architecture



The services level makes it possible for applications to process data from several and heterogeneous sources as if they were homogeneous and unique. It has three sub-layers: application, network, and data services. Application services work as interfaces to communicate the applications with the system. Network services facilitate the location of the data. These services offer a repository of the data model, which contains the definitions of all the objects of the system and also their related attributes and relationships. Network services also provide a security repository, a catalog about the location and function of the existing network systems, and the identification of the service of each group that can obtain data related to an object. Finally, data services carry on the maintenance, consultation, and access control to data.

The client applications level is an XML Web service that facilitate the access to the system and the provided services.

QL is a simplified SQL language developed for this platform. It can perform management operations such as to create, erase, or modify platform elements, as well as to grant or deny privileges regarding these elements. It can also launch consults about a previously defined data source or call on the execution of algorithms offered by a "Web service."

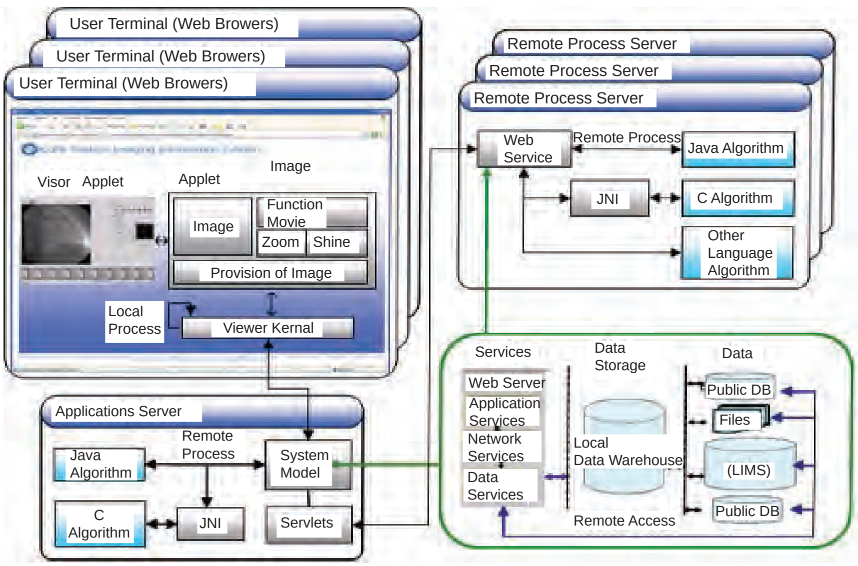
Data Processing

The computational cost of the algorithms used for data processing (data mining, image processing, etc.) varies largely depending on their own complexity and the type of data used. This is the reason why the proposed solution is based on a distributed architecture that enables two ways of processing: local and remote.

Local processing has simple computational requirements. They are basic algorithms that do not require powerful systems for their execution, but the mere ability of the user's computer.

Remote processing needs a more complex computational requirements. The procedures may last for a couple of hours, so it is necessary to use more powerful systems. Examples of this type of algorithms are segmentation and 3D reconstruction.

Figure 2. Structure of remote processing servers



As there are a number of libraries containing different algorithms for data analysis and processing, our architecture enables the reuse of algorithms in different programming languages. This architecture has three well-differentiated parts (see Figure 2): user terminal (viewer), applications server, and process server. These parts communicate among them by using the described platform as central node.

The user terminal is loaded on the user Web browser for visualization, processing, and data analysis. The user selects the data to process and the system will display several environments (known as tools), which are groups of functions to handle a given type of data. These functions (named as components) may vary from information management to digital image processing.

The applications server performs three main tasks. First, it communicates with the tools for both information transfer and obtaining to assure the independence of data access. The second task is data processing by using complex computational requirements (remote processing). Finally, the last task is to give access to the different servers of remote process.

The remote process server is last part of the architecture and its goal is not only to achieve a greater ability for processing and its distribution, but also to enable the integration of algorithms developed in different programming languages.

If processing algorithms have simple computational requirements, their execution would be performed at the user terminal (by using Java Algorithms) as Figure 3(a) shows.

In contrast, the algorithms that need complex computational requirements or the ones in other languages different from Java will run at the applications server or at several process servers. This remote processing bases in Web services developed in *.Net* (MacDonald, 2003; McLean, Naftel, & Willians, 2003; Ramer & Szpuszta, 2005) and *Java*. In this case, the component consists of two parts: local and remote. The local part of the component (i.e., user browser) is where to establish the parameters of the algorithm. See Figure 3(b).

The remote process servers can run together to perform load distribution. In that case, the remote part of the component requests the Web service, which would execute the algorithm and send the result back to the remote server for it to send this result to the local viewer. Figure 3(c) shows this process.

The remote process server proposed here consists of three parts (see Figure 4). The *Web services proxy* collects the external processing requests coming from the information system. These requests are sent to another server where the *processing manager* is located. This part redirects the requests attending both to the type of processing they need and to the original object that sent the request to the Web server. The last part comprises the *remote processing servers*, every one of which can execute one or more different processing algorithms

The ideal processing scenario would involve three or more machines for the processing server.

Figure 3. Processing types

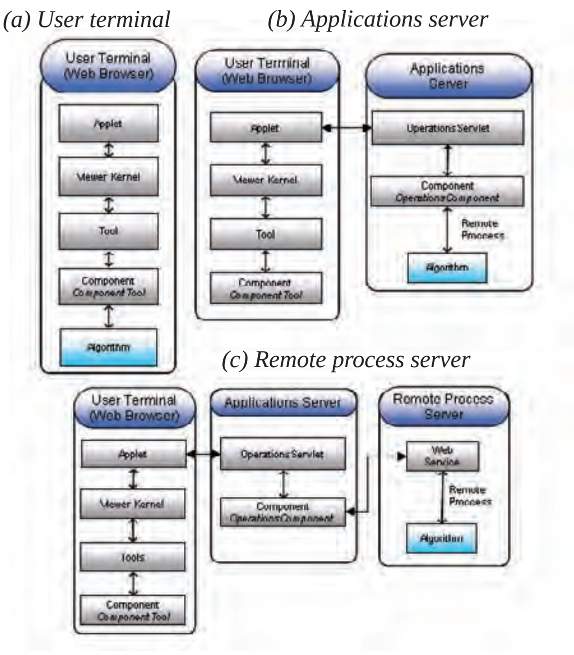


Figure 4. Structure of remote processing servers

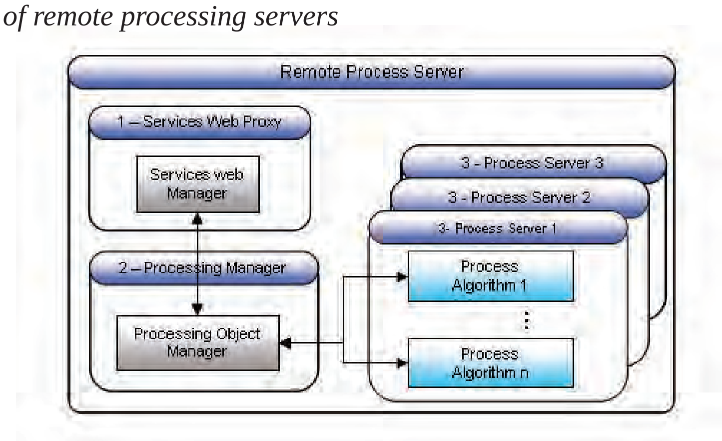
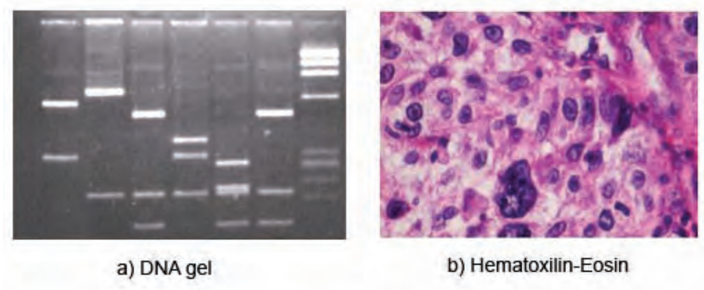


Figure 5. Samples of images



A first machine with a Web server would receive the incoming requests from either the applications server or a stand-alone application by means of Web services. Secondly, a server would keep a processing manager and the rest of the machines would lodge the processing servers.

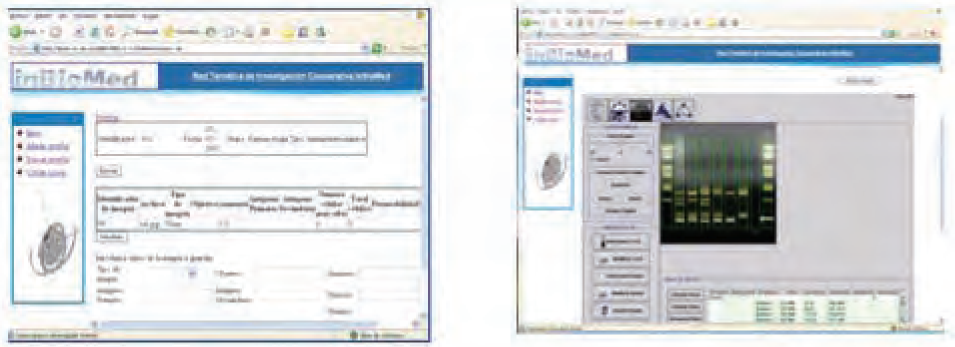
TRIAL BIOMEDICAL APPLICATIONS

In order to validate the concept of platform for the integration of medical data, there have been instituted two data processing systems, each for a different medical specialty.

The Pharmacology group of Santiago de Compostela University (Spain) carries on the research for the development of new drugs, validation of new therapeutic targets, etc. Among the information this group needs for its work, there are images of electrophoresis and DNA gels (see Figure 5(a)). The developed system helps with the storage and visualization of data and images of performed tests. It provides suitable tools for working with this type of information (i.e., image fitness, automatic count, automatic detection of bands and tracks, etc).

The group of epidemiological, environmental, clinical, genetic, and molecular research for urinary bladder cancer disease of Barcelona (Spain) carries on the study of those environmental, genetic, and molecular factors associated with the etiology and prognosis of urinary bladder cancer disease. The developed system uses information and images of Hematoxilin-Eosin (HE) stained slides (see Figure 5(b)). It provides a series of suitable tools for storing, managing, and visualizing data and images extracted from the tests. Tools for automatical cell count or image fitness are some examples.

Figure 6. Information System of the Pharmacology group



The DBs for both groups were federated at the platform, which is physically located at Jaume I University of Castellón (Spain). DBs are located at La Coruña and Barcelona respectively.

Some of the most important tools for information analysis are those relating to image processing. This field has already gone over a long way and currently there are specific bookcases for working with images, numerous algorithms for their processing, etc. As an example, the algorithms implemented with advanced API for JAI (Java Advanced Imaging) (Rodrigues, 2001, Sun, 2005) image processing can be used when performing local processing of images. Some of the processing algorithms federated at the platform are 2D segmentation algorithms, which use region growing algorithms for binary image conversion, algorithms for region extraction, etc.

The system users have a Web environment inside the Web portal, which offer them relevant information and provides them with suitable tools for processing as well as for analysis, attending to the data to visualize. Figure 6 shows two pages of this system. The first one has the data required by the user, the second one shows a tool for the analysis and processing of data coming from DNA gels images.

The information system at the Pharmacology group can be accessed at www.inbiomed.udc.es.

This platform for data management and processing is currently being developed through a project financed by the Health Research Fund (FIS) of the Carlos III Health Institute (Spain). These funds were obtained after the constitution of a thematic network for cooperative research in biomedical computer science, which has been named as Inbiomed. Thirteen research groups of different Spanish autonomous communities, such as Andalucía, Cataluña, Euskadi, Galicia, Madrid, and Valencia, with more than 100 researchers taking part in the network. More information can be obtained at http://www.inbiomed.retics.net/

CONCLUSION AND FUTURE TRENDS

The developed portal intends to relieve the increasing demand for an integrated information system to deal with all the information levels (population, disease, patient, organ, tissue, molecule, and gene) of health studies by integrating all the associated information. Therefore, biomedical computer science fulfills a preponderant role, which also should play with new emerging data and viewpoints.

The federation of databases, files, etc., provides an access to public and private databases by means of a “data warehouse” with consolidated information, which is oriented to requests. Besides, there is a common access to all the available data sources due to the access to the portal by means of a client tool or through Web services. Moreover, as the portal facilitates the federation of processing and analysis tools, there are a group of algorithms that can be either applied to data through the system interface or used as modules of previously developed applications.

Finally, the Web architecture proposed here provides the platform with hardware and software independence due to the integration of algorithms in different programming languages. At the same time, the incorporation of tools and new types of data supports the growth of the system. Another benefit is time and location independence provided by Web applications.

The group will continue the validation of the platform through several systems for data processing, oriented to other medical specialties. This will imply the development of new tools and algorithms for analysis and management of different DBs.

REFERENCES

Bidgood, W. D., Jr., & Horii, S. C. (1996). Modular extension of the ACR-NEMA DICOM standard to support new diagnostic imaging modalities and services. *Journal of Digital Imaging, 9*(2), 67-77.

Bio-Rad Laboratories. (2007). Retrieved December 22, 2005, from www.bio-rad.com

BOE (Boletín Oficial del Estado, número 151, de 25 de junio de 1999). *Real Decreto 994/1999, de 11 de junio, por el que se aprueba el Reglamento de medidas de seguridad de los ficheros automatizados que contengan datos de carácter personal.* (Spanish Order for security about prívate data files).

BOE (Boletín Oficial del Estado, número 298, de 14 de diciembre de 1999). *Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal.* (Spanish Order for security about private data).

Clunie, David. *Medical image format site*. Retrieved December 21, 2005, from http://www.dclunie.com

Colin, Adan. *Web services (WS).* Retrieved November 30, 2005, from http://www.webservices.org/

Garfinkel, S., & Spafford, G. (2001). *Web security, privacy, and commerce* (2nd ed.). Seastopol, CA: O'Reilly.

HapMap Project. (n.d.). Retrieved December 14, 2005, from http://www.hapmap.org/index.html.en

Huang, H. K. (2004). *PACS and imaging informatics: Basic principles and applications.* Hoboken, NJ: John Wiley & Sons.

LabImage. (n.d.). Retrieved December 21, 2005, from www.labimaging.com

MacDonald, M. (2003). *Distributed applications: Integrating XML Web services and .NET Remoting.* Redmond, WA: Microsoft Press.

McLean, S., Naftel, J., & Willians, K. (2003). *Microsoft .NET remoting.* Redmond, WA: Microsoft Press.

NCBI (National Center for Biotechnology Information). (2007). Retrieved December 14, 2005, from http://www.ncbi.nih.gov/

Nema. *DICOM Home Page*. (n.d.). Retrieved December 22, 2005, from http://medical.nema.org/dicom/

Nonlinear. (n.d.). Retrieved November 30, 2005, from www.nonlinear.com

Ramer, I., & Szpuszta, M. (2005). *Advanced .NET remoting.* Berkeley, CA: Apress.

Rodrigues, L. (2001). *Building imaging applications with Java technology.* Addison Wesley.

Sun Developer Network (SDN). *Java technology and Web services.* Retrieved November 30, 2005, from http://java.sun.com/

KEY TERMS

Applet / Servlet: Software components that run embedded in another program, for instance a Web browser (applet) or a Web server (servlet) to extend its functionality. Unlike a program, the applet is not able to run independently.

Component: Each concrete function developed to perform a definite function in the system. It will be possible to find the same component in different tools.

Digital Imaging and Communications in Medicine (DICOM): It is a standard for the storage and transmission of medical images. This standard provides TCP/IP transport among the modalities and the systems of image storage (PACS).

Digital Image Processing (DIP): Group of computer techniques applied to digital images to facilitate their study by the expert.

Local/Remote Processing: Different kind of processing depending on needs of computational requirements. While basic algorithms do not require powerful systems for their execution (allowing local processing), more complex actions require more powerful systems than the usual user machine (remote processing).

Picture Archiving and Communications System (PACS): It captures, stores, distributes and displays static or moving digital images such as electronic x-rays or scans, for more efficient diagnosis and treatment.

Tools: Group of concrete and simple functions (i.e., zoom, brightness, filters) and more specific algorithms (i.e., segmentation procedures, tagging) needed for working with a type of image.

ENDNOTES

[1] Inbiomed research network groups: Fernando Martín —Área de Bioinformática e Informática en Salud Pública (ISCIII) (Madrid, Spain); Ferran Sanz—Grupo de Investigación en Informática Biomédica (IMIM) (Barcelona, Spain); Nurial Malats – Grupo de Investigación en Epidemiología Clínica y Molecular del Cáncer (IMIM) (Barcelona, Spain); Victor Maojo – Grupo de Informática Médica (UPM) (Madrid, Spain); María Isabel Loza – Unidad de Farmacogenómica Aplicada a I+D de Fármacos (USC) (Santiago de Compostela, Spain); Dolores Corella – Unidad de investigación en Epidemiología Genómica y Molecular (UV) (Valencia, Spain); Oscar Coltell – Grupo de Integración y Re-Ingeniería de Sistemas (UJI) (Castellón, Spain); José Antonio Heredia – Ingeniería de Sistemas Industriales y Diseño (UJI) (Castellón, Spain); Montserrat Robles – Grupo BET- Informática Médica (UPV) (Valencia, Spain); Xavier Pastor – Grupo de investigación en Informática Médica (CSC) (Barcelona, Spain); Alejandro Pazos – Laboratorio de Redes de Neuronas Artificiales y Sistemas Adaptativos – Centro de Informática Médica y Diagnóstico Radiológico (UDC) (Coruña, Spain); Mª Gloria Bueno – Ingeniería Eléctrica, Electrónica y Automática (UCLM) (Ciudad Real, Spain); Juan Díaz – Servicio de Informática (HVN) (Granada, Spain).

This work was previously published in the Encyclopedia of Portal Technologies and Applications, edited by A. Tatnall, pp. 1144-1150, copyright 2007 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 8.4
Exploring a UML Profile Approach to Modeling Web Services in Healthcare

Wullianallur Raghupathi
Fordham University, USA

Wei Gao
Fordham University, USA

ABSTRACT

The Web services paradigm offers numerous potential benefits to the health care area. These include interoperability, portability, scalability, and compliance with universal standards. However, development of large-scale Web services applications in health care is relatively scarce. Multiple challenges are slowing the diffusion of information technology practices in health care, starting with the absence of modeling tools and methodologies that are the bases for building such technology. This article describes a UML profile approach to developing Web services applications in the health care area, emphasizing the use of such extensions as stereotyping and tagging to model the unique elements of health care processing. We apply these ideas to a health clinic example. From here, one can imagine building incrementally upon these ideas to develop frameworks for Web services applications in health care. The benefits would enhance the overall quality of health care delivery and lower costs.

INTRODUCTION

Web services are self-contained, Internet-enabled applications capable not only of performing healthcare-related activities independently, but also possessing the ability to engage other Web services to complete higher-order business (health) transactions (Papazoglou & Dubray, 2004). The key is modularity characterized by open standard and internet-oriented interfaces. Examples of such patient services include verifying appointment availability, producing a bill, ordering lab tests, prescribing medication, and discharge planning. The platform-neutral nature of Web services (Kreger, 2001) creates the opportunity for building composite services using existing elementary or complex services, possibly offered by different service providers (Meredith & Bjorg, 2003; Yang, 2003). On a larger scale, these services can be considered a set of interoperable technologies and standards, designed to support the integrating of several autonomous and heterogeneous systems. Web services are particularly geared toward the service-oriented computing paradigm, which can be considered a collection of services, coordinating, and communicating with each other to support a system's specific function or concept. Service-oriented computing has the potential to be the healthcare industry's new foundation for distributed systems and internet-based processing. Web services specifications, such as SOAP, WSDL and UDDI, facilitate open, XML-based methods to support application interoperability, service description, and service discovery (Estrella, McClatchey, Rogulin, Amendolia, & Solomonides, 2004).

The rapid adoption and diffusion of service-oriented architecture (SOA) and Web services in various industries indicate that the advantages and benefits (componentization, interoperability, platform-independence, modularity, reusability, etc.) therein could provide similar benefits for healthcare. Overall, there has been a major change in the way large scale software applications are designed, modeled, implemented and used (Stojanovic, Dahanayake, & Sol, 2004). Services are autonomous, platform-independent processes that can be identified and defined, published, discovered, choreographed, and programmed using universally accepted protocols for building health applications that work together seamlessly, both intra-enterprise (i.e. the healthcare delivery organization) and inter-enterprise (i.e. the consortium of healthcare delivery participants). When services are delivered via the Web we call them Web services. These have emerged as a framework for application-to-application interactions, making these applications available as Web services (Vara, De Castro, & Marcos, 2005). While Web applications via the Internet have been used in the healthcare industry for a few years now, these applications are mostly intra-enterprise, that is, services are within the specific healthcare organization using them. There is limited capability for supporting cross-functional and inter-enterprise applications. Recently, however, practitioners and researchers have started to address the distributed and inter-enterprise nature of healthcare delivery (the interaction arising out of the exchange of information among physicians, insurance companies, hospitals, labs, and pharmacies, for example). The desire to exploit the scalability of Web services within the SOA is understandable. Catley, Petriu, and Frize (2004) describe a prototype Web services architecture for physicians designed to provide clinical decision support. The objective is to integrate and access clinical decision support systems (CDSSs) and medical databases from various medical domains. Their rationale for integrating distributed databases is that different medical domains often exhibit complementary abilities in predicting medical outcomes. They give the example of the collective domains that include obstetrical, prenatal and neonatal databases. The overall goals have been multifold: the Web services are expected to integrate different types of CDSSs; provide for a way to link disparate distributed databases on multiple platforms; and making the CDSSs available locally as well as remotely (Catley et al., 2004).

Estrella et al. (2004) describe another application based on grid computing and the SOA. The MammoGrid project is expected to support collaborative medical image analysis. An important objective is to manage federated mammogram databases distributed across Europe. The framework would include regions with varying protocols, lifestyles and diagnostic procedures and would enable a range of functions including data mining, epidemiological studies, statistical analyses, and deployment of a standardized Standard Mammogram Form. The goal is to federate the distributed mammogram sources into a single virtual organization. While these applications are in very specific medical domains, they highlight the potential of SOA-enabled Web services in the health domain. The health domain is more encompassing than the medical domain because it includes business-related services such as patient billing, insurance approval, payment processing and case management. Even so, the groundwork is in place to scale up applications for more encompassing healthcare delivery systems, delivering services via the Internet. In this way, the industry may exploit the natural alignment of SOA-enabled Web services in healthcare processes. At the enterprise level, one can envision the strategic alignment of information technology with a healthcare organization's goal of delivering quality care.

But while the potential benefits are great, particularly the use of the Internet to deliver important health services, so are the challenges. A key issue is the need for robust modeling and development of Web services applications. Healthcare delivery is a highly complex process characterized by a multi-provider, multi-payer system. Modeling and representing healthcare processes—amidst the need for data integrity, privacy, security, real-time processing, interoperability (open system), portability, and scalability—is daunting. One solution is to extend existing techniques, such as the UML, which many consider the de-facto standard. A starting point is the standard UML modeling notation base defined by the W3C Web services group and OMG (Amir & Zeid, 2004). Aspects of the Web services applications across various platforms—such as the roles of service providers, requestors, and registries as well as the three operations of publish, find, and bind—need to be conceptualized in the SOA format. Further, a notation can be defined to support the actions. Moreover, the semantics of healthcare domain concepts have to be described to enable interpretation of their behavior. Additionally, the ontology of healthcare services must be developed.

These challenges in mind, we explore in this article the potential of a UML profile to model Web services in healthcare. We develop an application for a hypothetical health clinic, showing various elements that would comprise such a profile. This clinic is positioned to be an integral part of a larger, distributed healthcare delivery system that encompasses hospitals, labs, pharmacies, and insurance providers, such as HMOs. The diagrams that follow illustrate our model and visually identify its various parts. In this regard, one can use collaboration diagram notation. We propose a hierarchy of service components as the primary building blocks of the Web services architecture via concepts, meta models, and granularity types. Health service components are modeled as contract-based service providers and support the health organization's health processes through a formal orchestration. Simultaneously, they are realized as a composition of lower-level application service components (Skogan, Gronmo, & Solheim, 2004). These can be further mapped to more encompassing software artifacts. Fashioned in this way, the Web services architecture represents layers of abstraction between the healthcare processes and the operational information, minimizing inconsistencies. We introduce the concept of service component as the primary Web services building block. An important aspect of the article is the enunciation of a modeling approach that emphasizes the identification and specification of the service components of varied scope and granularity. This approach enables the evolution of a Web services healthcare solution.

The article is organized as follows. First, we discuss the potential of Web services in healthcare. Next, we explain the usefulness of the UML, particularly the profile in developing models. Then we develop the notion of a partial UML profile for a distributed health clinic example. Finally, we offer our conclusions for future applications and research

WEB SERVICES IN HEALTHCARE

Web services are defined as functional components (processes) situated in a particular domain and available over the Internet. In the healthcare area, these could include online query/response consultation with a physician, scheduling a patient visit, confirmation of a referral, and so on. The architecture would comply with a set of standards, such as those used with HTTP, XML SOAP, WSDL, or others (Vara et al., 2005). Alternatively, it is any service accessible over the Internet which takes XML as input and produces an XML result. The rationale behind a Web service composition is that many smaller tasks already identified as Web services can work together to execute a larger task. Such commands as VERIFY PATIENT DATA, CONFIRM PHYSICIAN, ROOM ASSIGNMENT, DATE/TIME AVAILABILITY, ASSIGN PATIENT TO SCHEDULE, and PROVIDE CONFIRMATION could collectively represent "PROCESS PATIENT VISIT SCHEDULE DATA." At the highest level a Web service implies delivery of high quality healthcare via the Internet. The services can be offered by multiple providers within the confines of the federation, a loosely coupled consortium of healthcare participants, and are typically published in a central repository. Finding the service is coordinated by a brokering mechanism. The Web services paradigm, then, enables the health services consortium to continue to function in a distributed fashion with localized control and use. In this way, applications extend beyond the intra-enterprise boundary to the inter-enterprise. Additionally, redundancy in services can be minimized by a repository. All participants in this "federation" can contribute to the creation of the services and use them. The benefits are numerous. Web services architecture mirrors the healthcare delivery mode and can be strategically dovetailed with the delivery organization functions. Also, middleware provides the interface so that legacy systems are not abandoned. Web services standards enforce compliance by vendors and platforms, thereby providing interoperability. This exploitation of the Internet would lead to maximal use of the health organization's capabilities and offerings, resulting quality patient care and satisfaction. That said, the application of Web services and SOA in the healthcare area is a relatively recent phenomenon. While knowledge about systems development can be transferred from other industries, the healthcare industry is more complex (fueled in part by the multiplicity of payers and providers) than most other industries. While the current trend towards universal electronic health records is a positive one, scalability and portability are important design challenges. Another hurdle is the minimal understanding of healthcare processes and workflow by application designers and vendors. Existing applications are ad hoc, incompatible system implementations. Plus, current standards are not at a level that Web services can be easily implemented. Methodologies and modeling tools are seriously lacking (Gronmo & Solheim, 2004).

But there is a bright side. UML-based approaches can be extended to the healthcare process modeling, providing a solid foundation for the implementation of Web services. For example, consensus on communication rules is important for all participants to succeed. This implies agreement on the ontology (Heuvel & Maamar, 2003). Anchoring the modeling of healthcare processes around the UML can minimize the learning curve because one can draw on the experience and knowledge currently available. The added advantage of bypassing legacy approaches and systems (especially in environments with no prior automation efforts) accrues.

Operationally, applications that are "services" must spell out their functional and non-functional (utility/procedural) requirements and capabilities in a prior agreement. Based on declarative service descriptions, automated service discovering, selection and binding provide the foundation for Web services applications. Therefore, an advantage arising from the dynamic binding is looser coupling between applications (Curbera, Khalaf, Mukhi, Tai, & Weerawarana, 2003). The componentized model emerging out of this process provides for "health services" forming the initial units for creating the application. Service composition and management become critical issues (Casati, Shan, Dayal, & Shan, 2003).

In healthcare, a service composition combines services and follows a certain pattern to achieve a healthcare goal, for example, properly discharging a patient from the hospital. An architecture that supports Web services includes the following aspects:

- The dynamic discovery of registered service, including searching for services that meet certain criteria, especially health criteria such as patient wait time, quality of care, and so forth;
- The intuitive organization of services so that one easily can understand what a service offers;
- The clear description of services, ensuring the proper use of each service.

This includes formats and protocols for invoking the Web service (Leymann, Roller, & Schmidt, 2002). A typical Web service follows this pattern: a service is created; the service is published; the service is located; the service is invoked. (Optionally, the service may be unpublished.) The complex services that a typical medical clinic provides vary widely across the patient group. An effective solution for the management of disparate medical data sources is federation of autonomous multi-center sites which transcends inter-enterprise boundaries. The resources in the federation—such as hospitals, diagnostic labs, and pharmacies—would be governed by the same sharing rules with respect to authentication, authorization, resource, and data access. These rules (protocols) create a highly controlled environment which dictates what is shared, who is allowed to share, and the conditions under which sharing occurs among the medical sites. Federation in this application implies cooperation of independent medical sites. Individually, these sites are autonomous in that they have separate and independent control of their local data. Collectively, those sites participate in a federation, and the federation is governed by the umbrella organization (let's say, a consortium of health providers) (Estrella et al., 2004). The ability of Web services to go beyond the health entity boundary (i.e. the firewall), the loose coupling between applications encouraged by Web service interfaces, and the wide support for core Web service standards by major enterprise software vendors are the key reasons why Web services have the potential to make integration of health applications both within and between health enterprises significantly more effective in terms of ease of implementation and cost savings. Loose coupling means that not only can applications be implemented on different platforms and operating systems, but also that the implementation can be changed readily without affecting the interfaces (Jiang, Xing, He, & Yang, 2005). One of the important modeling and design issues in building Web services is determining the type, scope, and granularity of service components necessary for representing the main architectural abstractions. Moreover, these different types of service components must be put in the wider context of a development process to provide that traceable business requirements are mapped to the software artifacts supporting them.

Having discussed the potential and implications of Web services in healthcare, we now turn to the examination of UML, its extensions and profile to model heath care processes in a robust fashion. How do we derive good service abstractions from high level healthcare domain requirements and health process models?

As a corollary, additional questions arise. What are good (health) services? For example, what is the right service granularity, and what does health process and information technology alignment mean from a modeling standpoint? How can the healthcare automation landscape be transformed into an integrated healthcare ecosystem? From a modeling standpoint, how can well-designed and meaningful service abstractions be characterized and constructed systematically? How are services in a Web service application identified and described? What is the process for developing Web services? How are health processes realized in terms of Web services? Which development approaches are relevant to a Web service? How can legacy systems and packaged applications be adapted as services? All of these questions emphasize the need to use an appropriate modeling approach.

UML PROFILE APPROACH TO MODELING WEB SERVICES IN HEALTHCARE

The identification and construction of a modeling language for Web services pose significant challenges. To provide rich and robust functionality there is a need to focus on the architecturally significant components particular to the Web services in a specific health system and to somehow model them with UML extensions. One way to manage the complexity of Web services is to abstract and model them. In a typical system there are multiple models, each representing a different perspective, layer of abstraction, and detail. Models help us understand systems by simplifying some of their details. But the decision as to what to model has a significant effect on the understanding of the problem (What are the health processes amenable to Web services modeling?) and solving it (How do you enable them?). UML has the potential to answer these challenges (Jiang et al., 2005). UML models and other architectural artifacts such as system context, component interaction and collaboration diagrams can play a key role during analysis and early design (Gronmo, Skogan, Skolheim, & Oldevik, 2004; Zimmermann, Schlimm, Waller, & Pestel, 2005). Also, use case, sequence, and class diagrams can help represent various modeling elements (Conallen, 2003).

A natural starting point for modeling Web services is component-based and interface-based modeling and the standard UML. But to what extent can component-based approaches and the UML fulfill requirements and provide necessary concepts and mechanisms for modeling the Web services, such as componentization, low coupling, high cohesion, interface-based design, hidden implementation, as well as flow of objects and actions? One can develop high-level models in UML, define conversion rules from UML to a target platform, and then use code generation to derive much of the implementation code for a desired platform.

Two key aspects have to be modeled: the service itself, and the workflow. Service modeling identifies which services are to be exposed with their interfaces and operation, while workflow modeling identifies the control and data flows from one service to the next. In a typical inter-enterprise application, one may first discover existing Web services, if any, from a registry. Then, service modeling identifies the interface of the composite Web service and the workflow focuses on reusability. Transformation rules from both models can be specified (Foster, Uchitel, Magee, & Kramer, 2003). The former is mapped to a WSDL document and the latter to a XML document. Finally, the XML document generates the code via a workflow engine. The WSDL document is published as a Web service registry. The UML, having evolved into a standard modeling language, is appropriate to model the service and the workflow. Given its visual format and wide use, the UML is a good candidate for diagramming comprehensible models. Diagrams are a natural part of UML and are used for workflow modeling of Web service compositions. Building composite Web services lacks sufficient support in traditional workflow modeling. Further, the UML seems suitable for expressing Web service patterns (Gronmo & Solheim, 2004). The diagrams and modeling elements can be extended to develop a high level profile of the healthcare process.

UML Profile

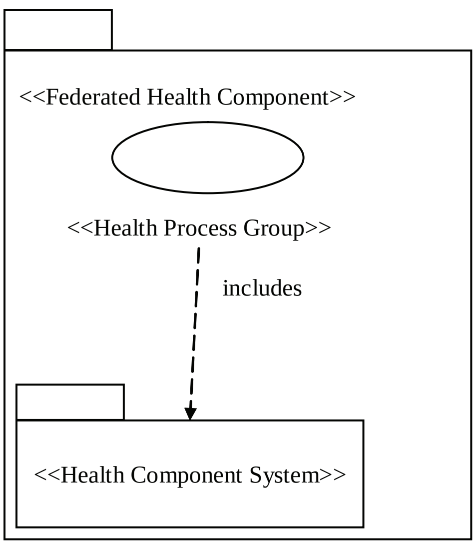
One strength of the UML is its adaptability to specific vertical markets with specific concepts and needs. In the UML standard these are called profiles (Heckel, Lohmann, & Thone, 2003). Various groups working with UML have started to address the customization of UML models for various domains by defining "profiles." The objective for the healthcare process domain is the definition of a generalized UML subset, embellished with UML-compliant extensions, which would permit the annotation of such concepts (Baresi, Garzotto, & Paolini, 2001). Thus, the resulting profile corresponds not to specific domains in healthcare (e.g. diabetes monitoring) but to healthcare processing wherein diabetes monitoring is one service component. While profiles may be standardized in the future, stable initiatives from various groups can form the basis for the definition and standardization of UML profiles (Thone & Varro, 2003). Profiles thus facilitate consideration of the key issues in healthcare processing, namely interoperability, portability, scalability, and universal standards. UML being a complex language, additional features cannot be described easily in the current version, though Version 2.0 may provide extensions. Therefore, UML provides mechanisms such as stereotypes and tagged values to enable extensions (Fontoura, Pree, & Rumpe, 2002; OMG.), and the defined extensions can be grouped into profiles. A UML profile is further defined as an extension of the UML standard language with specific elements. The profile provides new notational elements and typically facilitates specialization of the semantics of some of the elements. Usually the profile builds a hierarchy. The creators of the UML realized that there would always be situations in which, out of the box, the UML would not be sufficient to capture the relevant semantics of a particular domain or architecture. To address this issue, a formal extension mechanism was defined to allow practitioners to extend the semantic of the UML. UML provides two extension mechanisms—stereotypes and tagged values—for describing specific properties of the elements, such as classes, associations and methods (for example, role of "human" as "patient", a patient "visit", and a patient "schedule" in the context of healthcare). The roles and artifacts of the service-oriented style are represented in several diagrams that follow. The mechanism enables one to define stereotypes, tagged values and constraints that can be applied to model elements. A stereotype is an elaboration that allows one to define a new semantic meaning for a modeling element. Stereotypes indicate that the class is an abstraction of the logical behavior of a Web service on either the client or the server. Tagged values are key value pairs that can be associated with a modeling element and allow us to tag any value onto that element. Tagged values are used to define the parameters that are passed along with a link request. The <<link>> association tagged value "parameters" is a list of parameter names (and optional values) that are expected and used by the server page that processes the request. Constraints are rules that define the structure of a model (Fontoura et al., 2002).

One could say each Web service is a UML class. The server side of a Web service can be modeled with one class and the client side with another, distinguishing the two by using UML's extension mechanism to define stereotypes and icons for each <<server page>> and <<client page>>. Using stereotypes makes its easier to model a service's actions and its relationships. Class stereotypes used to model the logical behavior of Web pages means their collaboration with the server side components can be expressed in much the same way as any other server side collaboration (Conallen, 1999). While UML was not originally devised to model health processes or services, one can use the stereotyping capability to develop new extensions to UML. The UML business profile, designed to facilitate business process modeling purposes, is one example.

The Health Clinic Example

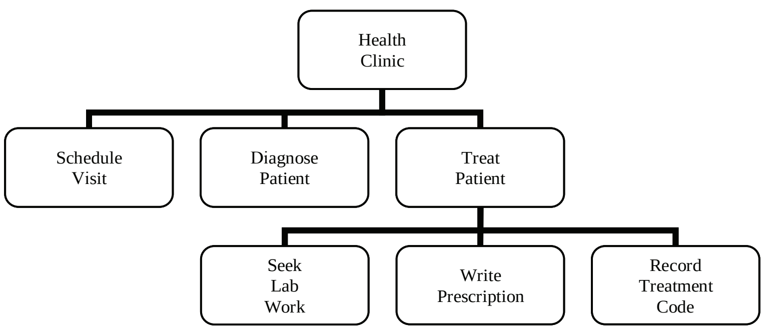
It is possible to provide a UML extension for health service modeling using typical stereotypes. Figure 1 shows the general relationships amongst some of the UML elements. A federated healthcare component performs specialty activities as its services. On the other hand, the specialty itself is decomposed to lower level services, encompassed within a medical entity. It is also similar for other levels of service/process granularity hierarchy. At the architectural level, other types of diagrams are involved, such as class diagrams with interfaces to define signatures and data types of operations, and sequence diagrams as alternative presentations of interactions (Heuval & Maamar, 2003). The Web service domain can be mapped by UML modeling concepts, providing platform independent representation as UML activity and component diagrams. Additionally, a collaboration diagram selected to represent the encapsulation of the binding mechanism as semantically precise as possible and provide all the necessary concepts and capabilities for representing integrations and collaborations of its participating parts. The diagrams, along with structured classes, can support the representation of the composition of interconnected elements representing run time instances, and the collaboration over communication links to achieve common objectives. In particular, collaborations provide the means to define common interactions between objects and other classifiers and assign roles and responsibilities to each. The interaction is similar to providing a pattern of communication between objects playing defined roles. Alternatively, the profile is equated with equivalent meta models. By defining their meta model relationships and interaction mechanisms, one can support the integration and interoperability of their models. In that way one can design complete system models having formalized interaction points and perform model transformations across multiple connected meta models (Staikopoulos & Bordbar, 2005). For example, UML and its standard profiles (CORBA, EJB, and EDOC) can be used to model PIM (Platform Independent Model) or PSM (Platform Specific Model) in the Model Driven Architecture (Brown, 2004). Design decisions can be defined by the UML profile, which provides a set of extensions to UML using the built-in extension facilities of UML, stereotypes and tagged values. The profiles provide a mechanism for managing stereotypes. The profile for healthcare would comprise the range of models derived from UML modeling. The same profile can be used across many projects. Using profiles in this manner (i.e. packages can be exchanged between models) will ensure a standardized approach to problem solving. UML extension mechanisms can help bridge between various models. Using the basics of the class diagram, the systems engineer can model the stack with relative ease. However, for the developer to implement it, he or she will need additional information. UML stereotypes can help clarify

Figure 1. A sample UML representation for health processes and services (adapted from Fatolahi & Shams, 2005)



the stack requirements. In this context, stereotype is a "label" that can be used to "characterize" a model item in some way. For example, an actor could be stereotyped as <<Patient>> <<Clinic>> or <<Service>>. A stereotype is a model element that defines additional values (based on tag definitions) and additional constraints. These can be applied to any model element. Stereotypes label a model element to denote that the element has some particular semantics. UML allows an application model to be constructed, viewed, developed, and manipulated in a standard way at analysis and design time. Just as blueprints represent the design for an office building, UML models represent the design for an application, allowing business functionality and behavior to be represented clearly by business experts at the first stage of development and away from vendor-driven preconceptions. The abstract syntax in UML's meta model is described by UML's class diagram which can represent the abstract syntax structure strictly and precisely (Jiang et al., 2005). For the purpose of health component identification, domain object analysis, and use case analysis are useful sources of input. Domain object analysis defines a domain vocabulary of the system being developed, in other words, information about health concepts in the problem area that should be handled by the system together with their attributes and relationships. Use case analysis is an effective mechanism for defining cohesive sets of features and services on the system boundary because they capture the intended behavior of the system without having to specify how that behavior is implemented. The use cases of the appropriate granularity correspond to the provider's health goals and activities, otherwise known as Elementary Health Processes (EHPs). The cases can be specified in detail according to the use case template that includes name, description, involved actors, goal in the context, scope, level, pre-condition, post-conditions, triggers, main success scenario, extensions, sub-scenarios priority, frequency, performance, etc. Figure 2 depicts a healthcare process hierarchy for this sample case. For simplicity only one thread is decomposed into its EHPs. The stereotyping helps translate the somewhat abstract class diagram into a domain specific platform independent model and platform specific model. For each use case, the use cases that precede it, follow it, perform in parallel with it or are synchronized in other ways with it should be defined. Furthermore, for each use case its super ordinate and subordinate use cases should be defined, providing a composite hierarchy of use cases, (i.e. corresponding business goals). This can be illustrated using a sequence diagram enriched to express the action semantics with the use cases on the horizontal axis of the diagram. Finally, domain information types resulting from domain analysis are cross-referenced with the use cases. In this way, for each use case the information types needed for its performance are defined. A healthcare service may be seen as

Figure 2. Healthcare process hierarchy



a component, which in turn is defined as a self contained health process or service with predetermined functionality that may be exposed through a provider or technology interface. This indicates an implicit relationship between the notions of service, process and component. One can then match the hierarchy of service-components with the hierarchy of healthcare processes in order to gain effective services. Since healthcare processes are significant components of the health delivery architecture and services are the major elements of a Web services' architecture, mapping may be seen as a bridging mechanism between the abstract health service models and the system architectures. A primary healthcare process, called a process group, may contain a series of sequencing process threads. Each thread results in a major added value for a healthcare delivery organization (called a provider), and is composed of its offspring elementary health processes (EHPs), which are the smallest meaningful units of activity for the end user, and which when complete leaves the information area in a self-consistent state. Each EHP may transform to some use cases when designing systems to support its holding thread. Therefore, a process group may decompose to a series of process threads each of which in turn may hold several EHPs. It is necessary to base the services on EHPs. An EHP can indicate which services are required by the enterprise and which of them are provided by the enterprise. Operationally, a health process may be a collection of services. The concepts are illustrated using a "Health Clinic" application that is distributed in residence of the data as well as functionality and use. Table 1 shows the service component hierarchy in the context of healthcare.

A federated health component is a set of services which are common to the various healthcare delivery organizations, including clinics, hospitals, HMOs, labs, and pharmacies. A health component system is a set of health service components assembled to deliver a healthcare solution, such as, providing quality healthcare to a patient. A health component is itself an independent concept, process or service. Distributed

Table 1. Service component hierarchy classifies health services through a hierarchy of granularity/service (adapted from Fatolahi & Shams, 2005).

Health Service Component Granularity	Example
Federated health component	Process group – Consortium of health care delivery participants
Health component system	Process thread – Health clinic
Health component	Use case (EHP) – Process Patient Schedule
Distributed component	Distributed database
Language Class	Method – Verify schedule availability

Table 2. UML stereotyping for health care (adapted from Fatolahi & Shams, 2005)

Service Component Stereotype	UML Element
Federated health component	Package
Health component system	Package
Health component	Package
Distributed component	Component (e.g., database)
Language Class	Class

components and language classes are physical software components (e.g. an underlying database) and classes designed for implementation. It is sufficient to follow the hierarchy of health processes in order to record included services at each level. But this does not mean every process is a service. Identifying the specific services to be implemented using a specific technology, such as a Web service, it may be enough to verify EHPs and decompose each EHP into Derived Logical Processes (DLPs). Each DLP can be a stand-alone process that does not require human intervention. The hierarchy provides a means to package services in well-defined components that are of interest to different users at various levels (Fatolahi & Shams, 2005). Table 2 shows the generic UML stereotyping for healthcare. An EHP included by the schedule is developed here. Figure 3 shows the UML stereotyping for the SCHEDULE VISIT thread for the clinic example. It indicates how the UML stereotyping mechanism may help us in modeling the services in the healthcare processes. The diagram presents the schedule visit thread and some of the linked services. Now consider the medical clinic application that serves as a manager (coordinator) of scheduling and coordinating patient visits. For simplicity, we restrict the functionality to scheduling a patient visit. For this purpose, the clinic software should be able to connect to third party (e.g. referring physician, hospital) information systems. After a patient requests a visit, the system has to query the third party systems as well as internal systems to select appropriate dates/place/location and schedule the patient visit. These requirements lead to a dynamic distributed system where new components can be brought up by the medical facility (e.g. hospitals) at execution-time. The application can be implemented as a Web service as modeled with the UML profile. Models can be differentiated at the various type levels (Baresi, Heckel, Thone, & Varro, 2004). Component diagrams can be used to show the component types of the applications, including the interfaces at the ports. Figure 4 is a partial component diagram of the health clinic example. The patient (or representative) requests to schedule a visit from the clinic. This action links to different delivery organizations. All healthcare delivery organizations need to publish Web services to accept the queries from external

Figure 3. UML stereotyping for the Schedule Visit thread (adapted from Fatolahi & Shams, 2005)

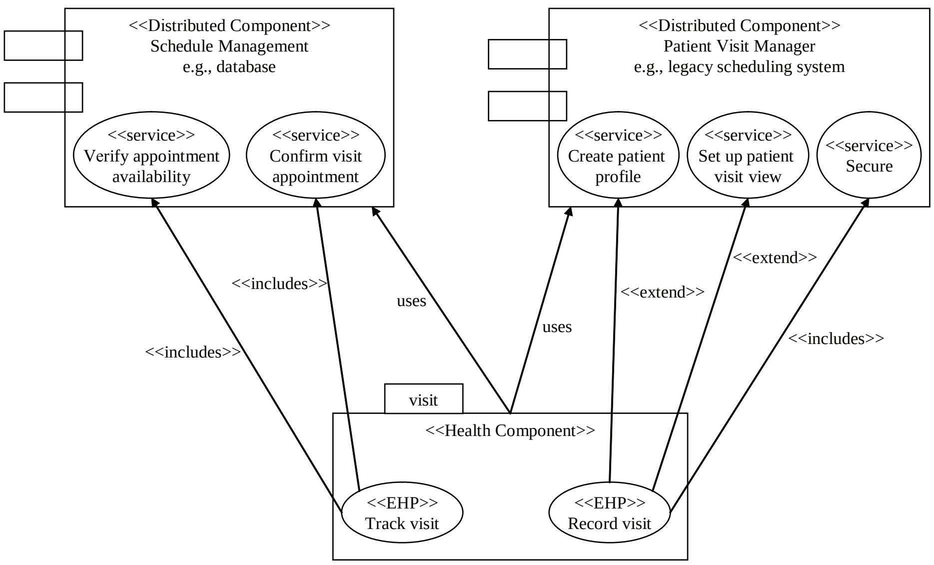
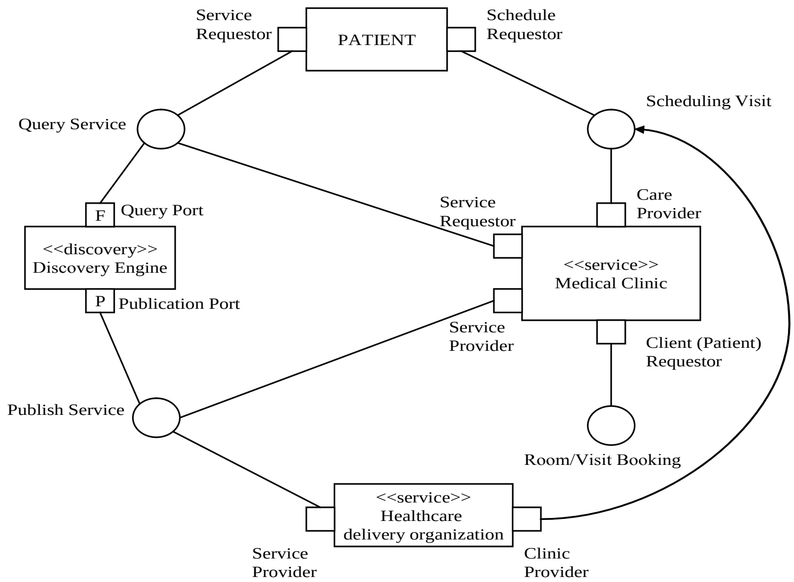


Figure 4. Web services specific component diagram (adapted from Baresi et al., 2004)



clinic systems in a healthcare UDDI Web services registry. When the clinic system queries third party systems, it discovers the Web services in the UDDI Web services registry, then binds the Web services at their particular Web location, and finally obtains the requested information by execution of the Web services. The UDDI helps integrate the healthcare service across various healthcare delivery organizations in an open environment. While WSDL and SOAP describe the service interface and communication protocol once a specific service is located, UDDI enables the service consumer to reach the service provider through a multitude of Web services available. The Service Provider registers the information about the service in the UDDI Registry by publishing the service, and the Service consumer queries the UDDI Registry to discover services matching certain criteria, all through standard defined UDDI APIs (UDDI.org, 2002).

UDDI becomes a key element in Web services runtime environment and life cycle when the services are dynamic and evolving, and the binding for a service is non-deterministic. It is conceivable that service consumer will eventually discover and use the most desirable service through a streamlined find-bind-execute process without human interaction.

Additionally, the details of the port types and the interfaces may be defined by class diagrams. A partial class diagram is shown in Figure 5. The class diagram also contains stereotyped associations. These define types for links that are available to connect to the various ports. We also need to model packages of the component instances in a run-time scenario. UML collaboration diagrams can be used for this purpose. Figure 6 is an example collaboration diagram for the health clinic. Various labels can be applied to the changed instances and links during run-time. The communication features are shown by assigning messages to the link symbols. Numbers are used to indicate the sequence of the messages. The displayed specification artifacts, such as service descriptions and service requirements, are needed to enable the Web service-specific capability of dynamic service discovery (Baresi et al., 2004). An objective of the health clinic system is to provide integrated scheduling and visit services for patients by loose coupling of providers, cus-

Figure 5. Web services specific class diagram (adapted from Baresi et al., 2004)

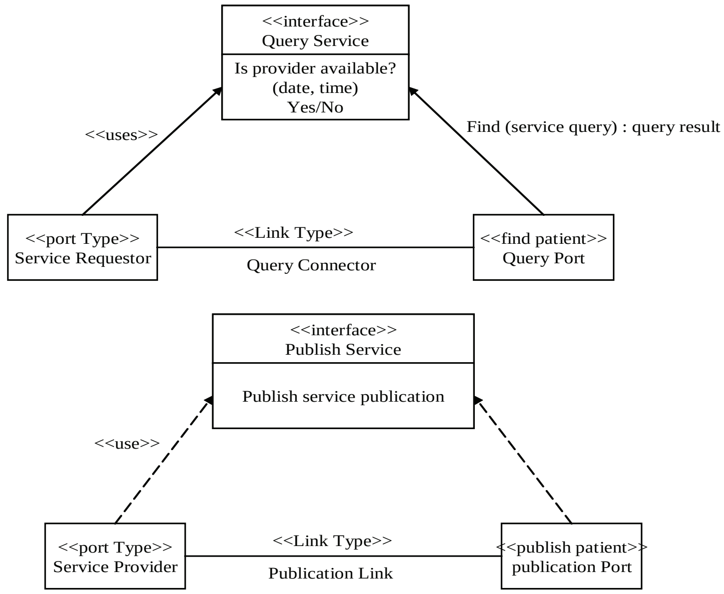
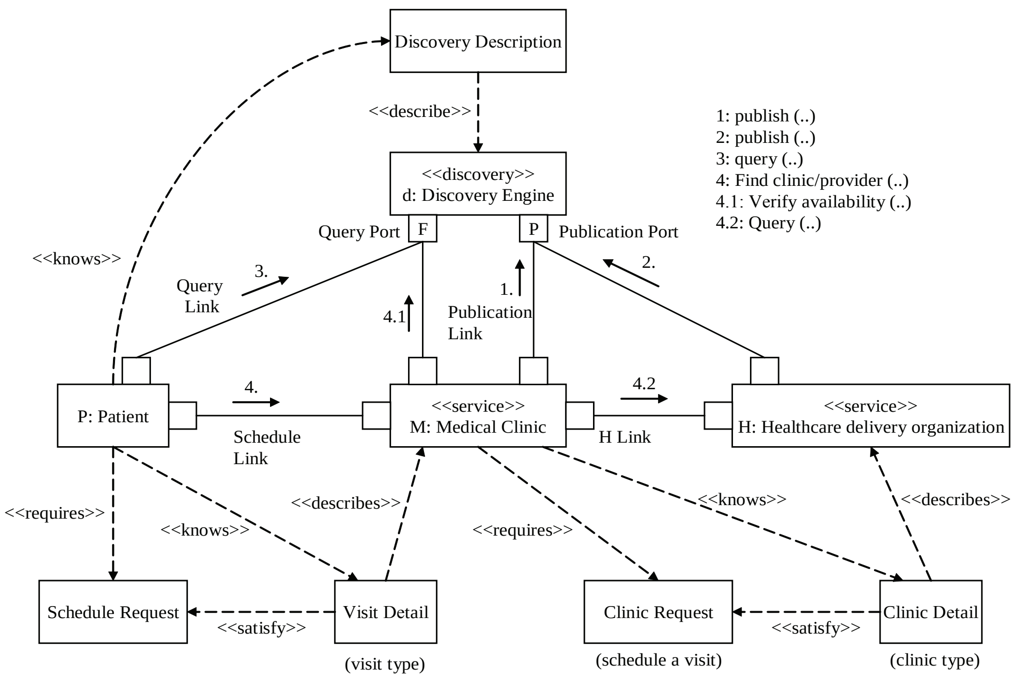


Figure 6. Web services specific collaboration diagram (adapted from Baresi et al., 2004)

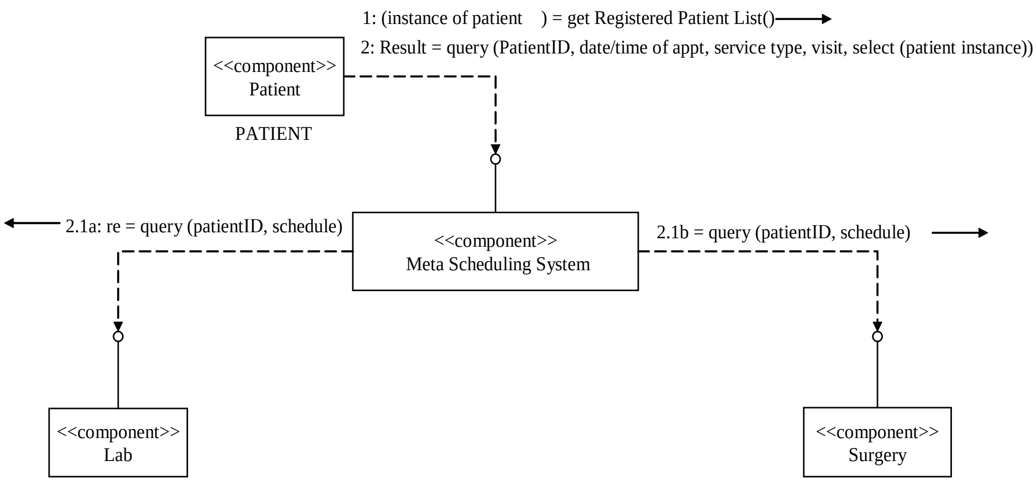


tomers (patients) and related-entities solutions already residing at their sites. Each of these can be accessed via Web service interfaces, and incoming requests are forwarded to all known services (medical; participants). Figure 7 shows a collaboration diagram modeling a hypothetical scenario to execute a query on the meta scheduling system (schedule/visit). Prior to dispatching a query, the client (patient) has to choose the local scheduling systems where the query ought to be run. Then the client retrieves a list of scheduling systems that are registered with the meta scheduling system ("discovery"), for example, the lab, and surgery. The search string is sent simultaneously to all selected scheduling systems. For this situation, a UML component diagram can identify the various components and interfaces for the given system. From the scenario in Figure 7 one can derive the UML component diagram in Figure 8 to identify the (types of) components and interfaces used in the application. The objective to develop UML profiles for healthcare frameworks is a potential definition of UML subsets incorporating extensions and annotating the artifacts. The examples represented by the various diagrams contribute to the overall development of profiles in healthcare.

CONCLUSION AND FUTURE RESEARCH DIRECTIONS

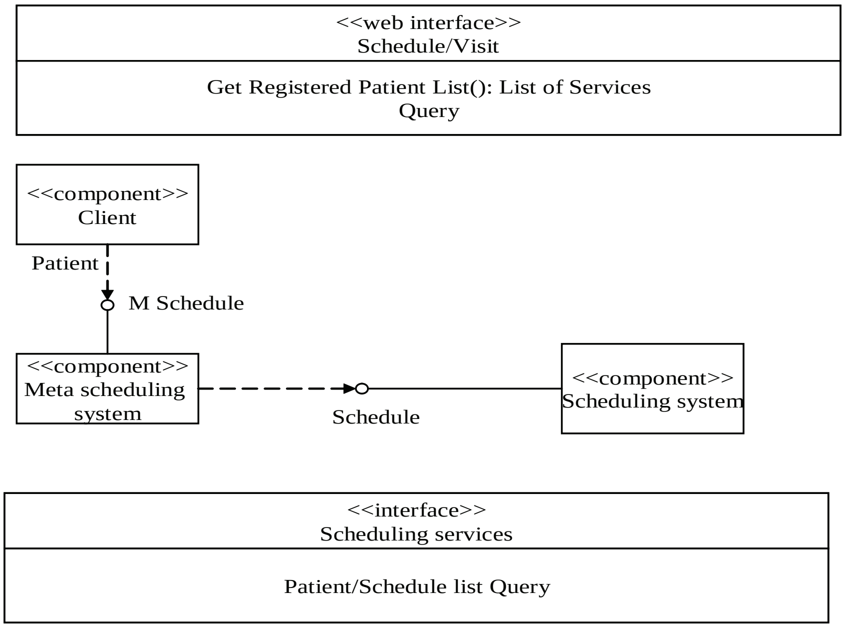
While the Web services paradigm is gaining attention in the development and integration of complex healthcare applications, there is an almost complete lack of modeling approaches. Web services modeling is a challenging task because the service concept further raises the level of abstraction and actually aims at narrowing the gap between the healthcare process and the software. Therefore, for the purpose of Web services modeling, the best practices from object-orientation and component-based design as well as work flow and health process modeling, must be considered and thoroughly integrated. In this article, we propose a UML profile for modeling Web services in healthcare. We describe the various components of such a profile for a hypothetical medical clinic whose interactions are characterized by distributed ness, need for interoperability, portability and scalability. Such profiles developed and standardized over time, will assist the healthcare industry in implementing robust Web services applications for a range of healthcare processes. The paradigm of Web services enables healthcare organizations to shift

Figure 7. Collaboration diagram modeling executing of a query (adapted from Heckel et al., 2004)



from tightly coupled and coarse-grained applications to more dynamic and loosely coupled ones. The UML profile outlined in this article requires further refinement especially if platform-specific details are added. The article has proposed a way to model the services components in a healthcare scenario that enable their realization and the orchestration as the primary building blocks of the Web services architecture. While the research is exploratory one can build on the concepts outlined here to develop comprehensive UML profiles for the various services in healthcare. This would advance the goal of uniform standards for interoperability as health related organizations rapidly develop electronic health record systems, clinical decision support systems and other applications. We advocate the notion of a profile along with the concept of a service component, a paradigm shift from components as static objects to components as dynamic service coordinators. In the future, specific mapping rules for transforming platform-independent models into platform-specific models have to be identified (Stojanovic et al., 2004). We envision health processes as reliable resources to discover services in SOA-enabled Web services (Fatolahi & Shams, 2005) architecture for healthcare delivery organizations (for both, profit and non profit sectors). Also, in the future the UML profile approach developed here can be extended and standardized to facilitate reusability and universal application of Web services in healthcare. Additional research can provide insight into UML extensions and stereotyping in health services as well as enable definition of standards for open source development of applications in healthcare. Finally, integration with HL7 standards has to be addressed. In this way, there is incremental contribution to the advancement of global applications of information technology in healthcare. Loosely federated Web-enabled systems at this level could help track and monitor health issues, such as pandemics, on a global scale.

Figure 8. Partial health services architecture derived from collaboration diagram (*adapted from Heckel et al., 2004)*



REFERENCES

Amir, R., & Zeid, A. (2004). UML profile for service oriented architectures. *Companion to the 19th annual ACM SIGPLAN conference on Object-oriented programming systems, languages, and applications*, pp. 192-193.

Baresi, L., Heckel, R., Thone, S., & Varro, D. (2004). Style-based refinement of dynamic software architectures. In *Proceedings of the Fourth Working IEEE/IFIP Conference on Software Architecture*, (pp. 155-164).

Baresi, L., Garzotto, F., & Paolini, P. (2001). Extending UML for modeling Web applications. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences (IEEE).*

Brown, A. (2004). *An introduction to Model Driven Architecture*. Retrieved from http://www-128.ibm.com/developerworks/rational/library/3100.html

Casati, F., Shan, E., Dayal, U., & Shan, M-C. (2003). Business-oriented management of Web services. *Communications of the ACM*, *46*(10), 55-60.

Catley, C., Petriu, d. C., & Frize, M. (2004). Software performance engineering of a Web service-based clinical decision support infrastructure. In *ACM SIGSOFT Software Engineering Notes, Proceedings of the 4th international workshop on Software and performance*, *29*(1), 130-138.

Conallen, J. (1999). Modeling Web application architectures with UML. *Communications of the ACM*, *42*(10), 63-70.

Conallen, J. (2003). *Building Web applications with UML*. Boston: Addison-Wesley.

Curbera, F., Khalaf, R., Mukhi, N., Tai, S., & Weerawarana, S. (2003). The next step in Web services. *Communications of the ACM*, *46*(10), 29-34.

Estrella, F., McClatchey, R., Rogulin, D., Amendolia, R., & Solomonides, T. (2004). A service-based approach for managing mammography data. In *11th World Congress on Medical Informatics (MedInfo '04)*, San Francisco, USA.

Fatolahi, A., & Shams, F. (2005). Service discovery based on business processes: a practical approach using UML. In *Proceedings of ICSSSM '05 Conference on Services, Systems and Services Management (IEEE)*, *1*, 417-421.

Fontoura, M., Pree, W., & Rumpe, B. (2002). *The UML profile for framework architectures*. Addision-Wesley.

Foster, H., Uchitel, S. Magee, J., & Kramer, J. (2003). Model-based verification of Web service compositions. In *Proceedings of the 18th IEEE International Conference on Automated Software Engineering*, (pp. 152-161).

Gronmo, R., & Solheim, I. (2004). Towards Modeling Web Service Composition in UML. *Paper presented at The 2nd International Workshop on Web services: Modeling, Architecture and Infrastructure*, Porto, Portugal.

Gronmo, R., Skogan, D., Solheim, I., & Oldevik, J. (2004). Model-driven Web services development. In *Proceedings of the IEEE International Conference on e-Technology, e-Commerce and e-Service*, (pp. 42-45).

Heckel, R., Lohmann, M., & Thone, S. (2003). *Toward a UML profile for service-oriented architectures.* Retrieved from http://www.hwswworld.com/downloads/9_28_05_e/MDAFA-2003-FinalVersion.pdf

Heuvel, Willem-Jan Van Den., & Maamar, Z. (2003). Moving toward a framework to compose intelligent Web services. *Communications of the ACM*, *46*(10), 103-109.

Jiang, Y., Xing, C., He, W., & Yang, J. (2005). On procedure strategy of constructing SOA's modeling language. In *Proceedings of the 2005 IEEE*

International Workshop on Service-Oriented Systems Engineering (SOSE '05), (pp. 71-76).

Kreger, H. (2001). *Web services conceptual architecture.* WSCA 1.0, IBM Software Group.

Leymann, F., Roller, D., & Schmidt, M.-T. (2002). Web services and business process management. *IBM Systems Journal*, *41*(2), 198-211.

Meredith, L. G., & Bjorg, S. (2003). Contracts and types. *Communications of the ACM*, *46*(10) 41-47.

OMG. http://www.omg.org

Papazoglou, M. P., & Dubray, J. (2004). *A survey of Web service technologies* (Tech.Rep. DIT-04-058). Informatica e Telecommunicazioni, University of Trento.

Skogan, D., Gronmo, R., & Solheim, I. (2004). Web service composition in UML. In *Proceedings of the Eighth IEEE International Enterprise Distributed Object Computing Conference* (pp. 47-57).

Staikopoulos, A., & Bordbar, B. (2005). A comparative study of metamodel integration and inter- operability in UML and Web services. In *Proceedings of the First European Conference on Model Driven Architecture, Lecture Notes in Computer Science*, 3748 (pp. 145-159).

Stojanovic, Z., Dahanayake, A., & Sol, H. (2004). Modeling and design of service-oriented architecture. *2004 IEEE International Conference on Systems, Man and Cybernetics*, pp. 4147-4152.

Thone, S., & Varro, D. (2003). A *UML-profile for service-oriented architectures.* Paderborn.

UDDI.org (2002). *UDDI Version 2 Specifications*, http://www.oasis-open.org/committees/uddi-spec/doc/tcspecs.htm

Vara, J. M., De Castro, V., & Marcos, E. (2005). WSDL automatic generation from UML models in a MDA framework. *International Journal of Web Services Practices*, *1*(1-2), 1-12.

Yang, J. (2003). Web service componentization. *Communications of the ACM*, *46*(10), 35-40.

Zimmermann, O., Schlimm, N., Waller, G.., & Pestel, M. (2005). Analysis and design techniques for service-oriented development and integration. *GI Jahrestagung*, 2, 606-611.

This work was previously published in the International Journal of Healthcare Information Systems and Informatics, Vol. 2, Issue 2, edited by J.Tan, pp. 36-52, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 8.5
Evidence-Based Medicine:
A New Approach in the Practice of Medicine

Nilmini Wickramasinghe
Cleveland State University, USA

Sushil K. Sharma
Ball State University, USA

Harsha P. Reddy
Cleveland State University, USA

ABSTRACT

The ongoing tension between certainty over uncertainty is the main force that is driving the evidence-based medicine movement. The central philosophy of this practice lies in the idea that one can never take for granted one's own practice, but by using a structured, problem-based approach, practitioners can logically manoeuvre their way through the obstacle course of clinical decision-making. Attending postgraduate educational events and reading various science journals are no longer sufficient to keep healthcare practitioners aware of all the new developments in practice. To gain this knowledge they need to accept that there are questions they have to ask about their practice. Having posed a number of questions, answers should be found to the most important, practitioners should appraise the quality of the resulting evidence and, if appropriate, practitioners should implement change in response to that new knowledge.

INTRODUCTION

Evidence-based medicine (EBM) can be thought of as the conscientious, explicit and judicious use of current best evidence in making decisions about the care of the individual patient. It means integrating individual clinical expertise with the best available external clinical evidence from systematic research (Sackett, 2003).

Medical practitioners see an overwhelming number of patients. Several questions concerning diagnosis, prognosis, treatment and general care may arise during each session. Hence, it is quite possible for a medical practitioner to make many thousands of clinical decisions a year. As per Covell (Covell & Manning, 1995), when asked, general practitioners say that they generate on average two questions per three patients. Forty percent of the questions were factual (e.g., What dose?), 43% were concerned with medical opinion and 17% were nonmedical. On average, the clinician is left with four unanswered questions per surgery or clinic. A colleague typically meets only 30% of physicians' information needs during the patient visit. Further, the reasons for not using printed resources includes lack of knowledge of appropriate resources, office textbook collections are often too old, as well as a lack of time to search for needed information. In addition, there is no proof that the information obtained is 100% up-to-date. Evidence-based practice can be of great help in this scenario of the background of information needed and clinical overload and the general feeling of helplessness. EBM aims to provide the best possible evidence at the point of clinical (or management) contact.

As per the study by Lundberg (1992), in 1900 there were about 10,000 scientific journals; in 1990 there were more than 100,000 scientific journals. Ninety percent of all major scientific advances are in only 150 of those 100,000 publications; 80% of the citations noted by Science Citation Index are to less than 1,000 journals. Not all of this information is valid or useful for patient care. There is a need to be able to identify relevant information and to be able to critically evaluate the scientific methodology and conclusions of the information.

Evidence-based medicine then focuses on converting the abstract exercise of reading and appraising the literature into the pragmatic process of using the literature to benefit individual patients while simultaneously expanding the clinician's knowledge base (Bordley, 1997).

The primary objective of this chapter is to discuss this new approach of practicing medicine. The need for evidence-based medicine becomes even more pronounced in an environment focusing on lowering healthcare costs. However, it is our thesis that in order to realize the full power and potential of EBM we must integrate the strategies, processes, tools, techniques and technologies of knowledge management. Given that the focus of this entire book is on creating knowledge based healthcare organizations, the goal of this chapter is to familiarize readers with the new and growing area of EBM. From this chapter we believe it will then be possible to appreciate the key role of both information and more importantly knowledge to EBM and thus why it is indeed imperative to incorporate a KM focus in EBM.

STEPS IN EVIDENCE-BASED MEDICINE

The practice of evidence-based medicine can be divided into the following components:

1. Ascertaining a problem or area of uncertainty.
2. Converting information into a focused, clinically important question that is likely to be answered.
3. Efficiently tracking down and appraising the best evidence.
4. Estimating the clinical importance of the evidence and the clinical applicability of any recommendations or conclusions.
5. Unifying the evidence with clinical expertise, patient preferences and applying the results in clinical practice.
6. Summarizing and caching records for future reference.

Ascertaining a Problem or Area of Uncertainty

This is the first step within EBM and must be done with utmost care, especially since all the other steps almost completely rely on this step. A bad identification will lead to form an irrelevant question and an incorrect clinical appraisal that leads directly to a wrong conclusion. This wrong conclusion can be very dangerous and even be harmful to the patient.

A considerable amount of time then should be spent to find the area which needs a literature search. The area of uncertainty should be properly focused and narrowed down before proceeding to the process of formulating the question.

Converting Information into a Focused, Clinically Important Question

The amount of work that the practitioners have to complete is sometimes overwhelming and they forget to ask what they are doing. The questioning attitude is what makes good professionals. Medical practitioners must think through carefully if they in fact are doing the right things and if so are they being done in the right way and at the right time. Further, they should be very accurate when forming a question because in extreme cases a question can be so badly formed that its answer is either unrelated or even meaningless.

Consider the following clinical scenario:

John is a new patient who visits a clinic for a routine physical exam. His medical records from his previous physician are available. John is in good health, although he has had hypertension for many years. There is a family history of stroke. He is 67 years old and he goes on long walks almost every day.

John's hypertension has been successfully controlled by Beta-blockers and he expressed satisfaction with this therapy and he has a feeling that he is getting better. However, John's daughter, who sees another physician, has recently been diagnosed with hypertension and was given Captopril. John wants to know if he can use Captopril. So the problem is whether or not Captopril would be a better medication for John.

A relevant question has to be formed for the above problem. Answers may be found more efficiently by structuring the question. It is, therefore, important to try to break the question down into several parts. This process is also known as "anatomy." The question should be framed in such a way that it facilitates the answer.

The question can be broken into:

- Patient or problem
- Intervention
- Comparison intervention
- Outcomes

Patient or Problem

The first part is to identify the problem or the patient. Healthcare problems may not always seem to be about patients. Sometimes the question may be regarding administrative work, other times the "problem or patient" may be the out of hours of the nurses. It is critical in EBM that the question being described is the patient or problem that is seen. The question must not be too specific at this stage as some important evidence might go unnoticed. There is also a balance to be struck between getting evidence about exactly the group of patients and getting all the evidence about all groups of patients.

Intervention

Intervention may in fact be a postponement of an action such as an operation. Most interventions are more straightforward such as types of dressings, drug therapies or counseling. Alternatively, there can be the provision of differing environmental factors. This portion of the question is important as it confirms the future action to be taken regarding the treatment.

Specificity must be as accurate as possible at this stage as there might be a need to back track if any evidence is not found. The way in which the treatment is offered to patients varies from both within and between primary and secondary care.

Comparison Intervention

Sometimes there is a comparison of intervention. For example, a search can be made for papers comparing the use of head lice lotion in children compared with placebo. Most of the treatments come to practice after performing such comparison tests. Comparative studies offer a great deal of information and help when searching for the evidence.

Outcomes

Outcome measures are particularly important when considering the question. It is worth spending some time working out exactly what it is you want. In some life-threatening or life-altering diseases, most of the times it happens that by concentrating on the mortality, important aspects of morbidity might be neglected. For example, the use of toxic chemotherapies for cancer may affect both aspects (Evidence-based Medicine, 2003).

In the case of John, in the example above, the question structure is detailed in Table 1.

Using this anatomy, the focused question for John would be:

In elderly patients, are ACE inhibitors more effective than beta blockers in controlling high blood pressure and minimizing adverse effects?

Efficiently Tracking Down and Appraising the Best Evidence

One of the questions is formulated. The next step in EBM is tracking down the evidence. This again requires attention to several key steps as described here.

Finding the Evidence

It is important that the most relevant and best available resources are to be searched and utilized for the different types of questions. Best information reserves can be gathered by following a systematic approach. The following five questions can be used as a guide for this process (*Evidence-based Medicine*, 2003).

- *Type of question being asked:* It is useful and also necessary to decide the category in which the question fits into. "Clinical findings," "Therapy," "Prevention," "Prognosis," "Harm or risk" and "Quality of life" are some of the varieties of questions.

Table 1. Structure of the question

Patient / problem	High rate of blood pressure, elderly
Intervention	Beta blockers
Comparison intervention	ACE inhibitors
Outcomes	Reduce blood pressure, minimize adverse effects

- *Choice of information that would provide evidence to answer this type of question:* Careful consideration must be given to the issues of qualitative and quantitative evidence. Sometimes qualitative factors that are impossible or difficult to quantify are involved for some types of questions, including some therapy questions. Though cultural perceptions, values, customs or other types of soft complexity make generalization between groups or population difficult, evidence is needed for decision-making.
- *Type of study that would provide such information:* For each particular type of question, various studies and investigations, which supply the information that can be used as strongest evidence to weakest evidence, must be accumulated. Qualitative research may provide information for complex health policy decisions.
- *Varieties of resources that would give access to the results of the above studies:* Resources can be classified into two fundamental types "primary" and "secondary." Primary sources are the ones in which the original studies are published, while the secondary sources are the ones which refer to or give bibliographical access in the form of references to material in primary sources. Among the primary sources, textbooks along with journals and publications occupy the most important position. With the growth of the Internet, the Web as a source of information is becoming a vital resource. Online information with the aid of powerful search engines such as MSN, Google, Altavista, Yahoo, OMNI, etc., provides links to various sites that have relevant primary and secondary sources of information in an organized way, while computer bibliographic databases like PubMed, MedLine, CINAHL, and EMBASE, etc., are the most important types of secondary resources. These online data sources have links to thousands of electronic journals and research papers from all around the world.
- *How can the best use of the available resources be obtained?* A search can be of two types:
  - High Sensitivity/Low Specificity Search: This search gives importance to comprehensiveness — finding everything of any relevance at all.
  - Low Sensitivity/High Specificity Search: This search gives importance to selectivity — finding only the most highly relevant material.

Clearly there exists a trade-off between the above two categories.

Appraising the Evidence

The main goal of a critical appraisal is to identify the quality of the article. There are three key issues to think about when appraising any paper:

- The validity of the results of the study
- The actual results
- The relevance of the results

Estimating the Clinical Importance of the Evidence

In this step, records are verified to see if the evidence obtained is relevant to the patient's problem and also to verify if the evidence or information obtained is of any clinical importance. Any evidence that is not clinically recognizable must be set aside. It can be considered as supplementary information, but must not be taken as the primary source of information to extract evidence.

The following section will describe how to make a decision regarding the usage of the acquired data.

Deciding Whether or Not to Act on Evidence

The process of evidence-based medicine is one that requires change to be monitored (*Evidence-based Medicine*, 2003). The evidence that supports the validity or truthfulness of the information is found primarily in the study methodology. Here, investigators address the issue of bias, both conscious and unconscious. The investigators or the patients must not overly influence the study results. To ensure this, study methodologies such as randomization, blinding and accounting are applied for all patients. To facilitate the rigor of the methodology, the audit cycle as depicted in Figure 1 is often utilized.

Evaluating the medical literature is a complex undertaking. Once it is determined that the study methodology is valid, the results and their applicability to the patient must be examined. Clinicians may have additional concerns such as whether the study represented people similar to his or her patients, whether the study suggested a clear and useful plan of action, and whether the study covered those aspects of the problem that are most important to the patient.

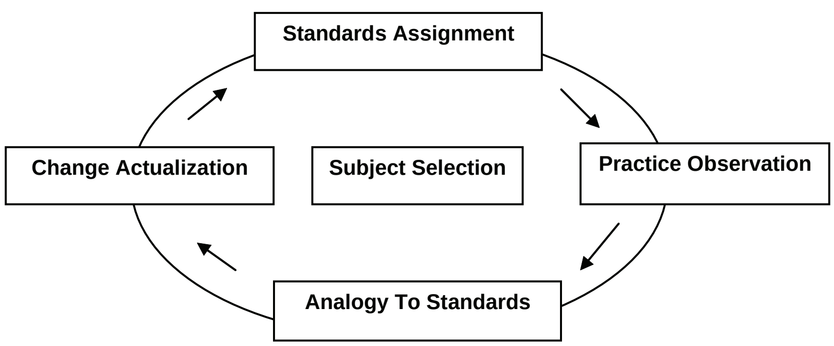
Distinct sources and techniques can be used to achieve and access the clinical importance. Randomized controlled trails are one of the widely used methods. The following are important considerations:

- For the prevention of bias, the groups must be randomized.
- The comparison groups must be similar as different groups might have different characteristics that may lead to bias.
- Patients must be unaware of whether their treatment is experimental or placebo. This is to be considered important as it affects the mental confidence of the patient and may even lead to an altogether different result.
- The length of study must reflect the type of study. It is necessary to have an idea of the natural progress of the disease for determining the appropriate length for a study.
- A good clinical trail requires a complete follow up of both groups. If less than 80% of the patients adequately followed up, then the results may be invalid.
- Appropriate outcome measures must be used in the research of the treatment. It must also be seen that the outcome measures are as accurate as possible.

Unifying the Evidence with Clinical Expertise, Patient Preferences and Applying it to Practice

The final step in the EBM process is to return to the patient and discuss the evidence and suggestions for further treatment.

Figure 1. The Audit Cycle



Suppose, for example, the analysis shows that the evidence is evaluated and accepted and that it meets the criteria for validity. To complete the analysis a review of the results must be performed. Care must also be taken to determine if they are applicable to the particular patient.

Let us return to the example given earlier for the patient John. The following treatment can be suggested. Suppose the results of this study indicate that ACE inhibitors are not more effective than beta-blockers in treatment of hypertension and that they lead to an increased risk of stroke. For John (based on his situation, the physician's clinical judgment, and the evidence) it does not seem appropriate to change to ACE inhibitors. Since there is a family history of stroke, his daughter is not a good candidate for ACE inhibitors, either. A copy of the study can be sent to her to share with her physician so that her physician will have a chance to go through the study to determine if the treatment being offered to John's daughter needs any improvements.

Summarizing and Caching Records for Future Reference

This is the last step in the evidence-based practice. It is not a mandatory step but is highly recommended. As there is a high probability for that physician to get cases very similar to the one he conducted research on, it is always better for him to store and organize all his research. His own work can become the most important resource of information for his future research.

MISAPPREHENSIONS ABOUT EVIDENCE-BASED MEDICINE

Misapprehensions arise naturally regarding any issue. They must be discussed and clarified. In addition, it is important that any of the unintended negative effects must also be taken care of and rectified immediately. While practicing or considering the practice of evidence-based medicine, the nature of the new paradigm is sometimes misinterpreted.

Recognizing the limitations of intuition, experience, and understanding of pathophysiology in permitting strong inferences may be misinterpreted as rejecting these routes to knowledge. Specific misinterpretations of evidence-based medicine, and their corrections, are provided in Table 2.

BARRIERS TO PRACTICING EVIDENCE-BASED MEDICINE

There are some difficult challenges to be faced regarding the practice of EBM. These include but are not limited to: (1) the appropriate literature may not be readily available; (2) economic constraints and counterproductive incentives may compete with the dictates of evidence as determinants of clinical decisions; and (3) sometimes the evidence obtained may be capacious, and there might not be a sufficient amount of time for careful revision relevant to a pressing clinical problem.

Some solutions to these problems are already available. Optimal integration of computer technology into clinical practice facilitates finding and accessing evidence. Reference to literature overviews meeting scientific principles (Oxman and Guyatt, 1998) and collections of methodologically sound and highly relevant articles (Haynes, 1991) can markedly increase efficiency. Evidence-based summaries are becoming increasingly available. Using online health information for obtaining evidence is already in practice. Furthermore, it is gaining importance. Increasingly, scientific overviews will be systematically integrated with information regarding toxicity and side effects, cost, and the consequences of alternative courses of action to develop clinical policy guidelines. The prospects for these developments are both bright and exciting. It is with respect to capturing key knowledge, in particular tacit knowledge, and storing this knowledge so that it can be eas-

Table 1. Evidence-based medicine and their corrections

Misinterpretation	Correction
Evidence-based medicine ignores clinical experience and clinical intuition.	This is not true. Untested signs and symptoms should not be rejected out of hand. They may prove extremely useful, and ultimately be proved valid through rigorous testing. Diagnostic tests may differ in their accuracy depending on the skill of the practitioner.
Understanding of basic investigation and pathophysiology plays no part in evidence-based medicine.	The dearth of adequate evidence demands that clinical problem solving must rely on an understanding of underlying pathophysiology.
Evidence-based medicine ignores standard aspects of clinical training such as the physical examination.	A careful history and physical examination provides much, and often the best, evidence for diagnosis and directs treatment decisions. Evidence-based practice considers the physical conditions of the patient while evaluating the evidence and also before applying treatment to the patient.

ily and efficiently accessed and hence used and reused that we believe the strategies, processes, tools, techniques and technologies of knowledge management have the greatest value in facilitating EBM, addressing its key barriers and enabling it to reach its full potential.

Concerns Regarding Adoption Of Evidence-Based Medicine

Although EBM has heightened awareness of the most effective management strategies, much of the evidence is not acted on in everyday clinical practice. In particular, five main themes have been identified that affect the implementation process (Freeman and Sweeney, 2001). These include:

Personal and Professional Experiences of the General Practitioners

Doctors' personal and professional experiences influence how clinical evidence is implemented. Despite being a relatively homogeneous group, the general practitioner's enthusiasm for the evidence and the way in which they implemented it varied. Thus, there is no proof that whatever approach is assumed as the best depending on the evidence is actually the best or not.

The Patient-Doctor Relationship

Implementation is influenced by the relationships that doctors developed with their patients. Perceived patient characteristics could have a positive or negative effect on implementation. It also depends on the confidence of the patient in the physician and also in the treatment.

A Perceived Tension Between Primary and Secondary Care

Implementation is also influenced by the relationships of doctors with secondary care doctors. Specialists approached evidence-based practice differently, treating "diseases rather than patients." Based on the same evidence, a general practitioner may think one approach as the best one, while a cardiologist might think the same approach as the least approachable.

General Practitioners Feeling about Their Patients and Evidence

Clinical evidence is not just an intellectually celibate commodity that is lifted out of medical journals and transferred to a patient. It has an emotional impact on practitioners and patients. Sometimes the knowledge that the evidence existed, waiting to be applied, was seen as a burden in itself. It may make the patient more anxious. Some practitioners may forgo applying the evidence due to the patient's unwillingness to a new kind of treatment.

The Logistical Problems

Some tricky logistical problems make the doctors less enthusiastic about implementing clinical evidence. Many physicians find it risky and a hassle both for doctors and patients to start a new treatment. Some patients might feel the same. Many of the problems can be resolved by the incorporation of knowledge management. For some problems, using the knowledge management concepts can reduce the intensity of its effect.

CONCLUSION

Based on an awareness of the limitations of traditional determinants of clinical decisions, a new paradigm for medical practice has arisen. Evidence-based medicine deals directly with the uncertainties of clinical medicine and has the potential for transforming the education and practice of the next generation of physicians. These physicians will continue to face an exploding volume of literature, rapid introduction of new technologies, deepening concern about burgeoning medical costs, and increasing attention to the quality and outcomes of medical care. The likelihood that evidence-based medicine can help ameliorate these problems should encourage its dissemination.

EBM is still in its infancy. Further, it requires new skills of physicians and other medical practitioners. While strategies for inculcating the principles of evidence-based medicine remain to be refined, initial experience has revealed a number of effective approaches. An evidence-based approach is robust, informative and feasible. It also promotes the quality and consistency of healthcare delivery, and it ensures that the services are cost effective. This chapter has served to highlight the key issues pertaining to EBM. By doing so, it highlights the integral role for KM within any EBM initiative to facilitate the generation, capture, storing, use and reuse of critical medical knowledge. We close by calling for further research into this key area of healthcare in today's 21st Century.

REFERENCES

Bordley, D. R. (1997, May). Evidence-Based Medicine: a Powerful Educational Tool for Clerkship Education. *American Journal of Medicine*, 102(5), 427-432.

Campbell, E. J. (1987, April). The diagnosing mind. *PubMed, Lancet*, 1(8537), 849-851.

Covell, D., Uman, G. & Manning, P. R (1995, October). Information Needs in Office Practice: Are They Being Met? *Annals of Internal Medicine*, 103(4), 596-599.

Dawes, M., Davies, P., Gray, A., Jonathan, M., Seers, K. & Snowball, R. (2003). Evidence-based practice — A primer for healthcare professionals. *Evidence-based medicine electronic tutorial*. Retrieved from http://www.hsl.unc.edu/lm/ebm/welcome.htm

Evidence-based Medicine Working Group. (2003). Retrieved from the World Wide Web: www.med.ualberta.ca/ebm/ebmintro.htm

Freeman, A. C. & Sweeney, K. (n.d.). *Why general practitioners do not implement evidence: Guidelines for reading literature reviews.*

Haynes, R. B. (ed.). (1991, January/February). ACP J Club: A18. *Annals of Internal Medicine*, 114, S1.

Introduction to Evidence-Based Medicine. (2003). Retrieved from www.hsl.edu/lm/ebm/welcome.htm

Lundberg, G. D. (1992, April). Perspective from the editor of JAMA. *The JAMA Bulletin of the Medical Library Association*, 80(2), 110.

Oxman, A. D. & Guyatt, G. H. (1988, April). Qualitative Study. *CMAJ*, 15, 138(8), 697-703.

Sackett, D. (2003). Evidence-based medicine — What is it and what it isn't. *The Origins and Aspirations of ACP Journal Club.* Retrieved from www.minervation.com/cebm/ebmisisnt.html

This work was previously published in Creating Knowledge-Based Healthcare Organizations, edited by N. Wickramasinghe, J.N.D. Gupta, and S. Sharma, pp. 125-135, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 8.6
Charting Health Information Technology Futures for Healthcare Services Organizations

Avnish Rastogi
Providence Health & Services, USA

Tugrul Daim
Portland State University, USA

Joseph Tan
Wayne State University, USA

ABSTRACT

As health organizations strive to improve operational efficiencies and increase worker productivity, new forms of health information technologies (HITs) are constantly being developed. This article surveys the extant HIT literature and adopts a case analysis approach to identifying emerging health information technologies. The understanding of HIT trend is further enriched through the applications of technology-forecasting techniques, specifically, scenario analysis and U.S. patent searches. The article focuses on five emerging HITs and their impact on the future of U.S. Healthcare Services Delivery.

INTRODUCTION

The advent of information technology (IT) and telecommunications has altered beyond recognition the traditional business processes in many industries. The impact of IT on the health industry is also gaining attention. In the past, researchers have noted the lag of IT diffusion across the health sector vis-à-vis most other industrial sectors. With the diffusion of health databases and automated data warehouses, routinely extracted health data from multiple sources can now be more fully integrated and mined to evolve a new type of knowledge that promises to transform the health industry. Physicians, aided by intelligent clinical support systems and automated devices, for example, can now provide better patient care by drawing on programmed reminders and alerts and other computerized protocols of guided treatments for specific diseases. Emerging IT has made physician practices more efficient and effective while new research in the area of optimizing user interface designs for physicians and other health professionals promises to further enhance such capabilities (Zheng et al., 2005).

In this age of knowledge explosion, health information technology (HIT) is beginning to permeate many areas of human life. Tan (2005) documented the various strategies and speed with which emerging HITs are transforming the healthcare industry. Telemedicine today encompasses a common infrastructure including physical facilities, and equipment used to capture, transmit, store, process and display voice, data and images. Technological advances such as fiber optics, satellite communications, and compressed video have minimized expansion costs and limitations, thereby initiating a growth of public and private sector interests. These newer, alternative modalities of healthcare delivery are beginning to affect the cost, quality, availability and accessibility of healthcare. HIT has become a crucial part of the routine practices of healthcare professionals. It is not unusual nowadays for health consumers to check medical conditions of their own or those of their significant others on the Internet. A recent Forrester survey of 1,300 doctors reports that 80% of physicians believe technology "makes life easier."

Specific to changing the patient-physician relationship and the role of organizational health services delivery in the United States (U.S.), emerging technologies are clearly dictating people to take a greater and more active role in maintaining their own health. The result is a proactive prevention-oriented, consumer-driven model for healthcare that includes capabilities such as "smart devices" that can "think," customized wearable devices, electronic patient records, and wireless Internet-linked systems. Over time, these technologies are expected to deliver more convenient, user-friendlier, and more consumer-driven, home-healthcare services, unlike traditional services delivered at a physical and distant medical facility. For consumers, this could mean added convenience, reduced travel time and costs. They will enjoy easy-to-use homecare systems and products that teach them to monitor themselves with gizmos that give timely warnings of illness so that they can turn to their physicians early when appropriate interventions will do the most good. For doctors, it could mean more timely and effective patient care.

In this article, a selection of emerging HITs with a focus on their potential impact on the future of organizational health services delivery is reviewed. The analysis of emerging HIT applications is based on key observations of Providence Health & Services (PHS) organizations, one of the leading health providers in Pacific Northwest that implements cutting-edge technologies to improve patient care and reduce healthcare cost.

The review of PHS presents the necessity of HIT applications. Onsite formal and informal interviews and discussions with key experts helped identify emerging HITs deserving attention. Experts were also consulted to verify the interview results. A twofold approach followed the interviews. Specifically, a combination of technology forecasting tools including "Patent Analysis" and "Scenario Analysis" was used to predict the future of these emerging technologies.

Each of the technologies identified is either replacing an existing one or evolving as a new technology, for example, radio frequency identification (RFID) in healthcare is replacing barcode technology whereas e-health is an emerging paradigm. This article therefore, represents the initial steps of a larger research framework for assessing and forecasting the diffusion of particular HIT and their potential impact on the U.S. healthcare services delivery system.

BACKGROUND

Research on technological assessments in the health sector has grown over the last several years. Table 1 highlights key but selective literature on efforts to implement HITs across the spectrum of healthcare services organizations. Three themes arise from this review: common HIT applications, health decision support tools and emerging HIT applications.

Unfortunately, the adoption of IT has not always been successful. Indeed, there is evidence to show that many system implementations across all industry sectors ended as failures when measured by the ability of top management to ensure timely delivery of a specific product within budget (Ein-Dor & Segev, 1978; Ginzberg, 1981). In particular,

Table 1. Selective literature on health information technologies

Area	Title	Year	Author
Introductory Reviews of Technology Management in Healthcare	Diffusion of innovations in health service organizations: A systematic literature review	2005	Greenhalgh, Robert, Bate, Macfarlane, Kyriakidou
	IT implementation in healthcare	2001	Berg
	The transfer and commercialization of university- technology	1999	Del Campo, Sparks, Hill, Keller
	Management of healthcare technology literature	1999	Kumar, Motwani
	Assessing technological barriers to telemedicine: Technology-management implications	1999	Paul, Pearlson, Mcdaniel
	Organizational management of telemedicine technology: Conquering time and space boundaries in healthcare services	1999	Sheng, Hu, Wei, Ma
	Technology in American healthcare	2004	Cohan, Hanft
	Technology and healthcare in an era of limits	1992	Gelijns, Institute of Medicine
	Technology, healthcare, and management in the hospital of the future	2003	Geisler, Krabbendam, Schuring
	Organizational factors that influence IT diffusion in academic health science centers	1997	Ash
Decision Support Tool	Health decision support systems	1998	Tan, Sheps
	A multi-attribute measure for innovation adoption	1999	Wilson, Ramamurthy, Nystrom
	Understanding technology adoption in clinical care: Clinician adoption behavior of a point-of-care reminder system	2005	Zheng, Padman, Johnson, Diamond
	Health technology and decision making	2005	OECD

continued on following page

Table 1. (cont.)

Emerging Health IT Applications	E-healthcare information systems	2005	Tan
	Ethics and information technology	2002	Anderson, Goodman
	The strategic application of information technology in healthcare organizations	2002	Glaser
	Evaluating informatics applications – some alternative approaches: Theory, social interactionism, & call for methodological pluralism	2001	Kaplan
	The diffusion and value of healthcare information technology	2005	RAND
	The behavioral side of information technology	1999	Dixon
	IT-adoption and the interaction of task, technology and individuals: A fit framework and a case study	2006	Ammenwerth, Iller, Mahler
	Assessing the level of healthcare information technology adoption in the United States: A snapshot	2006	Poon, Jha, Christino, Honour, et al.
	Individualization, globalization, and health – about sustainable information technologies and the aim of medical informatics	2006	Haux
	The impact of converting to an electronic health record on organizational culture and quality improvement	2006	Nowinski, Becker, Reynolds, et al.
	User interpretations of future information system use: A snapshot with technological frames	2006	Karsten, Laine
	Expanding multi-disciplinary approaches to healthcare information technologies: What does information systems offer medical informatics	2006	Chiasson, Reddy, Kaplan, Davidson

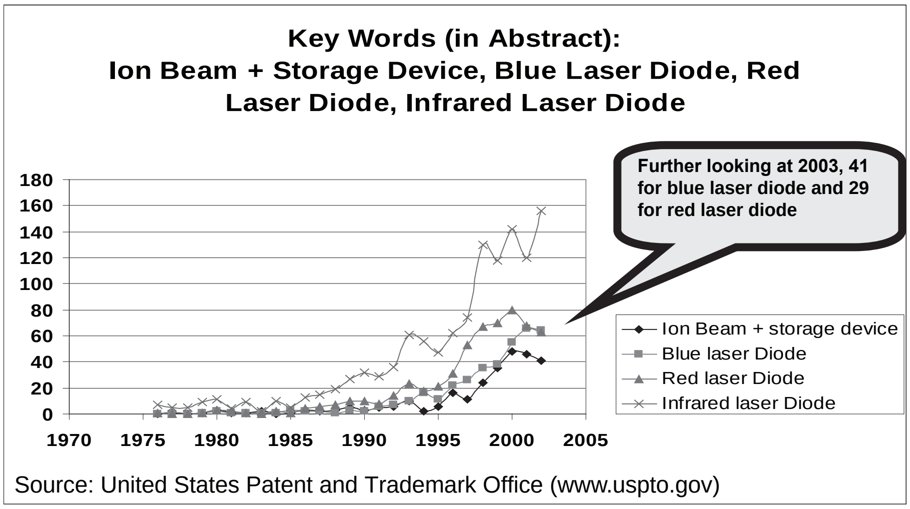
a factor that can also cause implementation challenges encountered among academic healthcare centers is the lack of "across-the-organization", top-down leadership philosophy (DeWitt & Hampton, 2005). Owing to the way in which healthcare provision is organized, Lorenzi and Riley (2000) observed that small group theory might be highly applicable to an evaluation of success factors for implementing HITs in these centers. Anderson (1997) further noted that organizationally mandated initiatives are not generally met with success, and in fact, might result in "evasive actions or even sabotage." The organizational dimensions of successful HIT implementation is of keen interest and has been noted as an area in need of further research (Larsen, 2003; Meijden, et al., 2003; Ash, 2004).

Technology Forecasting

Little research has focused on the application of technology forecasting methods to HIT applications. To narrow the focus of our analysis, we home on those emerging HITs that may significantly impact health services based on experts' opinions. Additionally, we are only interested in HITs that will directly improve patient care or result in rapidly improved organizational efficiencies within the U.S. healthcare services delivery system. Technology forecasting is aimed at predicting future technological capabilities, attributes, and parameters. It is a step in the process of anticipating and bringing about tomorrow's evolving technologies (Martino, 1983).

IT forecasting has become a widely valued technique during the past decade from the perspectives of both individual companies and entire industries. Innovation, based on new technology,

Figure 1. The number of U.S. patents by technology



plays an enormous role in any company's success. Innovative companies generally enjoy greater sales, profitability and global market share. To accelerate the pace of innovation, it is essential that there is sharing of greater collaboration among partners having common innovation goals. Several alternative methodologies are used in technology forecasting. Cauffiel and Porter (1996) use Delphi method to explore Electronics Manufacturing in 2020. Linstone (1999) uses multiple perspectives to assess any decision regarding future applications of a technology and demonstrated its use for technology forecasting. Kayal (1999) is a leading researcher who has used patent analysis for forecasting technologies. Watts and Porter (1997) introduce the use of bibliometrics for the same purpose. Sager (2003), Winebrake et al. (2003) and Silberglitt et al. (2003) use scenario analysis to explore biotechnology, hydrogen fueling systems and energy futures.

Daim et al. (2006) use patent analysis and bibliometrics to trend emerging technologies. Figure 1 depicts a graphical analysis of one of their study results. As shown, emerging technologies under exploration reach a plateau in terms of granted patents at a certain time period. In the case of blue laser diode, it was in and around the year 2000 that this plateau was reached and in late 2006, we are witnessing commercial applications just being released, which may be explained as the time required for commercializing the technology.

Governmental decisions to support some technologies and not others will also impact differently on the speed of technological innovation and development to market. If technological forecasting predicts that a certain capability is technologically feasible and is within our reach in the near future, and if the government chooses to support research and invest in this area, it is more likely that the technology will be developed and marketed successfully.

The major techniques for technology forecasting may be divided into two categories: (a) methods based on numeric data and (b) judgmental methods. The tool applied to forecast each technology was based on individual technology's characteristics. "Patent Analysis" was used for technologies grouped under the category of emerging technologies whereas "Scenario Analysis" is appropriate for studying technologies that are more suggestive of an evolving concept. In our analysis, both of these approaches will be applied to HIT forecasting.

THE PROVIDENCE HEALTH AND SERVICES (PHS) CASE

Providence Health & Services (PHS), with information systems centered in Seattle, is a nonprofit health provider organization with presence in Oregon, Alaska, California, Washington and Montana. PHS includes 28 hospitals, more than 35 non-acute facilities, physician clinics, a health plan and education facilities. Figure 2 maps out PHS services in the different regions.

PHS is the sixth largest Catholic system in terms of net service revenue and the largest healthcare provider in Oregon, Alaska and Washington as well as the largest employer in many communities. In 2005, PHS had seven million visits, which is equivalent to filling every NFL stadium in the country to three times their capacity.

In the Oregon region, PHS offers the following services:

1. **Benedictine Nursing Center:** Founded in 1957 by Benedictine Sisters, the center offers different level of services, including in-home services, nursing center, garden home, orchard house, rehabilitation services and home health
2. **Home Services:** This is a fully integrated Division of Providence Home & Community Services, including home infusion, offering an alternative to lengthy hospital stays for medically stable patients requiring IV

Figure 2. Maps of providence health & services

therapies; home health, providing round-the-clock skills, support and assistance for patients and in-home caregivers; home oxygen and medical equipment, servicing a wide variety of products and related patient services at home

3. **Outpatient Diabetes Education Program**
4. **Behavioral Health Services:** The program offers a complete array of services, from acute inpatient stay to outpatient follow-up
5. **End of Life Program**, including hospice care: Tailored to meet each individual resident and residents' family needs and desires at the end of life; personal and palliative care; and center for healthcare ethics
6. **Children's Services:** This program provides specialized medical care and support services to meet the needs of children and family
7. **Financial Assistance:** The financial aid program offers several ways to help patients with their medical bills
8. **Center on Aging:** The geriatric program offers wide range of services to meet differing needs of seniors
9. **Volunteer Services:** This program offers an avenue for the involvement of volunteers
10. **Brain Institute:** The brain program provides services for treatment of stroke, Alzheimer, Multiple Sclerosis, neurosurgical conditions and epilepsy
11. **Employer Services:** This program offers a variety of services to meet the needs of every workforce

At PHS, a significant number of different HIT applications are supporting key healthcare services. Table 2 lists and maps various support HIT applications to PHS services.

Today, the rapid rate of technology evolution demands technology planning in terms of both forecasting and management. Forecasting technologies is among one of the most formidable challenges that IT managers face. Indeed, it may be argued that getting ready for future HIT is key to managing well in any organization and to gain a market competitive advantage. Our

Table 2. PHS key services and supporting IT applications in the Oregon region

Service	Application	Function
Behavioral Health Services	BHCD	Provide Client Encounter Database for BHS Programs.
Benedictine Nursing Center	PathLinks	Financial and Clinical. Patient-centered, process-oriented software system that helps long-term care facilities manage clinical and financial business operations
Radiation Oncology	Impac	Primary function to monitor dosage and record radiation treatments for Oncology Patients. Secondary functions Physician Notes, Scheduling, Report delivery, & Transcription
Medical Foundation	Donor-II	Used for tracking donor information and contribution.
Diagnostic Imaging	STAR Radiology	Supports daily operational functions of the Radiology departments.
	Mammography Reporting System	Radiology application used to store patient demographics, orders and results for the mammography department.
	UltraPacks	Ultrasound image archive and review system with electronic reporting
Surgical Services	HSM	Surgical Services Case Management, Perioperative Charting, and Supply Management System.
	ORTracker	Manages daily surgery cases, actively tracks equipment and patients in surgery area, and displays nurse call info.

interviews with the IT experts concur with this observation.

Emerging HIT Applications

Americans are living longer and better: by 2050, it is projected that there will be one million centenarians, up from 71,000 today (usatoday.com, 2005). Hence, it is becoming critical to identify key technologies to help improve healthcare services, especially for the U.S. aging population.

Emerging information-intensive technologies in healthcare are converging to revolutionize self-care. This leads to consumers playing a greater role in maintaining their own health. Technologies such as wireless, sensor, wearable and radio frequency identification devices (RFID) are clearly coming in support of a prevention-oriented, consumer-driven model of healthcare. The U.S. patents database is searched to identify emerging HITs and to understand emerging technological trends, along with a comprehensive review of current literature and the Internet. Interviews with several key industry experts yield the following five emerging HIT as an extrapolation from our PHS case analysis: (a) E-Health; (b) Semantic Web; (c) Wearable Monitoring Devices; (d) RFID; and (e) Artificial Intelligence (AI).

E-Health

E-Health is the new paradigm for convenient medical information exchange from one site to another via electronic communications to improve and monitor patient health status. The World Health Organization (WHO) defines "e-Health" as "the cost-effective and secure use of information and communications technologies in support of health-related fields, including health-care services, health surveillance, health literature, and health education, knowledge and research." If sensibly deployed, e-health is expected to improve healthcare quality, cost-efficiency, and patient care. Key benefits for e-Health initiatives include improved provider inefficiencies, greater care accessibility, cost containment, better performance reporting and process efficiencies, enhanced patient safety and quality, reduced distance barriers in healthcare, and increased speed and quality of information capture, storage, organization and retrieval.

E-Health adoption involves overcoming several key barriers. The lack of interoperable systems is a major barrier. Reimbursement of e-visits is another source of concern to physicians as they face consumer demands for greater access to a limited resource—their time. The startup cost for e-Health may also be high as it often requires significant change in health IT infrastructure, training of healthcare professionals as well as anticipated changes in patients' behaviors. Cross-state licensure is another issue; in essence, a health provider or pharmacy must be licensed in every state in which it serves clients. Moreover, one of the areas, which remain challenged when it comes to the future of e-Health, is aligning economic interests across the U.S. healthcare delivery system. For example, there remains a real need to dissect the entire value chain of traditional prescribing and truly understand where digitizing the process saves money, time, or human lives. Understanding who benefits, when they benefit, why and to what extent they benefit is critical to understanding who should invest, why and to what level. For example, if the provider physician receives little to no economic value from digitizing the prescription order entry, why should s/he pay for this benefit? Finally, privacy standards, guidelines for creating healthcare sites, and language barrier are all key challenges facing the adoption and diffusion of e-healthcare. Medical Web sites, more that any other type of site on the Internet, should follow strict personal privacy codes to prevent individuals' personal medical information being compromised.

Scenarios and Strategies

The domain of e-Health is very broad, ranging from Electronic Medical Record (EMR) to e-Prescription and e-Visit, to telemedicine and tele-care services. Each of these areas may be further conceptualized in terms of a series of related projects. Using patent analysis in such a broad domain will not be very helpful in providing any insight into technology forecasting. Instead, scenario analysis for e-Health forecasting is proposed.

The healthcare environment is very volatile. The uncertainty surrounding market forces such as employee benefit structure, the future of Medicare managed care, and the impact of consumer behavior will all affect healthcare future. In the last decade, consumer expectations and their needs are one of the driving forces for change in healthcare. Other forces driving change are genetics, medical technology, population demographics as well as profile of disease. Today, consumers are taking greater responsibility for their own health. They are increasingly demanding choice, quality, convenience, and personalized services in healthcare. Scenario analysis overcomes this limitation by acknowledging the uncertain healthcare environment and articulating a set of plausible alternative futures. Scenario analysis is especially suited for analyzing the current healthcare climate as characterized by the significant level of uncertainty surrounding critical market forces. Key market forces today include: (a) collective bargaining of physicians; (b) health plan consolidation; (c) employee benefit structure; (d) Medicare managed care; (e) healthcare inflation; (f) consumerism; (g) provider payment structure; and (h) physician practice structure.

As depicted in Table 3, with PHS and other health services organizations, there are four possible scenarios that could provide insights into e-health technology forecasting. Analysis of alternatives under these various scenarios will

Table 3. E-health scenarios

Increasing Healthcare Cost	Choice of Freedom
Scenario Healthcare cost is exceeding general inflation. Employers would move to a defined and fixed healthcare coverage. Solution Consumers would make their own decision regarding the purchase of insurance and healthcare services.	Scenario Again, employer healthcare cost wouldn t be growing faster than general inflation. Employers would continue to offer limited choices. Solution Consumer-driven choice and low inflationary pressure have several strategic implications for hospitals and physicians.
Federal Government Role	**Medical Inflation**
Scenario Federal Government may pass series of reforms to manage healthcare costs for both providers and payers. Solution Federal scrutiny would force providers to shift their focus on corporate compliance and policy development. It would shift focus from payers to hospitals and physicians.	Scenario Managed care payers would be the dominant market force, setting contracting and coverage parameters and, thus, making the greatest profits. Solution Payers would require providers to submit information that would allow quality evaluation to occur.

provide better future readiness for healthcare service providers.

- **Increasing medical cost:** In this scenario, the increase in medical costs greatly exceeds general inflation, and the consumer chooses to take an active role in healthcare decision making. Employers would move to a defined-contribution approach to health coverage and provide employees with a fixed dollar amount every month. Consumers would make their own decisions regarding the purchase of insurance and healthcare services, seeking value from hospitals, physicians, and insurers. The combination of individual purchasing discretion and greater demand for information would lead to new models of e-Health contracting.
- **Choice of freedom:** Because employer healthcare costs would not be growing significantly faster than general inflation, companies would continue to offer their employees a choice of plans and providers with a defined benefit. Consumers would take an active role in healthcare decision making within the defined limits of their coverage. Although freedom of choice would exist, the market forces of supply and demand would serve to ration care and access. Consumer-driven choice and low inflationary pressure have strategic implications for hospitals and physicians such as competing on health services portability.
- **Federal government role:** Concerned with the high rate of medical inflation, the Federal government would pass a series of reforms that would establish active Federal oversight and regulation of both providers and payers. Faced with increased Federal scrutiny, providers would focus a significant portion of their resources on corporate compliance and policy development. It would shift risk from payers to hospitals and physicians primarily through capitation. Hospitals and physicians would re-evaluate the role of the integrated delivery system to optimally match their organizational structure with the industry's risk-based payment system.
- **Medical inflation:** In this scenario, managed care payers would be the dominant market force, setting contracting and coverage parameters and, thus, making the greatest profits. In an attempt to define and measure provider quality, payers would require providers to submit information that would allow quality evaluation to occur. Because of consumer indifference to choice among healthcare providers, there would be a shift from open-access products to closed-panel models presided over by gatekeepers. Federal legislation would wane because employers would be content with the relatively low rate of medical inflation, and consumers would not demand governmental intervention, as perceived problems would be minor. Hospitals and physicians would consider aggressive responses to the payers' strong market position. Hospitals would consider consolidation in an attempt to increase their bargaining power.

Semantic Web

Traditionally, pharmaceutical companies have relied on research about how specific diseases work and drawn on knowledge that is in the public domain as its starting point for the discovery of new medicine. In *Pharma 2010: The Threshold of Innovation*, IBM described how the molecular sciences will help the industry define diseases more accurately and create healthcare packages for patients with specific disease subtypes, rather than making one-size-fits-all drugs for patients with similar symptoms but essentially different diseases (Arlington, 2003; IBM, 2004). Similar to the views expressed by IBM, Dr. Allen Roses, worldwide VP of genetics at GlaxoSmithKine, clearly believes that the future of pharmaceuticals lies in being able to target drugs to a smaller number of patients with specific genes and disease subtypes, rather than in selling as many drugs as possible to as many patients as possible (Connor, 2003).

Still, the development of targeted treatment solutions hinges on a much more precise understanding of disease, including the mechanisms and pathways involved in particular disease states and the differences between related disease subtypes. Tools that can map the genes associated with susceptibility to particular illnesses will be vital in developing targeted treatment solutions. A British researcher, Tim Berners Lee, who also introduced the concept of the Internet in 1989, first introduced the "Semantic Web" concept in 2001. The Semantic Web would represent a World Wide Web of connected data, radically different than today's Web of discrete data. It could be the answer to affordable electronic health record.

The Semantic Web makes it easier for data on the Web to be shared and reused by people and applications. The technology allows taking better advantage of information on the Web. It allowed users to perform machine-readable queries and aggregate similar data and information. It is based on the W3C's Resource Description Framework (RDF), which uses XML to integrate applications. The technology has the potential to change the way the health industry behaves; it can also change the way health researchers share and access data. But like any new technology, there is an adoption cycle that must be overcome. With Semantic Web, data from different Web sites are collected into a network, which knows how to gather and assemble data. It allows users to change the information presented based on the interest. With one click, users can view all the information collected about the target; moreover, with another click in the dashboard, users can explore the relationships between various data entities.

The World Wide Web Consortium (W3C) (2005) has developed standards:

- RDF - data description and identification
- Web Ontology Language (OWL) - an ontology language
- A Semantic Web Rule Language (SWRL)

The Semantic Web uses these standards in conjunction with existing data formatting and tagging such as XML and Life Science Identified (LSID). In Semantic Web, elements are defined in statements called Semantic Web triplets, which contain a subject, predicate, and object.

According to W3C, embedding of semantics into medical and research information will offer researchers and health providers better access to information needed to find cures for diseases, make drugs safer and more affordable, and enable them to offer individualized clinical management. On November 22, 2005, W3C announced the formation of the Semantic Web for Healthcare and Life Sciences Interest Group, which is to deploy standardized Semantic Web specifications to provide services defined by the user community. The Semantic Web contains three layers: XML as a syntax layer, a RDF to provide machine-readable descriptions of data that can be parsed; and a third layer, OWL, to combine ontologies, or descriptions of specialized knowledge on the Web. The Semantic Web leverages some existing technologies as well as defines some new technologies. It leverages XML (Barners Lee et al., 1998), a data representation format that allows data to be tagged so that it is easier to distinguish among different pieces of data. It also leverages uniform resource identifiers (URIs) (Butler, 2003), a way of distinguishing elements not just within a document, but also within the space of all documents. URIs make it possible to identify resources and properties.

Major challenges of Semantic Web include:

- **Metadata:** One prerequisite for Semantic Web is that authors of Web sites and Web documents provide richer machine "processable" information, that is, metadata.

Until 1999, there were several ways to link metadata to Web documents. The W3C tried to unify these different approaches in the form of RDF. The RDF allows describing a concept rather than a Web document. Yet, Web content is still largely devoid of metadata labels.

- **Semantic word:** The word semantic is interpreted in different ways, oftentimes causing confusion. People are often willing to argue the advantage of RDF over XML. This different interpretation is illustrated in Note (Butler, 2003). In Computer Science, "Semantic" refers to a description that explains how to interpret a particular syntactic structure in the context of an abstract language (Butler, 2003). In linguistics, "Semantic" connotes formal symbol manipulations, which do not have any intentionality; they are quite meaningless; they are not even symbol manipulations, since the symbols do not symbolize anything. Hence, in the linguistic jargon, there is only syntax but no semantics.
- **Multiple vocabularies:** The Semantic Web supports the philosophy that central naming authorities are unrealistic. It is based on the fact that multiple vocabularies are reality of life and it will be necessary to interoperate among them. It should be possible to map property from one vocabulary to another vocabulary if both are associated with URIs but there is no standard way of doing this in RDF.
- **Technology barrier:** One barrier in the adoption of Semantic Web is RDF complexity, leading to another issue. Programmers, when faced with complexity, often adopt dirty and workaround tricks. When people encounter complexity, it prejudices them against Semantic Web. Implementing RDF parsers is complex and difficult and the best tools are hopelessly slow. These are the most basic and fundamental tools the Semantic Web needs to operate.
- **Business barriers:** If the Semantic Web is implemented, the current e-business models will be intensely disrupted. Ebay, Google,

Table 4. Semantic Web scenarios

Integrated Access to Biological Data Scenario Provide an unified point of access to different biological data repositories accessible through the Internet. Solution Application of semantic technologies to solve the inherent features of the biology field	**Hospital Information System** Scenario Data in a healthcare information system is dispersed and heterogeneous in a setting where speed of access and common presentation are important. Solution Integration and subsequent mediation of medical databases at the semantic level.
Data Warehousing in Healthcare Scenario A large health insurance data warehouse to ensure data integration and consistency. Solution Introduce a common terminology for healthcare data and wrap all legacy data in this terminology.	**Managing Biomedical Research Notes and Raw Data** Scenario Biomedical research notes and raw data are difficult to share because they are not understandable to others unless they are well structured. Solution A flexible and simple to use system for entering notes in both a structured and flexible manner, with complex interlinking.

and Amazon—virtually all mainstays of web-businesses—will have to significantly adjust their business and technology models.

As with other industries, these same challenges in applying Semantic Web apply to the healthcare industry. To look into its future in the healthcare industry, scenario analysis is suggested for Semantic Web because of the need to analyze the current healthcare climate, which is characterized by a significant level of uncertainty regarding IT adoption.

Table 4, which is self-explanatory, provides the scenarios and potential solutions for applying Semantic Web in healthcare. No matter how slow the health industry is in adopting new IT solution, these scenarios will drive Semantic Web technology in mainstream healthcare.

Just as with the e-Health domain, Semantic Web is a topic requiring a very broad search and Patent Analysis would provide little, if any, insight.

Wearable Monitoring Devices

The U.S. and international patient monitoring markets have grown rapidly in recent years due to factors such as population demographic trends, increased incidence of heart disease and other conditions requiring close patient monitoring, along with technological developments such as wireless remote monitoring. Shorter hospital stay and better community care are expected to be the future trend of national health services. Wearable devices represents the next big frontier in the push for the always-on, always-active, and always-connected environment of the future healthcare.

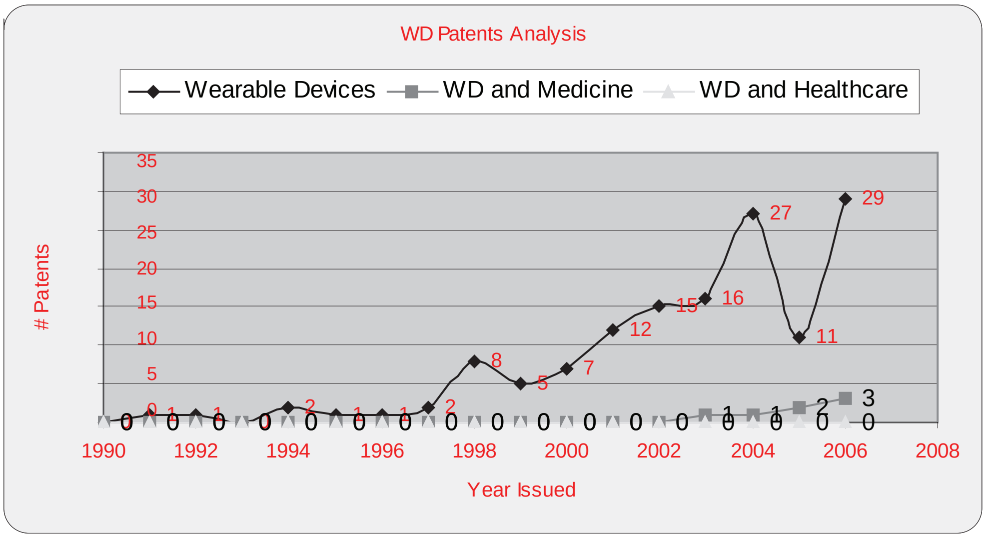
Wearable healthcare monitoring devices aim at a continuous monitoring of vital signs to provide assistance to patients. They are particularly vital to the world's growing aging population and their potential clinical value, particularly for chronic disease management, is increasingly being recognized. Advances in sensor networking and communication technology form the fundamental basis for a new generation of healthcare assistance systems. Patient monitoring products measure, display and document physiological information obtained at regular intervals over time from sensors attached to the patient or other input devices.

Several issues play a critical role in determining the success or diffusion of wearable monitoring devices. In order to make these devices more practical, a series of technical, legal, and sociological barriers need to be overcome. Just as with challenges faced in the context of e-Health, the list of challenges for wearable monitoring devices include: (a) battery life, (b) privacy and security, (c) invisibility and social acceptance, (d) usability, (e) language barriers, and (f) data standards. The inability to overcome these challenges can lead to increased costs, and difficulty in building the needed system-based answers, presenting a significant barrier to entry for new businesses.

The idea of wearable computers dates back to the 1960s, but early attempts to create these types of systems were hampered by the size of the hardware. In order for a computer to be "wearable", it has to be fairly small, lightweight, be able to be attached to clothing, or even be integrated into clothing fibers. A third or more of the 78 million baby boomers and 34 million of their parents may be at risk for the development of devastating diseases including cardiovascular disease, stroke and cancer. Experts predict that pre-symptomatic testing could save millions of lives and dollars in the coming decades (Stemberger, 2002). The implications and potential of these wearable health-monitoring technologies are paramount. These technologies assist healthcare providers in early detection of health deterioration and will alert them in critical situations.

Figure 3 depicts a segment of U.S. patent database to forecast the trend in wearable device technology. Between 1990 and 2005, there

Figure 3. Wearable devices patent analysis



were a total of 173 patents filed in the U.S. The maximum number of patents was filed between 1999 and 2003, totaling 115. The 1990s saw the widespread acquisition and use of laptop computers. Gartner predicts that by 2010, 40% of adults and 75% of teenagers will be utilizing wearable devices, and 70% of the population will spend ten times longer per day interacting with people in the e-world than in the physical world (John McCarthy, 2004).

Apparently, the next decade may be dominated by the production and use of wearable IT systems. Wristwatches, pagers, cell phones, pocket calculators, PDAs, and Blackberries are all examples of simple wearable information systems that are already in use. However, the next generation of wearable systems is still being incubated. While still on the "leeding edge", it may be time for some of the larger, more technologically advanced healthcare institutions to begin collaborating on further research, development, and pilot testing of wearable IT systems in the healthcare setting.

RFID

Approximately between 44,000 and 98,000 people die every year in the U.S. because of medical errors (Institute of Medicine, 1999). In the case where the patient survives due to a medical mishap, it costs hospitals about $47,000 per Adverse Drug Event (ADE). Apart from medical errors, there are people, events, assets and decisions in hospitals that need to be tracked. There are doctors, nurses, patients, visitors and medical equipments, which must be tracked in the event of emergency. Radio Frequency Identification (RFID) technology will facilitate such tracking and help in reducing healthcare costs.

RFID is a technology that uses radio waves to automatically identify people or objects. It has been around since as early as the 1970s, but then it was too expensive and limited to use in commercial applications. RFID has both positive and negative implications for companies as well as consumers. Unlike bar code technology, radio waves can travel through most non-metallic materials. If tags can be produced inexpensively, it can solve most of the problems associated with bar codes. The RFID tags can hold hundreds of characters of information, whereas bar codes can hold only 12 to 15 characters. These tags can be used to trace entire lifecycle of goods, including how and where the items were made, shipped, consumed, and disposed of. RFID technology supports two kinds of implementations: active versus passive. In an active RFID application, signals are continuously transmitted between transponders and transceivers. Passive RFID systems require a reader to be waved near a transponder with an RFID chip.

Implementation of RFID technology in healthcare may be done through sequencing applications in the context of particular hospital settings. The RFID technology requires a semi-conductor chip that emits digital signals that can be picked up by a scanner. Its implementation, therefore, requires hospitals to have basic computer infrastructure that supports an RFID network. Although not a requirement, experts recommend using a wireless network. Indeed, healthcare is one of the top four industries in which RFID is going to have a significant impact: while the bar code has been given a boost through patient safety initiatives and has helped to automate many hospital tasks, it requires a line of sight and manual (physical) scanning in order to register information. In other words, someone must scan the bar code. In contrast, active RFID takes automation one level beyond and decreases the potential for human error because it requires no human interaction and can be read automatically by sensors that may not be seen physically.

The RFID technology is proven but still not fully exploited. This is due partly to the early stage of global software and network standards. As well, tags and readers are still expensive to produce. A real challenge lies in building the hospital-wide infrastructure and overcoming all the attendant back-office hurdles as back-end infrastructure is required to support RFID technology. The integration of everything, from patient and clinicians tracking to medical equipment tracking, is another major task.

Altogether, key barriers to RFID adoption stem from high technology cost to making changes in the healthcare business models. The following factors should be considered in order for RFID technology to mature:

- **Cost:** Cost is a major barrier as RFID tags may be expensive to produce. However, that price is expected to decrease significantly. Cost of RFID tags had dropped from $1 in 2000 to 20 cents in 2004, and by 2006, it is expected to fall to 5 cents. In 2004, readers cost about $1,000 but are expected to fall to $200 by 2006 (3M, 2004). Cost is a major barrier especially in South Asian countries where labor is very cheap compared to the U.S. In these countries, the costs of implementing expensive technology infrastructure far exceed the costs of hiring extra labor.
- **Infrastructure:** While high technology costs is a barrier, the payback periods for RFID have typically been longer than desired. Tagging medical devices and placing readers in doorways at hospitals will require an initial deployment and a change in protocols for locating assets. Lean IT budgets mean that new technologies need to demonstrate compelling business cases and short paybacks on investments. Companies are skeptical if the costs cannot be offset by the promised benefits. Hospitals are therefore hostile to radio frequencies of all kinds. The construction materials and the typical layout of the hospital result in the propagation of RF signals within the structure that is highly attenuate, reflected, and refracted. High infrastructure cost include tag readability, tag price points, system integration costs, hardware and setup costs, management, ownership and sharing of data, business process changes, privacy concerns, and standards.
- **Wireless Network:** Many healthcare organizations are concerned about the wireless network infrastructure. These organizations want to be able to use existing wireless net-

works rather than a dedicated RFID network for tracking.

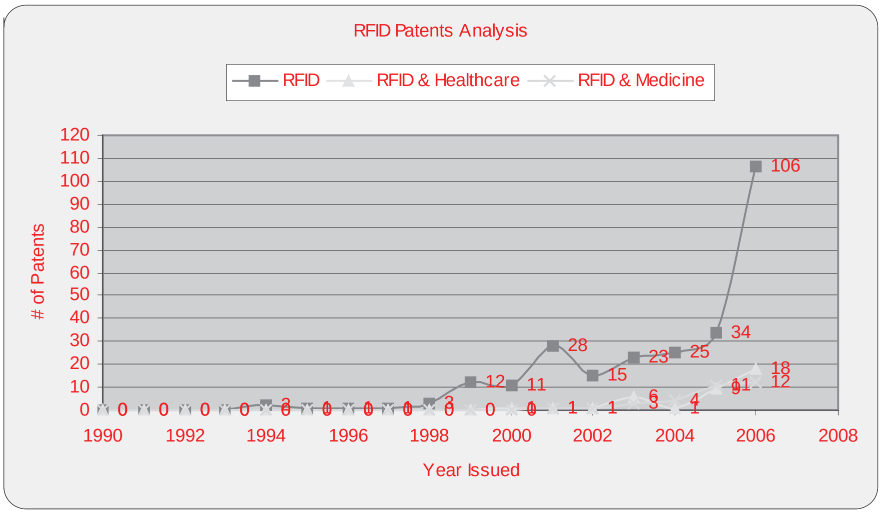
- **Security:** The idea of putting something in your head or in your arm frightens people and stirs up privacy worries. Americans have been warned to protect their personal privacy. Companies should be able to query tags securely while unauthorized parties should not be able to trace them. The possibility that a business loosing control of the privacy of its information is one of the largest risks associated with RFID. The information inside RFID tags is vulnerable to alteration, corruption and deletion.
- **Global Standards:** An additional barrier to RFID technology investment stems from the view that applicable standards are lacking. Currently, there is a tremendous push for consolidation of RFID technology with regards to capability. Yet, with multiple standards and developers seeking to gain the upper hand through their own intellectual property, many businesses are adopting the technology before it is fully ripened, which includes any unresolved issues surrounding security. Some questions include: "What type of number should be used as a unique identifier: the embedded NDC number or a randomly generated number to protect privacy and mask product identity?" and "What common fields/information should be included in an electronic pedigree?"
- **RF Impact:** It is a concern among consumers that radio frequency may have negative impact on drugs. However, based on preliminary information and hypothesis, RF does not have significant thermal impact on solid dosage form. An FDA CDRH lab has done some preliminary research analyzing the heating effect of 915 MHz RFID fields on a placebo, simulating a pharmaceutical. Results showed a very modest rise in temperature, of unknown significance. We are just learning to know the impact of RF on certain products.

A patent is a property right granted to an inventor "to exclude others from making, using, offering for sale, or selling the invention throughout the U.S. or importing the invention into the U.S." for a limited time in exchange for public disclosure of the invention when the patent is granted (www.uspto.gov,). The U.S. Patent office has granted a handful of RFID patents. Today, a growing number of new establishments are applying for RFID patents. Between 1970 and 2003, 1,256 U.S. patents containing the specific term 'RFID' were issued. Of these, 62% were issued in just the last four years. Also, surprisingly, 22% of these U.S. patents went to foreign firms, mostly in Germany and Japan. Even though the number of patents filed and issued for RFID technology is increasing every year, a recent survey by InformationWeek (www.informationweek.com,) shows that technical skill shortage is slowing RFID adoption. Seventy-five percent of business professionals believe there is an insufficient skilled workforce to design and deploy RFID technology strategies, and 80% think adoption rates will suffer.

The U.S. Patents database (www.uspto.gov,) search on "RFID and Healthcare" , as shown in Figure 4, indicates that the number of patents filed in the last ten years is inconsistent. Between 1990 and 2006, there were a total of 262 patents issued, but interestingly, 50% of these patents were issued between 2005 and 2006 alone. This indicates that RFID in healthcare will take another 3-5 years before it matures and replace bar code technology.

In spite of this, the progress of RFID towards maturity in commercial market has been forecasted. As forecasted by IDTechEx, in 2006, IDTechEx expected 1.3 billion tags to be sold, with 500 million RFID smart labels for pallet and case level tagging but the majority falling into a range of diverse markets from baggage and passports to contact-less payment cards and drugs. By 2016,

Figure 4. RFID patent analysis



IDTechEx expects the value of the total market including systems and services to skyrocket to $26.23 billion from $2.71 billion in 2006.

Most health executives believe that RFID technologies are strategic to their business in a number of ways such as patient safety, operational management, and supply chain management. The U.S. healthcare companies are increasingly seeing the adoption of RFID technology as a major potential contributor to safety across the industry. Yet, it is the biggest healthcare players that are generally able to move the quickest and invest in large deployments of the RFID technology.

Artificial Intelligence

Artificial Intelligence (AI) is the study of how computer systems simulate intelligent processes such as learning, reasoning, and understanding symbolic information in context. Human memory is very short and limited. We also have very limited capability to incorporate information in decision making. Our short-term memory can retain and utilize optimally four to seven data items. If we try to utilize more data items or groups, then our decision process may not be very effective (www.ai-depot.com). AI helps in overcoming this limitation and its use combined with informatics can contribute to elevate patient care in many ways. For example, computer-based robotics can help to discover ways to reduce cost, optimize clinical effectiveness, and improve patient care.

Humans have always speculated about the nature of mind, thought, and language, and searched for discrete representations of their knowledge. Traditional computer software is based on adding and multiplying binary number. AI, in healthcare, is an architecture that is designed more closely to a biologic neural system. The basic processing unit of an AI system is a series of neuron-like registers that sum inhibitory and excitatory stimulus, which in turn excites or inhibits a second set of sensors. The strength of AI is its ability to learn by example; this means that programmed procedures may not be necessary and rules do not have to develop by clinical experts. The AI systems in healthcare are intended to support health providers in the normal course of their duties, assisting with tasks that require manipulation of knowledge and data.

The field of AI in medicine has been slow to make its mark and the adoption of AI in healthcare involve overcoming the following barriers:

- **Ethical Challenge:** The efficient and effective use of computers in healthcare requires training, experience and education. Inadequate education and training in the use of computer systems and decision support tools will only result in disaster. Trying to determine who should be allowed to use HIT application is another ethical challenge. Major ethical issues related to AI in healthcare include confidentiality of electronically stored patient information; appropriate selection and use of informatics tools in clinical settings; the determination of who should be assigned to use such tools; the role of system evaluation; the obligation and reliability of system developers, maintainers and vendors; and the use of computers to track clinical outcomes to guide future practice and informatics engenders many important legal and regulatory questions.
- **Physician Resistance:** In medicine, a patient's life may be at stake. Inappropriate use of computer devices or decision influenced by decision support systems may result in an uncontrollable situation. To diagnosis a patient or render a prognosis requires more than a statistical operation performed on raw data. Understanding complex issues and how these issues relate to the patient are not elements that can easily be programmed or computerized. The key problems when using clinical AI to determine policy or aid in patient care are: (a) human cognition is still superior to machine intelligence; and (b) decisions about whether to treat a given patient are often value-laden and must be made relative to treatment goals.
- **Consciousness:** Computers follow algorithms in their thought process, much like the human minds. The difference is that humans can adjust their algorithms to new situations in very complex ways, whereas computers are only beginning to gain the ability to "learn" and adjust their algorithms to new situations. There is no evidence that computers have self-awareness, subjective emotions, abstract thinking, or strategic thinking. Another interesting question is whether these attributes are possible, or if there are real barriers to ever developing AI.
- **Cultural Differences:** Humans have a very wide range of emotions. Can we define each emotion? Is it possible to a non-human machine to differentiate between disappointment and sadness? The intelligence attributes could be a part of emotion, creativity, and spirituality. Although the concept of emotional machine is day-by-day becoming more convincing and compelling, the concept is still very raw.
- **Rule-Based Approach:** Many of the early efforts to apply AI to medical challenges have used a rule-based approach. Such programs are typically easy to create, because their knowledge is cataloged in the form of 'if, then, else' rules in chains of deductions to reach a conclusion. However, most serious clinical problems are so complex and broad that using the rule-based approach will encounter major difficulties. Problems arise principally from the fact that rule-based programs do not embody a model of disease or clinical reasoning. Given the difficulties encountered with rule-based systems, more recent efforts to use AI in medicine have focused on neural networks and programs organized around models of disease.

AI in medicine is created to support healthcare workers in the normal daily duties; they are best at assisting with tasks that rely on the manipulation of data and knowledge. An AI forecasting in healthcare environment describes a future healthcare possibility. It helps industry leaders and consumers to understand where technology will be in the next 5-10 years. Accordingly, we used U.S. patent database to forecast AI technology.

The search against "AI" keyword in the U.S. Patent database resulted in 14,971 hits while that against "AI and Healthcare" keywords resulted in 96 hits and "AI and Medicine" resulted in 652 hits. Figure 5 suggests that AI technology is being incorporated into a growing number of commercial products.

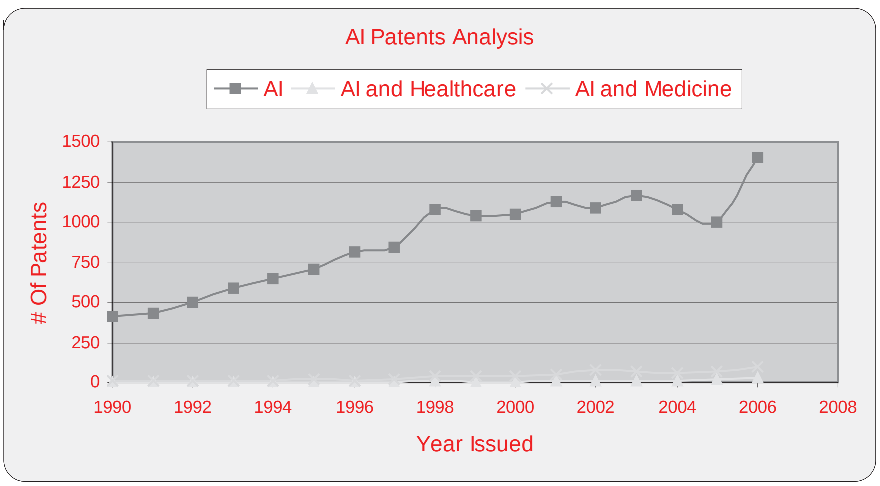
We also searched AI patents filed in other areas. The results indicated that the number of patents in AI, expert systems, and neural networks jumped from fewer than 20 in 1988 to more than 120 in 1996, and the number of patents citing patents in these areas grew from about 140 to almost 800. The number of AI-related patents in the U.S. healthcare system increased exponentially from approximately five in 1990 to more than 15-20 in 2000. In summary, the technological investment in AI and healthcare is enormous and we can expect to see more successful commercial AI products in the near future.

CONCLUSION

Apparently, IT is changing the way we deliver healthcare. New capabilities of IT have the potential to transform health services from primarily reactive to a proactive mode. In a recent press release, for example, a passenger became the first traveler on a commercial airline to have an electrocardiogram (ECG) recorded in flight and transmitted via satellite to physicians on the ground, who correctly diagnosed a heart attack and supervised successful emergency treatment administered by the flight attendants. The convergence of wireless, sensor, wearable and RFID will make scenarios such as this increasingly common.

In spite of rapid technological advances and breakthroughs, the high cost of healthcare in the U.S. prevails. This may be attributed to a number of factors, ranging from the rising cost of medical technology, aging population, to increasing administrative costs resulting from the complex multiprovider and multipayer system in the U.S. healthcare system. The growing shift from non-profit to for-profit healthcare is also contributing to the increasing cost of healthcare. Interestingly, the U.S. is the only developed country in the world, except South Africa, where all of its citizens do not have access to affordable healthcare. According to the most recently available figures, 42.6 million people in the U.S. do not have access to healthcare in 1999. The lack of health insurance

Figure 5. AI patent analysis



for a significant number of Americans has also other far-reaching consequence, as hospitals or providers are forced into cost shifting.

In this article, we discussed five emerging HITs. Our research supports that all the five technologies (RFID, Semantic Web, Wearable Devices, e-Health, and AI) are anticipated to have a future in the healthcare industry. In the U.S. and other developed countries, the healthcare industry is moving from an era of incremental cost control to an era in which the overarching need will be managing the growing demands of an aging population in a creative and affordable way. The state of the art in forecasting and futuristic thinking is in business entrepreneurship rather than in academic research. There is no doubt that these HITs can make a difference in the quality of care. While these technologies can help, impact is limited as long as healthcare remains the poster child for dysfunctional business models. However, these technologies are clearly coming in support of a prevention-oriented, consumer-driven model of healthcare.

The United States has by far the most expensive healthcare system in the world. As discussed, one reason for the high cost in U.S. can be attributed to medical errors. If we deploy emerging HITs intelligently and effectively, it will assist in reducing medical errors and healthcare administration cost. Automation is the key for success as it reduces manual work and errors. Like other healthcare industries, PHS is also trying to reduce administrative cost and medical errors by deploying these emerging HITs. Providence is doing a pilot project in the Oregon region to automate data capture at the point-of-care to reduce paper work and improve efficiency for clinical and nursing staff. Providence is also initiating several projects internally in e-health arena to improve patient-to-provider capability in the Oregon region.

As IT aligns more closely with medicine, new ideas will emerge to challenge entrenched care delivery paradigms. IT technology will pervade clinicians' everyday work lives. Today, automation of redundant and repetitive tasks and the elimination of paper still offer potential payoffs for healthcare organizations. Collaboration around direct electronic data interchange transactions between payers and providers, eliminating manual steps in the business office and in denial analysis, and facilitating electronic clinical documentation are good examples. In the future, the pace of combined medical IT innovation will quicken and be nonlinear. Figure 7 and Table 5 summarized the discussions provided in this article.

Figure 7. Patent analysis summary

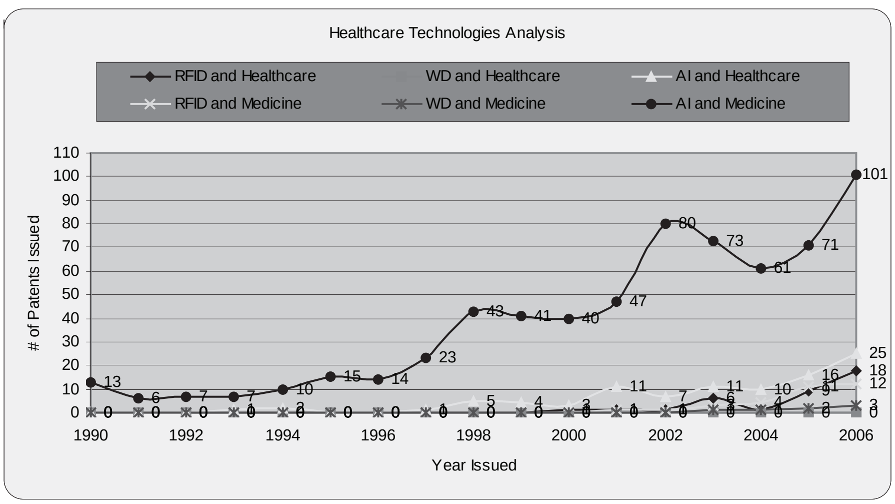


Table 5. Analysis summary

Technology Name	Forecasting Method	Forecast
RFID	Patent Analysis	5-7 Years
Artificial Intelligence	Patent Analysis	5-7 Years
Wearable Devices	Patent Analysis	3-5 Years
Ehealth	Scenario Analysis	3-5 Years
Semantic Web	Scenario Analysis	5-10 years

As seen, our approach for technology forecasting is based on literature search, Internet resources and proven forecasting techniques. In this article, we placed considerable emphasis on searching various literature sources. Typically, each literature reference cites to other references that lead to a constantly expanding database. Patent analyses are indicating that the major R&D efforts are completed for these technologies as the patent curves have started to go down. This indicates that future commercial products are close to being released.

As mentioned previously, many of the technologies we chose embrace a wide range of technologies and market opportunities and, thus, a number of different projections are required for a meaningful forecast of that field. Our next step will be to gather more data for forecasting these technologies. One consideration is to use Delphi Method and incorporate experts' advice in the forecasting process. It would be interesting to incorporate the downfall of existing technologies and emergence of some of these newer technologies. A further step would be to analyze how these newer technologies help in direct patient care, reducing healthcare cost and providing healthcare in remote and rural areas.

REFERENCES

3M Company. (2004). *The true cost of radio frequency identification (RFID).* High Jump Software.

Ammenwerth, E., Iller, C., Mahler, C. (2006). IT-adoption and the interaction of task, technology and individuals: A fit framework and a case study. *BMC Medical Informatics and Decision Making, 6*(3).

Anderson, J., & Kenneth, G. (2002). *Ethics and information technology.* Springer.

Anderson, J. (1997). Clearing the way for physicians' use of clinical information systems. *Communications of the ADM, 40*(8).

Arlington, S., Barnett, S., Hughes, S., & Palo, J. (2003). *Pharma 2010: The threshold of innovation.* IBM Business Consulting Services IBM Futures Series 2003.

Ash, J. (1997). Organizational factors that influence information technology diffusion in academic health sciences centers. *Journal of American Medical Informatics Association, 4,* 102-111.

Ash, J., Berg, M., & Coiera, E. (2004). Some unintended consequences of information technology in healthcare: The nature of patience care information systems-related errors. *Journal of The American Medical Informatics Association, 11*, 104-112.

Berg, M. (2001). Implementing information systems in healthcare organizations: Myths and challenges. *International Journal of Medical Informatics.*

Berners-Lee, T., Fielding, R., & Masinter, L. (1998). Uniform resource identifiers: Generic syntax. Retrieved from http://www.ietf.org/rfc/rfc2396.txt

Bower, A. (2005). *The diffusion and value of healthcare information technology.* Rand Corporation.

Butler, M. (2003). Barriers to real world adoption of semantic web technologies. Hewlett-Packard Labs Technical Report.

Cauffiel, D. & Porter, A. (1996). Electronics manufacturing in 2020: A national technological university management of technology mini-Delphi. *Technological Forecasting and Social Change, 51*, 185-194.

Chiasson, M., Reddy, M., Kaplan, B., & Davidson, E. Expanding multi-disciplinary approaches to healthcare information technologies: What does information systems offer medical informatics?. *International Journal of Medical Informatics, 76*, S89-S97.

Cohan, A. & Hanft, R. (2004). *Technology in American healthcare*.

Connor, S. (2003). Glaxo chief: Our drugs do not work on most patients. *The Independent*. December 8.

Daim, T., Grueda, G., Martin, H. (2006). Forecasting emerging technologies: Use of bibliometrics and patent analysis. *Technological Forecasting and Social Change*, *73*(8), 981-1012.

Del Campo, A., Sparks, A., Hill, R., Keller, R. (1999). The transfer and commercialization of university-developed medical imaging technology: Opportunities and problems. *IEEE Transactions on Engineering Management, 46*(3), 289-298.

Dewitt, J. & Hampton, P. (2005). Development of a data warehouse at an academic health system: Knowing a place for the first time. *Academic Medicine*, *280*(11), 1019-1025.

Dixon, D. (1999). The behavioral side of information technology. *International Journal of Medical Informatics, 56* (1), 117-123.

Ein-Dor, P. & Segev, E. (1978). Organizational context and the success of management information systems. *Management Science, 24*(10), 1064-1077.

Geisler, E., Krabbendam, K., & Schuring, R. (2003). *Technology, healthcare, and management in the hospital of the future*.

Gelijns, A. (1992). *Technology and healthcare in an era of limits.*

Ginzberg, M. (1981). Key recurrent issues in the MIS implementation process. *MIS Quarterly, 5*(2), 47-60.

Glaser, J. (2002). *The strategic application of information technology in healthcare organizations.* Jossey Bass.

Greenhalgh, T, Robert, G., Bate, P., Macfarlane, F., & Kyriakidou, O. (2005). *Diffusion of innovations in health service organizations: A systematic literature review*. Blackwell BMJ Books.

Haux, R. (2006). Individualization, globalization and health--about sustainable information technologies and the aim of medical informatics. *International Journal of Medical Informatics, 75*(12), 795-808.

IBM Business Consulting Services. (2004). *Pharma 2010: Silicon Valley.*

Institute of Medicine. (1999). *To err is human: Building a safer health system*. Washington, D.C.: National Academy Press.

Kaplan, B. (2001). Evaluating informatics applications- Some alternative approaches: Theory, social interactionism, & call for methodological pluralism. *International Journal of Medical Informatics*, *64*(1), 39-56.

Karsten, H. & Laine, A. User interpretations of future information system use: A snapshot with technological frames. *International Journal of Medical Informatics, 76*, 136-S140.

Kayal, A. (1990). Measuring the pace of technological process: Implication for technological forecasting. *Technological Forecasting and Social Change, 60*, 237-245.

Kumar, A., & Motwani, J. (1999). Management of healthcare technology literature (1979-1997): A multidimensional introspection. *IEEE Trans-*

actions on Engineering Management, 46(3), 247-264.

Larsen, K. (2003). A taxonomy of antecedents of information systems success: Variable analysis studies. *Journal of Management Information Systems, 2*(2), 169-246.

Linstone. (1999). *Decision making for technology executives: Using multiple perspectives to improved performance*. Boston: Artech House.

Lorenzi, N. & Riley, R. (2000). Managing change: An overview. *Journal of the American Medical Informatics Association, 7*(2), 116-124.

Martino, J. (1983). *Technological forecasting for decision making*, 2nd ed. New York: North-Holland.

McCarthy, J. (2004). *What is artificial intelligence*. Computer Science Department Stanford University.

Meijden Van Der, J., Tange, H., Troost, J., & Hasman, A. (2003). Determinants of success of inpatient clinical information systems: A literature review. *Journal of The American Medical Informatics Association,10*, 235-243.

Nowinski, C., Becker, S., Reynolds, K., Beaumont, J., Caprini, C., Hahn E., et al. The impact of converting to an electronic health record on organizational culture and quality improvement. *International Journal of Medical Informatics, 76*, 174-183.

OECD Health Technology and Decision Making. (2005). OECD Health Project.

Paul, D., Pearlson, K., & Mcdaniel, R., Jr. (1999). Assessing technological barriers to telemedicine: Technology-management implications. *IEEE Transactions on Engineering Management, 46*(3), 279-288.

Poon, E., Jha, A., Christino, M., Honour, M., Fernandopulle, R., Middletoon, B., et al. (2006). Assessing the level of healthcare information technology adoption in the United States: A snapshot. *BMC Medical Informatics and Decision Making, 6*(1).

Sager, B. (2003). Scenarios on the future of biotechnology. *Technological Forecasting and Social Change, 70*, 109-129.

Sheng, O., Hu, P., Chih-Ping, W., & Pai-Chun, M. (1999). Organizational management of telemedicine technology: Conquering time and space boundaries in healthcare services. *IEEE Transactions on Engineering Management, 46*(3), 265-278.

Silberglitt, R., Hove, A., & Shulman, P. (2003). Analysis of U.S. energy scenarios: Meta-scenarios, pathways, and policy implications. *Technological Forecasting and Social Change, 70*, 297-315.

Stemberger, S. (2002). *Wearable wireless devices*. Retrieved from http://www.ibm.com/developerworks/wireless/library/wi-wear.html

Tan, J. (2005). *E-healthcare information systems*. Jossey Bass.

Tan, J., & Sheps, S. (1997). *Health decision support systems*.Jones & Bartlett Publishers.

Watts, R. &. Porter, A. (1997). Innovation forecasting. *Technological Forecasting and Social Change, 56*, 25-47.

Wilson, A., Ramamurthy, K., & Nystrom, P. (1999). A multi-attribute measure for innovation adoption: The context of imaging technology. *IEEE Transactions on Engineering Management, 46*(3), 311-321.

Winebrake, J. & Creswick, B. (2003). The future of hydrogen fueling systems for transportation: An application of perspective-based scenario analysis using the analytic hierarchy process. *Technological Forecasting and Social Change*, *70*, 359-384.

Zheng, K, Padman, R., Johnson, M., & Diamond, H. (2005). Understanding technology adoption in clinical care: Clinician adoption behavior of a point-of-care reminder system. *International Journal of Medical Informatics.*

This work was previously published in the International Journal of Healthcare Information Systems and Informatics, Vol. 3, Issue 1, edited by J. Tan, pp. 1-23, copyright 2008 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 8.7
Towards Knowledge Intensive Inter-Organizational Systems in Healthcare

Teemu Paavola
LifeIT Plc, Finland and Helsinki University of Technology, Finland

Pekka Turunen
University of Kuopio, Finland

Jari Vuori
University of Kuopio, Finland

ABSTRACT

The aim of this chapter is to share recent findings and understanding on how information systems can be better adopted to support new ways of work and improve productivity in public funded healthcare. The limits of transferring explicit and tacit knowledge are discussed and moreover, the chapter elaborates barriers to the widespread use of knowledge management tools among clinicians. The impacts of an inter-organizational system used for remote consultation between secondary and primary care providers are examined. Furthermore, the authors suggest that issues related to clinical knowledge management such as the varying information and knowledge processing needs of clinicians from various medical expertise domains should be examined carefully when developing new clinical information systems.

INTRODUCTION

It is reasonable to expect information technology to bring benefits to healthcare organizations just as it does to any other business. But recent studies demonstrate that the introduction of information technology does not in itself improve employee effectiveness in healthcare (Littlejohns, Wyatt, & Garvican, 2003). Moreover, it has proved particularly difficult to evaluate information systems in healthcare. This may involve the "evaluation paradox": we refuse to use a new technology until an evaluation study of its use has proved it useful. Particularly true in the healthcare sector, this cautious approach can be seen as a virtue, considering that hospitals fortunately were never taken in by all the information technology hype; but on the other hand the healthcare sector is regrettably slow in adopting even the best new practices. Although successful adoption of inter-organizational systems in healthcare still lacks a substantial body of research we argue that a key issue to be addressed may well be the natural logic of information and knowledge processing in various medical areas.

The term "telemedicine" covers the application of information technology to medical care chains. Conventional operations can always be pushed up a notch with state-of-the-art technology. In many cases it would also be more efficient to reorganize the operations altogether. For example, an e-mail application is at its most useful when its users are considered as active users of information rather than passive recipients. An organization can learn new ways of profiting from this new communications channel, which lies at the medium level according to the information richness theory introduced by Daft and Lengel (1986). A good example may be found in the healthcare sector, where new technology has managed to loosen the shackles of conventional thinking.

The electronic consultation model in medical care emerged in the early 1990s, beginning with an experiment at the internal medicine department of a hospital in southern Finland, where referrals were returned with care instructions in cases where lab results coupled with the information on the referral were sufficient for an accurate diagnosis. If the specialist receiving the referral considered that the case had not been examined sufficiently, the referral would be returned with instructions as to further tests or examinations. The new model evolved naturally from this practice. The information entered on a referral is often sufficiently comprehensive to enable remote consultation, and under the new model about half the patients referred to the hospital could be treated at their local health center on the basis of instructions provided by specialists through an information network.

There are many phenomena at play in shaping the practices of the healthcare sector. Identifying and allowing for these phenomena may be the key to successful telemedicine projects, indeed even more important than the technology itself. In the case of the new model described above, the innovation was supported by studies in the field: over half of all referrals in Finland and, for example, in Britain contain enough information for making an accurate diagnosis. So why are these patients being sent to hospital if there is already enough information for a diagnosis to be made at the health center? Convention, insecurity, financial incentives and similar factors have been cited as possible reasons.

This chapter aims to share recent findings and understanding on how information systems can be better adopted to support new ways of work and improve productivity in public funded healthcare. The effects of an integrated electronic referral system used for remote consultation between secondary and primary care providers is examined in a case study of two healthcare units in southern Finland. The study demonstrates how costly investments in videoconferencing in orthopedics yielded lesser benefits than the cheaper investment in e-mail-type application in internal medicine. Evidently internal medicine relies on fixed-format information, whereas orthopedics is more dependent on direct sensory inputs and tacit knowledge (Sternberg & Hovarth, 1999). Consequently, the natural logic of information and knowledge processing needs to be examined carefully before investing in information technology.

DEVELOPMENT TOWARD KNOWLEDGE INTENSIVE SYSTEMS IN HEALTHCARE

The information system presented in this chapter is representing information systems of new generation. The amount and meaning of these knowledge intensive systems will be highly expanding in the field of healthcare in the future. The progress of information system presented here is concluded based on theoretical frameworks of information systems in different disciplines. Phases of information system development in organizations have been dived from two to five phases by information system science researches. Most explanatory model seems to be Friedman and Cornford's (1989) intensive study (Checkland & Holwell, 1998; Korpela, 1994).

Friedman and Cornford argue that information system development can be seen as the interaction of changes in the core technology with changes in applications, linked by activities which mediate between them. These three elements are acting upon agents of change. The agents of change are technical development, changes in applications, market pressures and internal pressures, such as management culture (Friedman & Cornford, 1989).

After given this analysis, researchers see the history of development of the information systems in terms of three phases, each defined by a "critical factor" that has limited the development of computerization during that period. These phases are (1) hardware, (2) software and (3) user relation constraints. According to Friemdan and Cornford (1989) the first phase was approximately until the mid-1960s, the second one lasted from the mid-1960s until the early 1980s and the third one was early in the 1980s at least until their research was published in 1989.

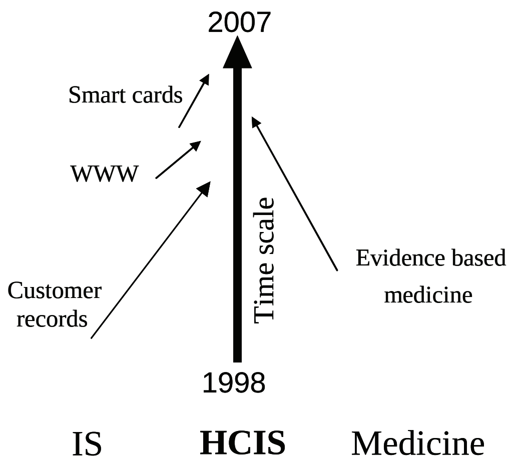
In the year 1989, Friedman and Cornford speculated that a fourth phase might be organization environment constrain. The fourth phase could be also divided into sub-phases. At the first issue of developing systems is to make large-scale organizational systems and after these problems have been solved the main concern is to build inter-organizational systems. These inter-organizational systems are often customer orientated information systems (Friedman & Cornford, 1989).

Not every organization is going causally through these four phases. The development of information systems in the organization depends also on time when computerization has started in the particular organization or on other issues such as size of organization and ability to change organization (Checkland & Holwell, 1998; Friedman & Cornford, 1989).

At the moment, the state of healthcare information systems resembles that found in other fields in the early 1990s. As a rule, the field seems to be 10 years behind other areas (Ragupathi, 1997). However, the development of healthcare information systems is not exactly linear. Some innovations are put into use in healthcare faster than others (that are not seen very useful in healthcare). For example, Intranet technologies have been applied fast in healthcare organizations at least for non-medical purposes. Also (medical) knowledge intensive solutions are implemented fast in the field of healthcare (see Figure 1) (Turunen, Forsström, & Tähkäpää, 1999). The use of knowledge intensive systems have seen already in other fields, thus this same development is expected to happen also in healthcare.

General development of information systems in the 1990s as well Friedman's genuine work is clearly concluding that healthcare information systems will develop into the direction of customer orientated systems. This progress is in some cases passing through organizational environment constraints. It is possible that organizational phase is skipped over, if the particular organization is innovative, small or organic ones (cf. Courtright, Fairhurst, & Rogers, 1989). The more organic organization is relation to its environment the more capable it is to react the demands of customers and adapt the new technology. Naturally, the

Figure 1. An example of diffusion time of IT innovations and medicine into the field of health care information systems (HCIS)



rapid development is also helped by new Internet technology.

Because of information intensive medical work (see right side of Figure 1) these new applications are in the healthcare strongly knowledge intensive ones. As a matter of fact, medicine can be considered as a knowledge-based business, which in practice means that experienced doctors use about two million pieces of information to manage their patients (Pauker, Gorry, Kassirer, & Schwartz, 1976; Smith, 1996; Wyatt, 1991). About a third of doctors' time is spent recording and combining information and a third of the costs of a healthcare provider are spent on personal and professional communication (Hersch & Lunin, 1995). Still most of the information doctors use when seeing patients is kept unrecorded in their heads and unfortunately some of this information is out of date or wrong. Thus the interest in applying networking media, like Internet in medical practice has arisen from the ever greater demand of meeting the needs of patients by drawing on the knowledge accumulated by medicine over 5,000 years.

New information may not have penetrated and the information may not be there to deal with patients with uncommon problems (Smith, 1996). These problems have become more serious as the rate of change in medical knowledge has accelerated. There are new scientific findings every day and at the moment the medical knowledge is estimated to increase fourfold during a professional lifetime (Heathfield & Louw, 1999; Wyatt, 1991). This inevitably means that doctors cannot practice high quality medicine without constantly updating their knowledge and finding information to help them with particular patients. The historical arrangement, in which doctors individually held the responsibility for ensuring an adequate supply of knowledge to guide their practice, is now failing to meet the knowledge needs of modern healthcare (Smith, 1996; Tannenbaum, 1994). Therefore diverse organizations, also health authorities, spread a vast amount of information to doctors.

It seems that knowledge intensive inter-organizational systems are the latest result of evolution in the field of healthcare. Healthcare organizations will adopt these kinds of systems sooner or later. Thus, it is important to understand the use of these knowledge intensive inter-organizational systems.

MANAGING EXPLICIT KNOWEDGE: TOWARDS OPTIMAL ALLOCATION OF SUPPLY AND DEMAND

The demand for healthcare services in both quantitative and qualitative terms is growing constantly. The aging of population will cause more pressure on healthcare organizations in several countries. The demand for healthcare services is greater, almost double among the old when comparing to average adults (Bethea & Balazs, 1997). Managing supply and demand between the primary parties, doctors and patients, seems to be a simple constellation of resources and requests. However, when managing supply and demand within the public funded healthcare system, the constellation changes into allocation of information resources of varying proficiencies and cost structures. By tradition, the tool used for this allocation has been the referring process.

The traditional referring process has its roots in the late 19th century England. There were three medical proficiencies operating in hospitals: surgeons, doctors and pharmacists. The third group was later called general practitioners, and before the turn of century they were pressured to leave hospital premises. As a result, citizens had no longer free passage to hospital care without seeing a general practitioner first.

REENGINEERING THE ONE-WAY INFORMATION PROCESS: A CASE STUDY

In the following example we concentrate to analyze the factors behind the startling change that took place when paper referrals were replaced with electronic referrals. Our example is based on a study that took place in southern Finland focusing on cross-functional processes within two healthcare units (Harno, Paavola, Carlson, & Viikinkoski, 2000). The process chosen for the study was simple: a health center general practitioner referring a patient to a specialist at a regional hospital. While one unit began to utilize information system when referring a patient to hospital, the other send referrals traditionally by ordinary mail. Parts of the study are reported in TQM and Human Factors conference, Sweden (Lillrank, Paavola, Harno, & Holopainen, 1999).

Both healthcare units, Peijas Hospital and Hyvinkää Hospital, consist of a hospital and several health centers serving 150,000 people. Both healthcare units use similar internal administrative processes, however, Peijas has an Intranet-type system connecting the hospital with health centers while the information process between the parties in unit Hyvinkää is done traditionally with paper documents. All internal medicine cases that were referred from health centers to the hospitals during the study are included.

The process of making diagnosis goes in short according to following. First, in health center a general practitioner begins to diagnose by taking the anamnesis, that is, a preliminary case history of a medical or psychiatric patient. Patient explains the matter in his or her own words and general practitioner uses interviewing technique to lead the conversation in order to confirm or dispatch any presentiments of diagnosis that well forth in mind. Based on the anamnesis and verbal communication general practitioner can also decide to take laboratory tests in order to make the diagnosis more accurate. The diagnosis can be then defined and the treatment process can begin.

In roughly 95 percent of visits, the health center can handle the whole case. The remaining five percent of cases require specialist's resources for diagnosis, tests, treatment or all of these. In such cases a referral is sent to a hospital. The behavior behind the sending of a referral can be explained by several factors that may all have an effect to tip the scale in favor of referring (Coulter, Noone & Goldacre, 1989).

At the hospital, after examining an incoming referral, the specialist decides its level of urgency, again in one out of the three categories. Within the given time period the patient will be invited to hospital for consulting. The other option would be to send the referral back to the general practitioner if the specialist decides the case does not warrant specialist care. In practice, however, almost all referrals lead to a hospital visit. During our study the specialists working with traditional paper documents estimated that one out of five cases could have been treated at a health center with the help of some consultation. However, they felt that inviting a patient in for consulting that takes normally 20 minutes was less troublesome and took less time than writing and mailing instructions back to the general practitioner. Further, there were no economic incentives to take this trouble, since the local governments paid hospitals by the number of patients that were actually called in. In sum, the traditional process had neither procedures nor incentives for optimal resource allocation.

The Peijas hospital was built in the early 1990s. The health centers in the corresponding healthcare unit Peijas were linked to the new hospital with a tailor-made Intranet-type system. Initially it was assumed that the system would simply replace paper forms and mail, and allow electronic filing. The effects were assumed to be speedier traffic of information. There had been some plans to use the new information route also to integrate secondary and primary care providers with an electronic referral system linked to an electronic medical record. However, it was not known what the new process would look like, or even less, if the procedures were going to work at all. The precautions were justified since the effects of an integrated electronic referral system used for remote consultations had not been studied.

For a few years the system was utilized as such, a speedway for one-way referral traffic giving the predictable, but rather minor impact on the performance of the whole system. Only after a few specialists in the internal medicine area started, on their own initiative, to utilize the system for remote consultation with promising results, the other specialists in the same area joined in. Eventually the local governments had to revise their economic system to hospitals to include remote consultations as well as actual visits as basis for remuneration. Thus, a new business process supported by remote consulting had emerged in internal medicine.

The basic assumption when the integrated referral system was installed was that some 20 percent of referral cases could be treated at a health center with the help of remote consultation. In practice, the amount was more than double. The ratio of referrals that were returned to the health center grew to 50 percent of all referrals.

The consequences from this were, first, that the number of patients who were asked to actually visit a specialist was reduced to half thus reducing the burden on specialist resources. According to the doctors involved, this happened in 75 percent of the cases without any observable problems to fulfill of the appropriate treatment. Second, it was discovered that the rate of remote consultations given by specialist varied depending on referral urgency. Only 10 percent of the most urgent, priority I referrals lead to the use of remote consultation, while the percentage was 30 for category II, and over 50 for the category III referrals. Consequently, the major cycle time reduction came to the least urgent cases. Third, the number of health center patients whose problems got the attention of a specialist through remote consultation, roughly doubled that is from five to ten patient visits out of hundred. This was due to the learning experience to utilize cooperation with specialists in borderline cases. As a fourth effect it can be added the hard to measure effects that frequent exchange of experiences had on learning on health center side.

Encouraged by the positive results achieved in internal medicine, the possibility of expanding the system into other disciplines was discussed at Peijas hospital. Orthopedic surgery was used as a comparative case, since its business process is in many ways different. Internal medicine is, by and large, an intellectual process where specialists make conclusions from pieces of information of symptoms, medical histories and laboratory tests. The information used is reasonably well structured and treatment is often medication. Orthopedic specialist's were not inclined to use e-mail consultations, therefore a videoconferencing system was installed allowing the specialist's to study x-ray images and the patient located at health center while interacting with the general practitioner. In half of the cases videoconferencing was found to work satisfactorily, however, it did not reduce the number of patients that were eventually sent to the hospital. Thus, the relatively cheap e-mail system in internal medicine yielded a better return on investment than the costly videoconferencing system in orthopedic surgery.

MANAGING IMPLICIT KNOWLEDGE IN MEDICINE: CHALLENGE FOR SHARED KNOWLEDGE

The limits of transferring knowledge as well as managing it only the basis of rational view of man has proved to be inconsistent with the reality of medical experts in many ways (Ferlie, Wood, & Fitzgerald, 1999; Patel, Arocha, & Kaufman, 1999). Medical professionals produce a great amount of knowledge, which a doctor has but is unable to articulate or quantify easily. In frame of transferring technology managing knowledge refers more explicit than implicit knowledge. To be more precise, most of medical experts transfer more data than practical know-how to each other. The problems with sharing knowledge among a specific districts of medicine rises from the fact that every each of doctor attach a different meanings to data on the basis of their experiences of previous patients. However, studies of social construction, learning and tacit knowledge gives us a opportunity to approach the dilemma in promising way (Argyris, 1999; Berger & Luckmann, 1966; Dixon, 1994; Koskinen, 2001; Polanyi, 1966; Polanyi & Prosch, 1975; Sternberg & Horvarth, 1999).

It is commonly accepted scholars that are two basic types of knowledge: explicit knowledge that can be verbalized, such as knowledge of facts and concepts; and implicit knowledge that cannot be made verbal, such as intuition and knowledge of procedures (Patel et al., 1999). Even if the conventional synonym of implicit knowledge is tacit knowledge, some of scholars use the concept of tacit meaning structures as well (cf. Dixon, 1994). Tacit knowledge, by definition, refers to the inarticulate aspects or meaning structures that cannot be taught explicitly and therefore are only learnt via direct experience (Dixon, 1994; Patel et al., 1999). According to Polanyi and Prosch (1975) textbooks of medicine are so much empty talk in the absence of personal, that tacit knowledge of their subject matter. Therefore evaluating explicit knowledge (e.g., education) is pretty easy to do, but to evaluate its impact on tacit meaning structures of orthopedic specialist's is almost impossible.

In management literature the meaning of intuition for leaders has not been studied very much even if the importance of it for management was found already in 1930's (Barnard, 1938). However, for example Patel et al. (1999) outlined a systematic approach to the study of medical expertise in order to understand the role that the acquisition of tacit knowledge plays in competent and expert performance. According to them intensive care decision-making is characterized by a rapid, serial evaluation of options leafing to immediate action. In this real-time decision-making, the reasoning is schema-driven in a forward direction toward action with minimal interference or justification (ibid.). Practically this means that doctors are tend to attach meaning to data more on the basis of their experience than logical interpretation of facts. On the other hand we may say that the scientific knowledge is in part socially constructed in different ways in a case of general practitioner who is sending data compared with medical expertise who is interpreting it (cf. Ferlie et al., 1999). Therefore managing knowledge should be always understood in the terms of socially construction reality among a certain group of professionals.

The importance of tacit knowledge in decision-making is proved to be more important than explicit knowledge (Leprohon & Patel, 1995). In one study (ibid.) decision-making accuracy was significantly higher in nurses with 10 years or more of experience than in nurses with less experience, which is consistent with that we know about acquisition of expertise in their domains.

Managing the tacit knowledge may easily sound the answer for the problems of electronic referring system. However, the tacit knowledge is also the issue of cooperation between medical professionals. To some extent there will be always deliberate amount of 'tacit knowledge' in healthcare organizations. Clinical professionals will retain a monopoly of knowledge as far as professionals are forced to compete each other in terms of budget and resources (cf. Ferlie et al., 1999).

Barriers to open communication between different medical experts is understandable when the goal of medicine is not integrated (cf. primary care vs. special care, science vs. clinical care in university hospitals). Findings of Smith and Preston (1996) are pretty common: senior management have problems regarding the interface between junior doctors and other professional groups. An elite of professional workers secures easily a high degree of autonomy for itself on the basis of possession of expertise (Ferlie et al., 1999). However, tacit knowledge is said to be the primary basis for effective management, and the basis for its deterioration (Argyris, 1999; Nonaka & Takeuchi, 1995). Understanding the management of knowledge in the frame of rational man leads us easily situation where the benefits of electronic referring system are limited (cf. Breite, Koskinen, Pihlanto, & Vanharanta, 1999).

KNOWLEDGE MANAGEMENT ISSUES FOR CLINICAL APPLICATIONS

When utilizing applications of information technology in product-line orientated organization aspects such as the level of utilization becomes important. For example if only every third salesman in a company types his or hers orders in a given database the benefits of using an enterprise resource planning system barely exists at the shop floor. In knowledge intensive organizations, on the other hand, the quality and usefulness of the information content seems to play a leading role.

Information collected and stored in product-line orientated organizations is normally fixed-format information, for example, ordering codes that typically consist of numbers and letter acronyms. When storing qualitative data in knowledge intensive organizations the case becomes more complex. Storing personal notes is a trivial case, since one understands all the matters that are relevant and connected with the stored information. On the other hand storing qualitative information for other people's later interpretation is a greater challenge. One must presume the form that the data will be later most useful in; that is, one must understand the context the data will be later utilized in. In order to fulfill the validity and relevance requirements of users the rationale behind the information must be somehow included in the stored data. Until we learn to do that we must grapple with the quest of various oracles and gurus to explain us the true nature of knowledge management.

Slawson and Shaughnessy (1997) determine the usefulness of medical information according to the following equation (Figure 2). Since in practice the fixed consulting hours of doctors set limits to the dispersion or flexibility of the work amount, we can argue that the more validate and relevant information the practitioners get within constant time to access it the better.

A cautious approach to introduce knowledge management issues in clinical applications can be seen as a virtue. Sackett et al. (1997) argue that still 80 per cent of all treatments are not based on scientific proof but on believes. Validity can be determined as the probability of the information being true. For example the fact that an article is published in an academic journal does not guarantee that the information is true. One such article presented findings from a research that demonstrated how sodium fluoride effectively prevents contracting osteoporosis. Later another

Figure 2. Usefulness of medical information

$$\text{USEFULNESS OF MEDICAL INFORMATION} = \frac{\text{RELEVANCE} \times \text{VALIDITY}}{\text{WORK}}$$

more validate research findings proved that there truly was causality between these two—only the effect sodium fluoride had on the illness was just the opposite (Slawson & Shaughnessy, 1997).

One barrier to the widespread use of knowledge management tools is the psychological need of clinicians. Weinberg, Ullian, Richards, and Cooper (1981) found that 81 American doctors in a distant geographical area consulted 23 experts, logging 11 calls to experts each month. Six of these experts received over 90 percent of the calls, about 90 each a month. Doctors consult experts, because it is a quick, cheap, and easy method. The other doctors can also provide the psychological benefits that are not available from books, journals, and computers. On the other hand, the doctors "answering" the questions may sometimes not be much more knowledgeable than the doctors seeking answers.

Also a review article by Smith in 1996 which looked at the information needs of general practitioners made several similar conclusions, one of which was that doctors are looking for guidance, psychological support, affirmation, commiseration, sympathy, judgement, and feedback. Smith (1996) argues that this aspect of doctors' information needs is particularly poorly explored, and yet it may well be the most important need and the biggest stumbling block to a technical solution. The electronic consultation and knowledge management tools, such as decision-support systems, could support each other in various ways. Doctors could use the knowledge system for retrieving required information before referring to the specialist. They could for example check what are the necessary laboratory tests that should be done before referring. This would save specialists' time in electronic consultation as wouldn't need to separately ask for those tests. In addition that would be convenient for the patient.

CONCLUSION

Within the healthcare sector, different types of process can be identified with different demands on the quality of information and knowledge required. The amount of reference information needed to make a diagnosis differs from one specialization to the next in a hospital. In some fields, the information required by a doctor consists mainly of basic data, observations recorded in numbers and text, whereas in other fields an image or video clip is considered crucial for drawing conclusions. While the significance of what is known as tacit information is important in some fields, in the majority of cases it is explicit information that is required; and it is in care chains relying heavily on explicit information that the potential of information technology can be put to the best use.

We examined the impacts on adoption of a new inter-organizational system in healthcare. The consequences from using the Intranet-type referral system in the example case were, first, that the number of patients who were asked to actually visit a specialist was reduced to half thus reducing the burden on specialist resources. This was due to the replying option in the system. According to the doctors involved, this happened in 75 percent of the cases without any observable problems to fulfill the appropriate treatment. Secondly, the rate of consultations was varied depending on referral urgency. The major time reduction came to the least several cases. Thirdly, 5 out of 100 general practitioners' patients normally treated also by specialists was doubled to 10. However, there were differences between the benefits in different medical areas according to nature of information related to education of different specialists. This should be considered more carefully in the future, when new information systems will be planned and developed.

Evidently in internal medicine, specialists make conclusions from pieces of information of symptoms, medical histories and laboratory tests. The information used is explicit, reasonably well structured and treatment is often medication. Intranet referral system both increased the utilization of specialist consultation and decreased the need for secondary care services by transferring information and knowledge to primary care. The system allowed more patients to be treated at lower expense. Because all patients were thoroughly examined beforehand, the numbers of repeat visits as well as direct costs remained lower.

In orthopedic surgery the situation appeared to be different, as the diagnostic process is based more on tacit knowledge and the specialist frequently rely on direct observations. In orthopedics the specialists were not inclined to use e-mail consultations, therefore a videoconferencing system was installed allowing the specialists to study x-ray images and the patient located at health center while interacting with the general practitioner. In half of the cases videoconferencing was found to work satisfactorily, however, it did not reduce the number of patients that were eventually sent to the hospital. The relatively cheap e-mail solution in internal medicine yielded a better return on investment than the costly videoconferencing system in orthopedics. Consequently, the natural logic of information and knowledge processing in various areas needs to be examined carefully before investing in new information technology solutions.

Although it has been said that it is possible to ease sharing of tacit knowledge by using computer environment that is simulating real world (see Breite et al., 1999), it seems to be expensive task in healthcare. To optimize relatively small resources it should be invested more for developing sharing knowledge systems in those medical areas (like in internal medicine) that are using explicit information. In those medical areas that are using more like implicit information (like in surgery) development knowledge sharing systems should be delayed until we have effective mass-products that are supporting multimedia features. However, even then the development of those systems that are supporting open communication between different specialists and more generally between different professional groups should be encouraged. This would ease the change of working culture into the direction of non-hierarchical knowledge sharing.

REFERENCES

Argyris, C. (1999). Tacit knowledge and management. In R.J. Sternberg & J.A. Horvath (Eds.), *Tacit knowledge in professional practice: Researcher and practioner perspectives* (pp. 123-140). London: LEA.

Berger, P. & Luckmann, T. (1966). *The social construction of reality.* New York: Penguin.

Breite, R., Koskinen, K., Pihlanto, P. & Vanharanta, H. (1999). To what extent does the tacit knowledge embodied in a technology product limit its electronic commerce potential? *Proceedings of IRIS22,* August 7-10 , Keuruu, Finland.

Buchan, I., Heathfield, H., Kennedy, T. & Bundred, P. (1996). Decision support for primary care using the Path.Finder System. *British Journal of Healthcare Computing, 13*(6), 20-21.

Checkland, P. & Holwell, S. (1998). *Information, systems and information systems.* Chichester: John Wiley & Sons.

Coulter, A., Noone, A. & Goldacre, M. (1989). General practitioners' referrals to specialist outpatient clinics I. Why general practitioners refers patients to specialist outpatient clinics. *British Medical Journal, 299*(6694), 304-306.

Courtright, J.A., Fairhurst, G.T. & Rogers, L.E. (1989). Interaction patterns in organic and mechanistic systems. *Academy of Management Journal, 32*(4), 773-802.

Daft, R. & Lengel, R. (1986). Organizational information requirements, media richness and structural design. *Management Science, 32*(5), 554-571.

Dixon, N. (1994). *The organizational learning cycle. How we can learn collectively.* London: McGraw-Hill.

Duff, L.A. & Casey, A. (1999). Using informatics to help implement clinical guidelines. *Health Informatics Journal, 5*(2), 90-97.

Ferlie, E., Wood, M. & Fitzgerald, L. (1999). Some limits to evidence-based medicine: A case study from elective orthopaedics. *Quality in Health Care, 8*(2), 99-107.

Friedman, A.L. & Cornford, D.S. (1989). *Computer systems development: History, organization and implementation.* Chichester, UK: John Wiley & Sons.

Haines, A. & Jones, R. (1994). Implementing findings of research. *British Medical Journal, 308*(6942), 1488-1492.

Hanka, R., O'Brien, C., Heathfield, H. & Buchan, I. E. (1999). WAX ActiveLibrary: A tool to manage information overload. *Topics in Health Informatics Management, 20*(2), 69-82.

Harno, K., Paavola, T., Carlson, C. & Viikinkoski, P (2000). Patient referral by telemedicine: effectiveness and cost analysis of an Intranet system. *Journal of Telemedicine and Telecare, 6*(6), 320-329.

Heathfield, H. & Louw, G. (1999). New challenges for clinical informatics: knowledge management tools. *Health Informatics Journal, 5*(2), 67-73.

Hersch, W.R. & Lunin, L.F. (1995). Perspectives on medical informatics: information technology in health care. Introduction and overview. *Journal of American Social Infrastructure Science, 46*, 726-727.

Korpela, M. (1994). *Nigerian practice in computer systems development. A multidisciplinary theoretical framework, applied to health informatics.* Doctoral thesis. Helsinki University of Technology, Department of Computer Science Reports 1994: TKO–A31. Helsinki.

Lillrank, P., Paavola, T., Harno, K. & Holopainen, S. (1999). The impact of information and communication technology on optimal resource allocation in healthcare. *Proceedings of the International Conference on TQM and Human Factors QERGO,* Linköping, Sweden (pp. 169-174).

Littlejohns, P., Wyatt, J. & Garvican, L. (2003). Evaluating computerised health information systems: hard lessons still to be learnt. *British Medical Journal, 326*(7394), 860-863.

Nonaka, I. & Takeuchi, H. (1995). *The knowledge-creating company.* New York: Oxford University Press.

O'Brien, C. & Cambouropoulos, P. (2000). Combating information overload: A six-month pilot evaluation of a knowledge management system in general practice. *British Journal of General Practice, 50*(455), 489-490.

Patel, V.L., Arocha, J.F. & Kaufman, D.R. (1999). Expertise and tacit knowledge in medicine. In R.J. Sternberg & J.A. Horvath (Eds.) *Tacit knowledge in professional practice. Researcher and practioner perspectives* (pp. 75-100). London: LEA.

Pauker, S.G., Gorry, G.A., Kassirer, J.P. & Schwartz, W.B. (1976). Towards the simulation of clinical cognition. Taking a present illness by computer. *American Journal on Medicine, 60*(7), 981-996.

Polanyi, M. (1966). *The tacit dimension*. New York: Doubleday & Co.

Polanyi, M. & Prosch, H. (1975). *Meaning.* Chicago: The University of Chicago Press.

Ragupathi, W. (1997). Health care information systems. *Communications of ACM, 40*(8), 81-82.

Sackett, D.L., Richardson, W.S., Rosenberg, W. & Haynes, R.B. (1997). *Evidence-based medicine: How to practice and teach EBM*. London: Churchill Livingstone.

Slawson, D.C. & Shaughnessy, A.F. (1997). Obtaining useful information from expert based sources. *British Medical Journal, 314*(7085), 947-949.

Smith, A.J. & Preston, D. (1996). Communications between professional groups in an NHS trust hospital. *Journal of Management in Medicine, 10*(2), 31-39.

Smith, R. (1996). What clinical information do doctors need? *British Medical Journal, 313*(7064), 1062-1068.

Sternberg, R.J. & Horvath, J.A. (1999). *Tacit knowledge in professional practice: Researcher and practioner perspectives.* London: LEA.

Tannenbaum, S.J. (1994). Knowing and acting in medical practice: The epistemological politics of outcomes research. *Journal of Health Politics, Policy and Law, 19*(1), 27-44.

Turunen, P., Forsström, J. & Tähkäpää, J. (1999). Visions of Health Care Information Systems in Finland by the year 2007 – At the point of intersection between different cultures. In C. Carlsson & R. Suomi (Eds.), *The State of the Art of Information Systems Applications in 2007.* TUCS Publications No. 16, University of Turku, Finland.

Weinberg, A.D., Ullian, L., Richards, W.D. & Cooper, C. (1981). Informal advice- and information-seeking between physicians. *Journal of Medical Education, 56*(3), 174-180.

Wyatt, J. (1991). Use and sources of medical knowledge. *Lancet, 338*(8779), 1368-1373.

This work was previously published in Clinical Knowledge Management: Opportunities and Challenges, edited by B. Bali, pp. 271-284, copyright 2005 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 8.8

Documents and Topic Maps: An Original Way to Manage Medical Records

Frédérique Laforest
LIRIS CNRS UMR 5205, France

Christine Verdier
LIRIS CNRS UMR 5205, France

ABSTRACT

Medical records have been used for a long time with different forms, aims, and usages. This heterogeneity is the result of different professions, ways of working, and needs. It is but prejudicial to querying and sharing data and documents. Moreover, we consider that the system must be as close as possible to a more classical, noncomputerized way of working, such as paper-based medical record, and should thus manage documents. Medical records are often loosely or semistructured documents, impeding easy retrieval. In our approach, a medical record is considered as a set of documents and a set of data. In this article, we propose a software system useful for extracting data from loosely-structured documents coming from different sources and for querying them in a hybrid way. Querying can be done in a navigation space which represents extracted data or entire documents. Two main parts are described: the extraction of data in loosely-structured documents and the navigation in a unified view of documents and data.

INTRODUCTION

Computerized medical information systems have been developed for a long time. One can find two main types of such systems. Each system of the first type has been defined for one specific medical service. It is highly suited to its organization, but it misses openness and generalization. Communication with other systems is difficult, and adaptation to other services or specialties is not trivial. Systems of the second type have been defined for a broad use (many services and many specialties). Some contain common storage for selected critical data, but most of the time they do not adapt to the specific needs of each specialty. Others have tried to offer everything that any specialty requires and have become unusable spaghetti dishes.

But the communication of medical information to each practitioner has become compulsory for quality of care and costs management. As the medical record of a person is dispatched in many places, and as legacy systems have provided very heterogeneous data formats, structures, and ways of working, the gathering of the whole is a big challenge today. Exchange platforms have been proposed. They base their exchange on semistructured data (e.g., XML documents) empowering e-mail concepts by proposing joining documents (letters, images, reports, etc.). Consulting such an exchange platform is easy and convenient for daily care. But it is not sufficient for a profound patient care analysis or even for inter-patient studies. Data are not structured and multiviews on data are not provided.

Such platforms have brought a pleasant way of working. Care practitioners use electronic documents to store and exchange information, strict forms are not provided. This can be viewed as a step backward to the paper-based patient record. But we believe this is rather a new way of imagining computerization, offering a user interface as close as possible to the noncomputerized way of working. We have to couple such systems with new engines that allow for the retrieval of data inside documents, for the gathering of distributed information, for its processing, and for navigation purposes. This is the subject of this article. We present here our ideas and prototypes for the extraction of relevant data in documents, for the gathering of data coming from heterogeneous sources, and for the hybrid navigation among extracted data and documents. Our objective is to make a presentation of the medical record based on user's needs, rather than in a chronological or thematic way.

COMPUTERIZING THE MEDICAL RECORD

Characteristics of the Medical Record

Degoulet and Fieschi (1991) provide a globally agreed definition of the medical record:

The medical record is not limited to doctor's written observations neither to nurses notes. It encompasses all that can be stored about a patient, from demographical data to electro-physiological data or to images. According to this role, the patient record is and remains the main tool for the centralization and the coordination of medical activities.

Another concept is interesting: the encounter concept permits knowing all actions on a patient. At each encounter, the care practitioner builds a document. Encounter documents allow following the temporal rhythm of the patient care. Each document has a factual nature. Contained information concerns a fact and describes this fact. Documents are independent the ones from the others; they are solely understandable. Each new document can be seen as a transaction (using the databases sense of this term) as it has the same ACID (atomicity, coherence, isolation, and durability) characteristics.

Documents may be of different types, according to the encounter type (surgery report, biopsy results, and clinical examination notes). Each document becomes a new piece of the medical record. One can thus provide a new definition for the medical record:

The medical record is a collection of documents, each document providing information about an event in the medical history of the patient.

The medical record has specific characteristics that make it a complex element to study. Among them, the two following elements interest us. First, the medical record is more and more a coordination and collaboration tool. All care practitioners enter the medical record to learn what they need to know and to provide what they need to transmit to others. Computerized tools must provide a view on information that is adapted to the care practitioner profile and to the patient case. Another important point concerns the distribution of information and the mobility of people. It implies two things. First, we require tools for the transparent access to distributed data. Second, the mobility of patients and care practitioners necessitates the introduction of pervasive information systems, including mobile terminals (e.g., PDAs or mobile phones).

Hospitals are today the less distributed organization of the medical domain. Care networks include all types of partners collaborating for patients care: hospital staff, private nurses, pathology laboratories, pharmacies, and nonprofessionals like family. Home hospitalization is built on top of the coordination of care practitioners. The medical record is used as the collaborative tool. Information needs and information management functionalities are different from one care practitioner to the other. Writing means and reading shapes should be adapted to each actor category. Documents-based user interface can provide flexibility and adaptability for end users.

General Architecture of the Project

Information exchange and distant access are now available thanks to the last advances in terms of communication, multimedia, and documents. New principles for numerical documents and information access have allowed thinking of the use of documents as a central computerization mean. These are worldwide topics today. But using documents as an end-user paradigm should not reduce data management. Structured data are still unavoidable for many purposes, as getting graphics on numerical values, activity reporting, but also simply for the navigation in paragraphs of text.

That is the subject of this article. Documents of the patient record are analyzed to extract important information that is used to fill in a database. This database can then be used in a classical way, or as an index for documents retrieval. The use of topic maps permits a combined information search, in documents and in data. Topic maps interest does not stop here, they also allow for the integration of data coming from different sources.

Our proposition consists of an architecture for (1) extracting data from loosely-structured documents and storing them in a database, (2) gathering distributed extracted data and merging them in a XML format for knowledge representation (XTM), and (3) querying the system in a simple browsing way through a graphical user interface based on topic maps. In this architecture, the user is at the beginning and the end of the system. At the outset, the user produces loosely-structured documents (medical record). At the end, the user receives data gathered in a navigation tool. Two types of data are considered: atomic values (fever temperature, glycaemia rate, etc.) and original documents themselves (encounter documents). The global architecture is provided in Figure 1.

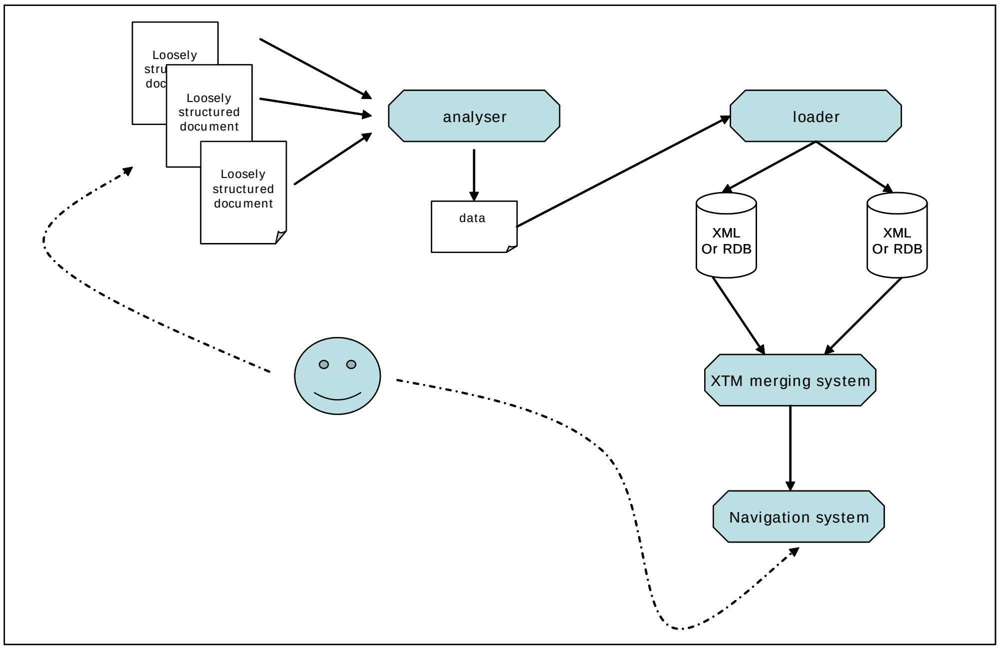
In the following, we first present documents analysis for database filling. Then, we explain what topic maps are, how they are used for data integration, and for the navigation in data and documents.

FROM LOOSELY STRUCTURED DOCUMENTS TO STRUCTURED DATA

Loosely Structured Documents Concepts

Semistructured documents are documents containing information and tags. Tags provide the se-

Figure 1. Global architecture



mantics of information chunks (Abiteboul, 1997). The most used language today is XML (Bradley, 2000; W3C, 1998). With XML, it is possible to define a grammar that defines the allowed tags and the composition rules of tags. Semistructured documents have allowed experts to write documents following a logical organization. For this reason, they are interesting for the care practitioner; medical documents capture is closer to documents writing than to forms filling.

Tags in *loosely-structured documents* do not embrace data but paragraphs. Each paragraph contains a portion of free texts (generally a few phrases) in which many atomic data are drowned. For example, one can define a paragraph to provide the diagnosis information, another paragraph provides a medication prescription, and past medical history is given in a third paragraph type. In a past history paragraph, one can read a date, a family connection, and a disease, but none of these data chunks is tagged. This definition differs from the classical semistructured documents used for data exchange in which all data chunks are tagged. We

Figure 2. Loosely-structured document vs. data centric documents

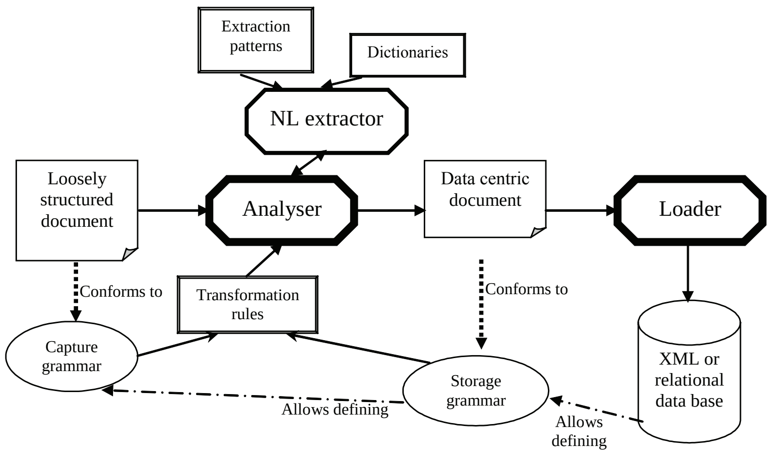
<pastHistory> the patient mother has diabetes. Known since 1990</pastHistory>

<pastHistory> the patient <familyConnection> mother </familyConnection> has <disease id=123> diabetes </disease>. Known since <date> 1990 </date> </pastHistory>

call such documents *data centric documents*, as they are used to organize data. Figure 2 presents the same information piece but written in two ways: first in a loosely-structured document style, and second in a data centric document fashion. This distinction has also been made by Bourret (2000).

Loosely-structured documents provide a new way for capturing information: the end user is guided by paragraphs, but free to write information in a personally preferred way. It is really more comfortable than filling forms or even than

Figure 3. Architecture for the extraction of data from loosely-structured documents



using data centric documents. Such an information capture releases the "computer-related" constraints, allows the user save time, and permits thinking of more intuitive user interfaces. But this also augments the computer work, as data are still required at the end. The computer has to extract data chunks from paragraphs. This extraction allows for building data centric documents and/or for filling a structured database and thus providing the processing power of standard data management systems.

General Architecture for the Transformation

We have designed and developed a prototype of a transformation tool that builds a data centric document from a loosely-structured document (Badr, 2003). Figure 3 provides a synthetic view of our extraction prototype. The analyser takes as input a loosely-structured document and transformation rules. It extracts paragraphs and interacts with a natural language extractor that uses extraction patterns and dictionaries for data chunks extraction in paragraphs. Transformation rules thus define how extracted data should be processed and structured to provide the data centric document. A database loader is finally used to fill in a relational database, or a XML database.

All kinds of documents must conform to a grammar. For each document type (e.g., surgery report or nursing report), two different grammars (DTDs) can automatically be processed from the database schema: a capture grammar that allows validating the loosely-structured documents, and a storage grammar that allows validating data centric documents (see section 4.3.1). The grammar for loosely-structured documents is a simplification of the corresponding data centric document grammar, removing all tags at depth greater than 1.

Steps of the Transformation Process

Data generation from a loosely-structured document contains the following steps:

Step 1: Select paragraphs
Step 2: Extract relevant information
Step 3: Process extracted data
Step 4: Restructure it according to a storage grammar

These steps are described under the form of transformation templates, similar to template-patterns in XSLT. Transformation templates are

Figure 4. Transformation template

```
//transformation template
<Transform paragraph='XPath_Expression' >
      //extraction, structuring and processing rules here
</Transform>

//structuring rule
<Construct frame='tagname' type='frag | defrag'>
//insert here extraction or processing rules
</Construct>

//extraction rule
<Extract pattern = 'Extraction_pattern_name' >
   <dataChunk  name='chunkName1'/>
            …..
   <dataChunk  name='chunkNamen'/>
</Extract >

//processing rule
<value-of operation = 'operation' />
```

gathered in documents called *transformation sheets.* We have built a *transformation processor* that reads a document and gets the transformation sheet associated with its capture grammar. The transformation processor is an augmented XSLT processor. Each transformation template is defined by a *transform* element. Rules are elements inserted in this element.

- A selection rule is a match pattern that identifies paragraphs by operating on the document logical structure. The *paragraph* attribute of the *transform* tag and its value specify the selection rule under the form of an XPath expression.
- An extraction rule (*extract* tag) can be viewed as a function call that specifies arguments and returned values. The function locates matching sequences of words within a selected paragraph. In other words, it makes the data chunks extraction from the selected paragraph. The body of this function is defined internally as an extraction pattern under the form of customized finite state transducers (FSTs) (Silberztein, 2000), as explained later in this article (section 3.4). Within an extraction pattern, each matching sequence of words marks out a *data chunk*. An extraction pattern can return many data chunks.
- Structuring rules (*construct* tag) are used to tag data chunks and insert them into the target document. They do operations like generating constant text, removing XML elements content, moving or reordering XML elements, or duplicating text. Structuring rules put data in a format that is easy to feed the database. They present a super-set of standard XSL elements (xsl:element, xsl:attribute, xsl:copy, and xsl:value-of). We have added the *construct* element type to gather a combination of extraction rules into the target output within a single element. Extracted information are tagged and copied to the target document. If the *type* attribute value is "*defrag*," the entire source text is put into the target document. Otherwise,

only extracted data are inserted.

- Processing rules (*operation* attribute of *value-of* tag) apply operations on extracted information. The result is tagged and inserted into the target document. The basic categories of operations are string operations (comparison, substitution, etc.), arithmetic operations (+, -, /, *, etc.), and mathematical functions (ceiling, floor, etc.).

At the end of the transformation of all paragraphs of a loosely-structured document, extracted data chunks are organized in a XML data centric document. This data centric document also keeps a reference to its source loosely-structured document. The data centric document is thus used to fill in a relational or XML database.

Extraction Details

The natural language extractor (Xtractor) makes the extraction of data chunks in selected paragraphs. After preprocessing the paragraph, it applies FSTs to point at data chunks. FSTs are built by compiling a set of user-defined extraction patterns.

Paragraph Preprocessing

In a loosely-structured document, elements are tagged paragraphs of natural language text. One motivation for document fragmentation is to construct small fragments of a mixture of grammatical and free text; identifying data chunks is thus easier. Since our work is an application oriented system, it uses natural language dictionaries and domain specific dictionaries. In extraction patterns, users can refer to these dictionaries by metasymbols, such as <disease> to refer to any disease in the disease dictionary.

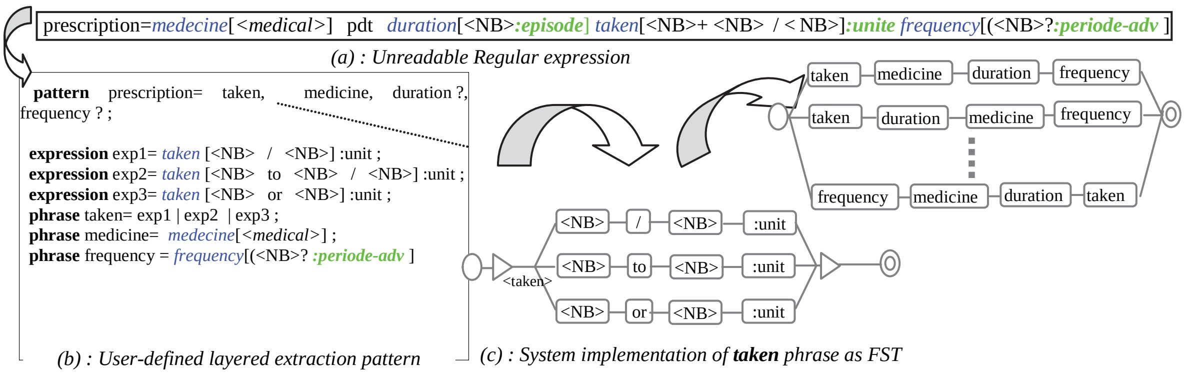
In order to extract relevant information, Xtractor applies its dictionaries against documents paragraphs and defines a scenario of text language preprocessing as follows: (1) validation ensures that the input is a text and it is clean of control characters, (2) preprocessing identifies sentences in paragraphs, delimits unambiguous compound words, and marks off special tokens such as elided words, (3) lexical analysis consists of applying language dictionaries and domain specific dictionaries, and (4) disambiguation removes ambiguity when a word corresponds to more than one entry in dictionaries. After the preprocessing stage, the result is a text vocabulary and indexed dictionaries. Various search operations can thus be done by applying extraction patterns.

User-Defined Extraction Patterns

We provide end users with a high level specification language to build dictionaries of extraction patterns. It is based on regular expressions but includes additional functionalities. We supply a compiler to check the syntax and to translate each extraction pattern into a finite state transducer (FST), as it is required by Xtractor. Figure 5(a) presents a complex extraction pattern. It is not easy to maintain such a pattern. It becomes more readable if it is decomposed into layers as presented in Figure 5(b). We have defined four layers:

1. The *term* layer is composed of a finite set of terms. A term is either (1) a form, that is, a sequence of letters delimited by separators, for example, whitespace, and so forth, and (2) a metawords to describe a number <NB>, a word <MOT>, an empty word <E>, or a reference to lemma in dictionaries.
2. The *expression* layer contains a finite set of expressions; we denote by an expression a concatenation of terms separated by a separator. We say that an expression holds if we find a sequence of words that matches the expression.
3. The *phrase* layer is made of alternates of expressions; we say that a phrase holds if one of its expressions holds. Unary operators

Figure 5. Extraction pattern for prescription paragraph and its equivalent FST



could be applied on expressions in a phrase such as zero or more (*), optional (?), and one or more (+).

4. An *extraction pattern* over a paragraph is a finite and unordered set of phrases. Conditions over phrases may modify the possible combinations. Unary multiplicity operators over phrases increase search paths and a restriction over a phrase sequence reduces possible combinations.

Extraction Patterns Implementation as FSTs

Xtractor uses FSTs to recognize some sequences of words or terms in the input, and associates them with some outputs. All possible combinations defined in an extraction pattern can be illustrated by a graph (Figure 5[d]) with one start node and one end node where every path represents one and only one possible combination of phrases. The skeleton graph of an extraction pattern can be translated into a FST and each phrase can be replaced by its equivalent elementary FST (Figure 5[c]).

Assuming that a FST can reproduce the matching sequence in the output, we can insert additional nodes within the FST to tag relevant sequences and merge their outputs within the original text. That is why we have introduced *named term* as an identifier assigned to a term in order to add semantics and to delimit it by tags in the output. Formally, we note a named term as identifier[term].

Applying FSTs on Documents

Creating a new document type requires effort for describing extraction patterns in the user language we have defined. It may also require defining new dictionaries.

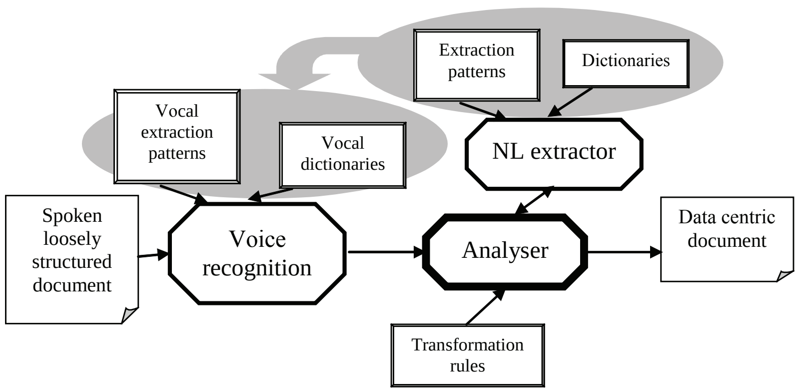
For each new document written by an end user, the analyser sends paragraphs to Xtractor. After preprocessing, it applies FSTs. It results in data chunks identification. These data chunks are returned to the analyser, which continues its transformation steps. For our prototype, we have used the Intex extractor (Intex, 2006). It is a French academic tool provided with French dictionaries and a java API.

User Interface for Documents Capture

The capture of information is here based on the capture of loosely-structured documents. We have studied two ways for the capture of such documents. The first way is the standard one, using a graphical user interface for the end user to type documents. Figure 6 provides a snapshot of our

Figure 6. Standard loosely-structured document capture user interface

Figure 7. Architecture of user interfaces for the capture of spoken loosely-structured documents



prototype. The main central part of the GUI allows the end user to write paragraphs in a free grammatical way. A right click on a paragraph permits the user to define the paragraph type, so that it gets tagged.

But such a GUI cannot be used on a mobile device (even if they nowadays get more and more processing power as well as a bigger and bigger screen). There are many other situations in which the use of a keyboard is not convenient: during surgery for example, or simply when used by a naive end user. That is why we have also prototyped a vocal user interface that allows for the dictation of paragraphs to build the loosely-structured document. We have coupled our basic architecture with a voice recognition tool, as shown on Figure 7. The voice recognition tool receives the dictated voice document. We call voice document a speech flow provided by the end user. The voice recognition tool receives a speech grammar. This grammar mainly contains the description of the capture grammar. Each time the speech tool has coupled a sentence and a paragraph, it sends it to the analyzer for further processing. We also have studied the recognition of more precise data chunks during speech analysis. To do so, we have also provided some (rewritten) extraction rules to the speech recognizer. It can easily identify words coming from a dictionary (like medication names). It thus reduces the further analysis made by the analyzer and the natural language extraction tool.

For the speech recognition prototype, we have used the IBM ViaVoice system. As it has a java interface, it has been easily coupled to our analyzer. The rules for the speech recognition are written in the Java speech grammar format (Sun, 1998). We have validated our speech recognition prototype on medication prescriptions. It has proved that the doctor can dictate the prescription in a "normal" sentence way, without indicating the logical organization of the sentence. The speech flow is continuous and regular. No stop is required between words or phrases.

FROM DISTRIBUTED DOCUMENTS AND DATA TO A UNIFIED VIEW

After a first part dealing with data extraction from loosely-structured document, we present the architecture of hybrid navigation in a space gathering data and documents. This architecture is based on topic maps, a paradigm useful for knowledge representation.

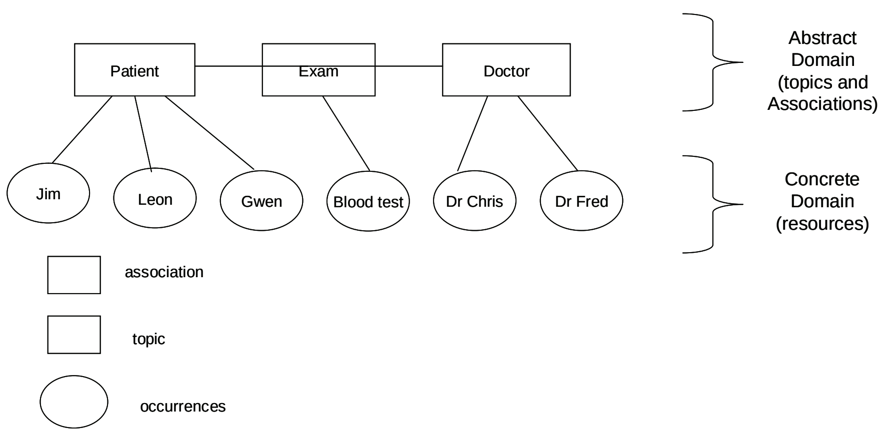
Topic Maps Concepts

In 1999, a standard defining the topic maps model and its syntax was edited by ISO. This standard has been completed in 2002 (ISO, 2002). According to the ISO definition, topic maps are a paradigm used to formalize and organize human knowledge to make creation and retrieval easier in computer processing. It is also used as a mechanism for representing and optimizing resources access. Other definitions put their aim forward. For Fresse (2000), topic maps build a structured semantic link network on resources. According to Widhalm and Mueck (2003), topic maps are a standardized modelling approach for the semantic annotation and description of resources. They enable an improved search and navigation among information stored in semistructured information spaces.

According to Biezunski (2003), topic maps represent a tool useful to organize information in a way that is optimized for navigation. It addresses the problem of infoglut that we are facing in the Web environment. Too much information is almost equivalent to no information, unless there are ways to filter and to extract it efficiently.

A topic map is built from topics linked by associations. Topics and associations represent

Figure 8. A topic maps representation



the abstract part of a topic map. The concrete part is represented by occurrences. A topic occurrence can be any information that is relevant to a given topic. Associations are used to express knowledge at the topics level and not at the occurrences level. Figure 8 shows the representation of a topic map.

The abstract (topics and relations) and concrete (occurrences) parts allow for coupling topic maps with any technique associating data and structure. The most natural approach, proposed in this work, is to couple topic maps with XML (documents structures and documents instances). Another is to couple topic maps with description logics (ABox and TBox).

We use the following main concepts of topic maps: subject, topic, association, and role:

- **Subject:** In the most generic sense, a subject is anything regardless whether it exists or has any other specific characteristics, about which anything may be asserted by any means. In particular, it is anything on which the author of a topic map chooses to discourse. In order to discourse on a subject within the topic map paradigm, that subject must be reified by a topic.
- **Topic:** A topic is a resource that acts as a proxy for some subject; it is the representation of that subject. The relationship between a topic and its subject is a reification relation. Reification of a subject allows topic characteristics to be assigned to the topic that reifies it.
- **Association:** An association is a relationship between one or more topics, each of which plays a role as a member of that association. Each individual association is an instance of a single class of association (also known as an association type) that may or may not be indicated explicitly. The default association type is defined by the "association" published subject.
- **Role:** A role expresses the nature of a topic's involvement as a member of an association.

General Architecture for the Unified View Construction

According to our architecture in Figure 3, the data integration part consists of using XML or relational databases to build a unified view under the form of a topic map schema. This schema represents the conceptual basis for the querying. Two main goals have thus to be achieved: (1) to offer a friendly graphical user interface to navigate through a coherent data space, and (2) to offer access to original documents from selected data.

The global architecture of this part is represented by three main steps: the merging of XML documents (representing the data sources), the creation of the global topic map (base of the navigation), and the interactive creation of a document resulting from a map navigation (see Figure 9).

Steps for Data Integration Based on Topic Maps

The integration of data coming from multiple data centric documents is a complex process. In this paragraph, we explain the main steps of this process, which include exporting the relational database into an XML format (for relational database storage of data), merging the XML documents, and building the topic map.

Database Exportation in a XML Format

Rules used to create a XML document from the DB conceptual schema are the following:

- An *element* is created for each table (the element's name is the table's name). For example, Table Patient produces the XML

Figure 9. Architecture of integration and navigation process

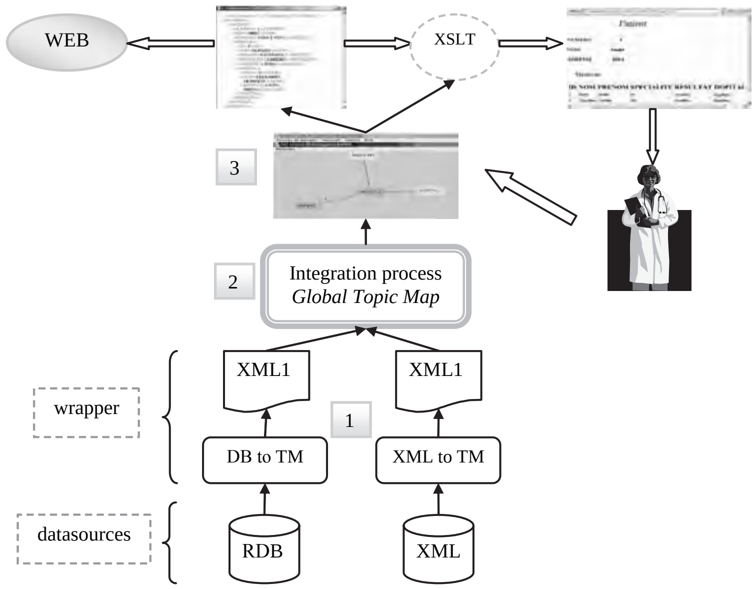
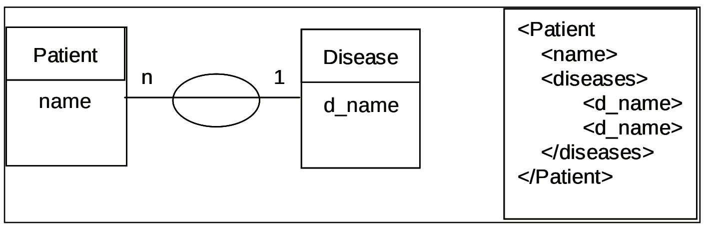


Figure 10. XML translation of 1-N association



element <Patient> …….</Patient>.

- For each *(1-n) relation*, the XML element of the table, which is the source of the functional dependency (cardinality 1), is a subelement of the table element (Figure 10).
- For each *(n-n) relation*, two representations can be done in two views (Figure 11).

XML Documents Merging

The process of integration is shown using an example. Let us consider two XML documents. The merging intends to produce a XML global document which keeps the data semantics and includes data without information loss.

A XML document can be represented using a graph structure G(V, E) in which the vertices represent the document elements (attributes, classes, etc.) and the edges represent the links between the vertices. An edge is a binary relation (association between 2 elements). In our example, (Patient, Doctor) represents an oriented edge from the Patient element to the Doctor element.

Let the G1(V1, E1) and G2(V2, E2) be two graphs representing the XML1 and XML2 documents.

V1 = {Patient, Doctor, Hospital, Treatment}

Figure 11. XML translation of N-N association

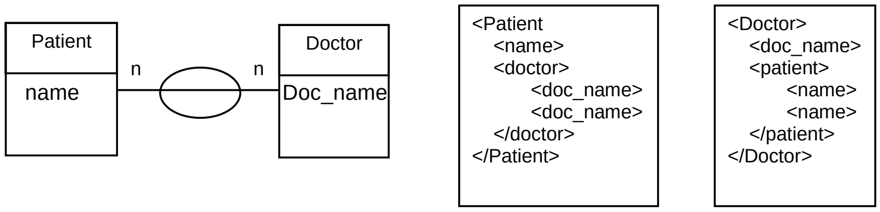


Figure 12. Graph representation of a XML structure

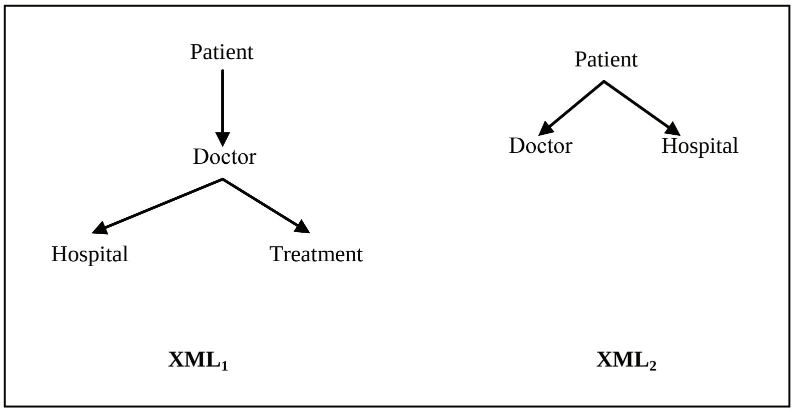
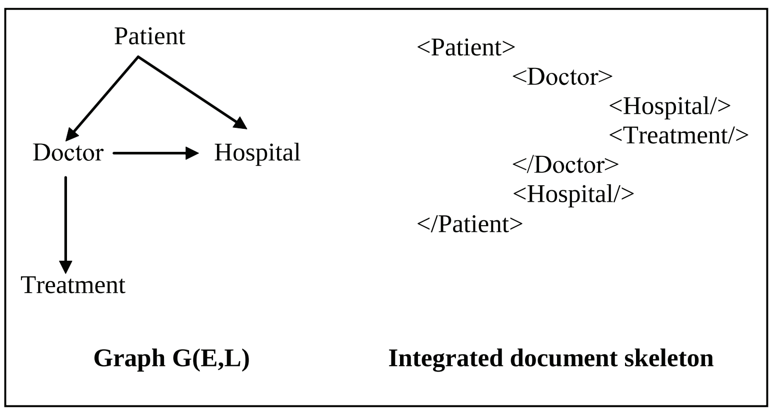


Figure 13. Data merging



E1 = {(Patient, Doctor), (Doctor, Hospital), (Doctor, Treatment)}
V2 = {Patient, Doctor, Hospital}
E2 = {(Patient, Doctor), (Patient, Hospital)}.

The merging results in one graph (Figure 13) called $G(V,E) = G1(V1, E1) \cup G2(V2, E2) = G(V1 \cup V2, E1 \cup E2)$ and gives a new XML document:

$V = V1 \cup V2$ = {Patient, Doctor, Hospital, Treatment}
$E = E1 \cup E2$ = {(Patient, Doctor), (Patient, Hospital), (Doctor, Hospital), (Doctor, Treatment)}

The final document can contain redundancies. In our example, the Hospital element is represented twice in two different places in the graph. The edge (patient→hospital) is included in the edge (patient→doctor→hospital). The first edge says "a patient *p* is cared by a doctor *d* in a hospital *h*." The second edge says "a patient *p* is cared in the hospital *h*." On syntactic aspects, the first edge is contained in the second one. But in case of both sets disjunction, the redundancy cannot be suppressed. A link will be suppressed only if intentional and extensional redundancies can be suppressed. Let us consider the example seen in Figure 14.

The link (patient→hospital) is structurally redundant but the edge values are different in the documents. The hospital h3 in XML2 is not an instance of XML1 and cannot be represented in the path (patient→doctor→hospital). There is no link between hospital h3 and any doctor in both data sources. So the suppression of the link (patient→hospital) is not possible in the resulting document.

Topic Map Building

The previous step gives a XML document representing the merging of database. This step represents the topic map building. Tags in the global XML document are used to produce our topics. We define two types of tags. A composition tag is a unit of data essentially corresponding to a relational entity. The tags <doctor>, <patient>, and <treatment> are composition tags. A data tag corresponds to a relational attribute (e.g., <id_patient> and <doc_name>). Data tags are not translated as topics, they are descriptors of topics.

A composition tag will be translated into a topic in the topic map. The display name of a topic is the name of the composition tag it comes from. These topics are represented in a graph structure as composed vertices. A relation between composition tags is translated into a link between the corresponding topics. The content of each topic is stored using a dedicated XML structure that includes its data tags. The result of this step is a global topic map in which the user chooses the topics to query.

User Interface for Hybrid Navigation

The user should browse the topics space in the easiest way. The interface is built dynamically. When the user selects a topic, the GUI presents this topic and other topics directly linked with it. The structure of a topic can be shown (Figure 15 upper) and its content is displayed as a table to the end user (Figure 15 lower).

Figure 14. Document's instances XML1 and XML2

```
<patient id=p1>
        <doctor id=d1>
            <hospital id=h1/>
            <treatment/>
        </doctor>
        <doctor id=d2>
            <hospital id=h2/>
            <treatment/>
        </doctor>
</patient>
```

XML_1

```
<patient id=p1>
        <doctor id=d1>
        <hospital id=h1/>
        <hospital id=h3/>
</patient>
```

XML_2

Valid Topic

At the beginning of a browsing session, the topic map is centred on a start topic which represents the valid topic at the moment. Only a part of the global topic map is displayed. This part, called valid part, is composed of topics directly linked with the start topic and of associations between the displayed topics. The construction of the valid part can be designed with a specific algorithm (not given in this article).

As the valid part has been created, it is displayed as boxes (topics) connected with straight lines (associations). The user navigates these boxes in two ways: using a graphical SQL-like querying system, or browsing through the global topic map.

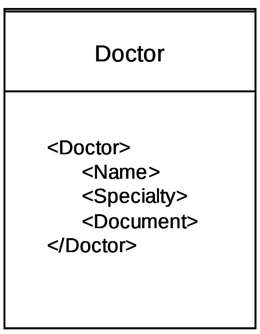
Each topic produces two links: one at the semantic level and one at the instances level.

- **Semantic level:** When the user navigates from a topic t_1 toward a topic t_2, the result is the topic t_2 **of** t_1. Navigating from the Patient box to the Doctor box through the care association, the result is doctors caring for one or more patients.
- **Instances level:** A knowledge base is built with the resources of the data sources. Relevant instances are extracted to be displayed according to their content and the query of the user. The cardinality is displayed on the links between topics and the content can be displayed. The instances are considered at two levels. The first level concerns data (relational or XML data) and the second level concerns the source document of data. The number of instances relevant for the query is displayed on the topic map.

Topic Map Querying

We have proposed algebra to manage dynamic querying (Ouziri, 2003). The operators of this algebra are used to translate the real time actions of the user into actions on the presented topic map. At each graph node, the system calculates the next node according to the user action. This algebra

Figure 15. Topic content



Structure of the Doctor topic (XML coding)

Name	Specialty	Document
Jim	Radiologist	http://www.medical_record/radio.gif
Paul	Cardiologist	http://www.medical_record/heart_image.gif

Content of the Doctor topic (XML coding) partially uses the Xpath syntax (Berglund, 2006) and extends it with operands *This, Next, Previous,* and *Currents* and with other algebraic operators of *Project, Select*, and *Access.* These last operators are useful for the graph querying and for the graph updating at each step of user navigation. The Next and Previous operators are not usually used in the graph theory. The main approach in this theory is to calculate optimal paths in a static graph. But in our case, the graph structure changes every time the user makes a navigation action. So, the navigation operators recalculate the graph structure and instances.

The Access, Next, and Previous operators produce a graph structure that is built according to the semantic links extracted from data sources and restricted to the constraints defined by the user during the navigation. The Access operator browses the graph to calculate the instances linked to the start concept. The Previous operator is the exact (semantic) opposite of the Next operator but does not produce necessarily the same results. For example, the Next operator used from Patient to Doctor gives for each patient the doctors. The opposite (Previous) operation gives for each doctor the list of patients. The Access operator is a generalization of the two operators Next and Previous. It allows for reaching all instances from the "source" topic following a chosen path.

These operators allow for the navigation in the graph. The graph structure is modified according to the user query so that only relevant information is displayed. The screenshot in Figure 16 shows in the background the topic map presentation to the end user. We can see two boxes representing two topics and links showing their associations. Values on links represent occurrences numbers. In the popup window there is a form which allows for the selection of an item.

RELATED WORKS

Works on relations between documents and databases can be seen under three directions: information retrieval, information extraction, and databases.

Information retrieval aims at selecting a subset of documents from a collection based on a user query (Witten, Moffat & Bell, 1994, p. 519). The user browses the results to get the desired information. Indexing techniques allow fast access. But the semantic ambiguity of natural language is not taken into account. Structured documents databases allow using the structure in queries. But anyway, information retrieval does not allow

Figure 16. Navigation between patient to doctor

applying manipulations on data inside documents. Relevant data hidden in text are not structured and they are not accessible for efficient processing.

Information extraction from textual documents intends to extract information and to encode data into a format suitable for populating a database (Muslea, 1999). They are based on extraction patterns built within a learning phase, such as Sundance (Riloff, Wiebe & Phillips, 2005), PALKA (Kim & Moldovan, 1995), and LIEP (Huffman, 1996). Extraction patterns are rarely accessible to the user. On another hand, wrappers are specialized information extraction systems dedicated to Web sites. TSIMMIS (Garcia-Molina, Papakonstantinou, Quass, Rajaraman, Sagiv, Ullman et al., 1997), RoadRunner (Crescenzi, Mecca & Merialdo, 2001), RAPIER (Califf & Mooney, 1999), and NoDoSE (Adelberg 1998) are well-known wrappers generators. Most of them focus on HTML delimiters and do not provide any specification about how to map extracted information to a database.

The database community has studied the retrieval, manipulation, and restructuring of structured documents stored in databases. Some have defined new XML databases, such as Timber (Jagadish, Al-Khalifa, Chapman, Lakshmanan, Nierman, Paparizos et al., 2002), and Tamino (Schöning & Wäsch, 2000), others transform the hierarchical XML model to a relational or object-relational model to store documents in such databases as XML-DBMS (Bourret, 2000), and X-Ray (Kappel, Kapsammer & Retschitzegger, 2001), and some works also provide built up documents as results to queries, but cannot get documents as input (SQLserver, 2000).

Works concerning navigation are essentially oriented towards the semantic Web and information retrieval. The Web is considered as a graph of documents connected by hyperlinks (Pinkerton, 1994). Some drawbacks have an impact on keyword search and on navigation. The user navigates a large collection of pages following the hyperlinks defined by the Web page designer. The resulting document represents the only and unique granularity of the result. Finding a specific term in a document is not possible. For example, getting all the blood pressure values of a patient is not possible. The smaller granularity is the Web page. To improve information search, many works have been done to face the problem of infoglut and irrelevant documents. We can cite Pant, Bradshaw, and Menczer (2003) who propose to link better a crawler and the search engine to improve the focusing on a specific community's interest. Some works are interested in information retrieval (Baeza-Yates, 2003), Web query processing (Diao, Lu, Chen & Tian, 2000), semantic Web (Koivunen, 2001), and adaptive Web navigation (Dolog, Henze, Nejdl & Sintek, 2003).

Other research works face the problem of querying. In these projects, queries are based on the content of Web pages and on the link structure connecting Web pages. A first group of works try to structure Web querying with ordered arc-labeled trees. Among them, we can find works by Arocena and Mendelzon (1999), Fernandez, Fisher, Gruber, and Mandelbaum (2006), and Konopnicki and Shmueli (1998).

CONCLUSION

In this article, we have described a software system useful for extracting data from loosely-structured documents. The extraction allows the expert (physician, computer scientist, etc.) to retrieve data or parts of medical documents. The extraction phase gives a very precise data and document representation. The retrieval phase is made easier with a navigation based on a topics cartography that presents the semantically correct information space to the user and helps with navigation. This navigation is intuitive, easy, and secure. As a further work, we intend to build a full integrated prototype to test our application on real medical data.

ACKNOWLEDGMENT

The authors want to particularly thank Nada Al Masri, Youakim Badr, Stéphane Frénot, and Mourad Ouziri for their precious help and their different works in the domains covered by this article.

REFERENCES

Abiteboul S. (1997). *Querying semi-structured data*. Paper presented at the International Conference on Database Theory (ICDT) (pp. 1-18), Delphi, Grèce.

Adelberg, B. (1998). NoDoSE: A tool for semi-automatically extracting semi-structured data from text documents. In *Proceedings of the ACM SIGMOD Conference*.

Arocena, G.O., & Mendelzon, A.O. (1999). WebOQL: Restructuring documents, databases and web. *TAPOS, 5*(3), 127-141.

Badr, (2003, December). Couplage documents et bases de données: Etude et réalisation. Unpublished doctoral thesis, INSA de Lyon.

Baeza-Yates, R. (2003). Information retrieval in the Web: Beyond current search engines. *International Journal on Approximated Reasoning, 34*(2-3), 97-104

Berglund, A., et al. (2006). *XML path language (XPath)*, Version 2.0, (XSL WG). IBM Research <alrb@us.ibm.com>

Biezunski, M. (2003). Introduction to the topic maps paradigm. In J. Park & S. Hunting (Eds.), *XML topic maps: Creating and using topic maps for the Web*. Addison-Wesley.

Bourret, R. (2000). *XML and databases*. Retrieved May 27, 2007, from http://www.rpbourret.com/xml/XMLAndDatabases.htm

Bradley, N. (2000). *The XML companion*. London: Editions Addison-Wesley.

Califf, M.E., & Mooney, R.J. (1999). Relational learning of pattern-match rules for information extraction. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99)* (pp. 328-334). Orlando, Florida.

Crescenzi, V., Mecca, G., & Merialdo, P. (2001). *RoadRunner: Towards automatic data extraction from large Web sites*. Paper presented at the 27th International Conference on Very Large Databases.

Degoulet P., & Fieschi, M. (1991). *Traitement de l'information médicale: méthodes et applications hospitalières*. Paris: Masson.

Diao, Y., Lu, H., Chen, S., & Tian, Z. (2000). Toward learning based query processing. In *Proceedings of the 26th International Conference on Very Large Databases*, Cairo, Egypt.

Dolog, P., Henze, N., Nejdl, W., & Sintek, M. (2003). Towards the adaptive semantic web. In *Proceedings of the International Workshop on Principles and Practice of Semantic Web Reasoning at International Conference on Logic Programming* (ILP), Mumbai, India. Springer Verlag.

Fernandez, M.F., Fisher, K., Gruber, R., & Mandelbaum, Y. (2006). PADX: Querying large-scale ad hoc data with XQuery. *Plan-X*, 24-35.

Fresse, E. (2000). *Using topic maps for the representation, management and discovery of knowledge*. Paper presented at the XML Europe 2000, Paris.

Garcia-Molina, H., Papakonstantinou, Y., Quass, D., Rajaraman, A., Sagiv, Y., Ullman, J., et al. (1997). The TSIMMIS approach to mediation: Data models and languages. *Journal of Intelligent Information Systems*, *8*, 117-132.

Huffman, S.B. (1996). Learning information extraction patterns from examples. Connectionist, statistical, and symbolic approaches to learning for natural language processing. *LNAI, 1040*, 246-260. Berlin: Springer Verlag.

Intex, (2006). Intex, linguistic development environment. Retrieved May 27, 2007, from http://www.nyu.edu/pages/linguistics/intex/

ISO/IEC 13250. (1999). *Topic maps.* Dec ISO/IEC FCD.

ISO/IEC 13250. (2002). *Topic maps, document description and processing language* (2nd ed.). Retrieved May 27, 2007, from http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0322_files/iso13250-2nd-ed-v2.pdf

Jagadish, H.V., Al-Khalifa, S., Chapman, A., Lakshmanan, L.V.S., Nierman, A., Paparizos, S., et al. (2002) TIMBER: A native XML database. *The VLDB Journal*, *11*(4), 274-29.

Kappel, G., Kapsammer, E., & Retschitzegger, W. (2001). *XML and relational database systems: A comparison of concepts*. Paper presented at the International Conference on Internet Computing (1) (pp. 199-20).

Kim, J.T., & Moldovan, D.I. (1995). Acquisition of linguistic patterns for knowledge-based information extraction. *IEEE Transaction Knowledge Data Engineering, 7*(5), 713-724.

Koivunen, M. (2001). *W3C semantid Web activity*. Paper presented at the Semantic Web Kick-Off Seminar, Finland.

Konopnicki, D., & Shmueli, O. (1998). Information gathering in the World Wide Web: The W3QL query language and the W3QS system. *ACM TODS*, *23*(4), 369-410.

Muslea, I. (1999). *Extraction patterns for information extraction tasks: A survey.* Paper presented at the AAAI 1999 Fall Symposium on Narrative Intelligence, North Falmouth, Massachusetts.

Ouziri, M. (2003, December). *Utilisation des Topic Maps pour l'interrogation et la generation de documents virtuels: Application au domaine medical*. Unpublished doctoral thesis, INSA de Lyon.

Pant, G., Bradshaw, S., & Menczer, F. (2003). Search engine-crawler symbiosis. In *Proceedings of the 7th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2003)*, Trondheim, Norway.

Pinkerton, B. (1994). *Finding what people want: Experiences with the Webcrawler.* Paper presented at the Electronic Proceedings of the 2nd World Wide Web Conference.

Riloff, E., Wiebe, J., & Phillips, W. (2005). Exploiting subjectivity classification to improve information extraction. In *Proceedings of the 20th National Conference on Artificial Intelligence (AAAI-05)*.

Schöning, H., & Wäsch, J. (2000). *Tamino: An Internet database system*. Paper presented at the EDBT 2000 (pp. 383-387)

Sun Microsystems, Inc. (1998, October 26). Java speech API programmer's guide. Retrieved May 27, 2007, from http://java.sun.com/products/java-media/speech/forDevelopers/jsapi.doc

W3C. (1998). Extensible markup language (XML). Retrieved May 27, 2007, from http://www.w3c.org/XML/#9802xml10

Widhalm, R., & Mueck, T.A. (2003). Merging topics in well-formed XML topic maps. *Lecture Notes in Computer Science*, *2870*, 64–79. Springer-Verlag.

Witten, I.H., Moffat, A., & Bell, T.C. (1994). *Managing gigabytes: Compressing and indexing documents and images* (1st ed.) New York: Van Nostrand Reinhold.

This work was previously published in the International Journal of Healthcare Information Systems and Informatics, Vol. 2, Issue 4, edited by J. Tan, pp. 22-40, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 8.9
A Prehospital Database System For Emergency Medical Services

Nada Hashmi
10Blade, Inc., USA

Mark Gaynor
Boston University School of Management, USA

Marissa Pepe
Boston University School of Management, USA

Matt Welsh
Harvard University, USA

William W. Tollefsen
Boston University School of Medicine, USA

Steven Moulton
Boston University School of Medicine, USA

Dan Myung
10Blade, Inc., USA

ABSTRACT

Emergency Medical Services (EMS) are not only responsible for providing prompt and efficient medical care to many different types emergencies, but also for fully documenting each and every event. Unfortunately, the vast majority of EMS events are still documented by hand. The documents are then further processed and entered manually into various billing, research, and other databases. Hence, such a process is expensive, labor intensive, and error prone. There is a dire need for more research in this area and for faster, efficient solutions. We present a solution for this problem: Prehospital Patient Care Record (PCR) for emergency medical field usage with a system called iRevive that functions as a mobile database application. iRevive is a mobile database application that is designed to facilitate the collection and management of prehospital data. It allows point-of-care data capture in an electronic format and is equipped with individual patient sensors to automatically capture vital sign data. Patient information from the field is wirelessly transmitted to a back-end server, which uses Web service standards to promote interoperability with disparate hospital information systems, various billing agencies, and a wide variety of research applications. In this chapter, we describe the current state of EMS, the iRevive application, a mini-trial deploying iRevive in real scenarios, the results, and a future direction for our solution.

INTRODUCTION

There are times when an individual's life may depend on the quick reaction and competent care of emergency medical technicians (EMTs). These highly trained, prehospital healthcare providers are dispatched by 911 operators to incidents as varied as motor vehicle crashes, heart attacks, near-drowning events, childbirth, and gunshot wounds. Their first priority is to stabilize a patient's cardiopulmonary status. They must then determine the nature and severity of the patient's condition and whether the patient has any preexisting medical problems. EMTs follow strict rules and guidelines in their provision of emergency care and often use special equipment such as backboards, defibrillators, airway adjuncts, and various medications before placing patients on stretchers and securing them in an ambulance for transport. At a medical facility, EMTs transfer the care of their patients to emergency department personnel by reporting their observations and actions to staff.

Equally important is EMS personnel documenting the care they provide. They do so in the form of a prehospital record, which must be completed for each patient who is treated or transported by them. The prehospital record is a medical and legal document used by emergency medical technicians to record a variety of data concerning a patient's current illness or injury, past medical history, treatment rendered, and subsequent improvement or worsening of the patient's condition (Mann, 2002). This type of prehospital documentation is used to support the actions of the crew, the transfer of care, and to justify reimbursement from various insurance companies; it is also used for quality improvement programs and research. Unfortunately, the vast majority of EMS events are still documented manually by hand on paper. This leads to an extensive amount of manual data processing as the often illegible handwritten data must sometimes be deciphered, then manually entered into various billing, research, and other databases. The whole process is expensive, labor intensive, and error prone.

The rest of this chapter is sectioned as follows: first, an overview of the current state of EMS workflow, documentation methods, and research is provided. This section emphasizes the National Highway Traffic Safety Association's goals for EMS in the future, including the call for a national EMS database and improved information systems, so that prehospital information can be linked with the hospital record. The next section is a description of one solution called iRevive, a mobile database for EMS professionals that streamlines data capture, communication, reimbursement processing, quality assurance, and research. It takes advantage of tiny wireless sensors to automatically record vital sign data. It permits multilevel decision support; the local EMT over his/her patient, the regional commander over a selected vicinity, and the central level of control over all the events occurring at a particular time. Actual deployment of iRevive, for live field-testing by Professional Emergency Services of Cambridge, Massachusetts, is examined and critiqued. This trial version was conducted without sensors or multilevel decision support. Finally, a future vision of iRevive is described, including the addition of many different types of sensors such as chemical sensors and GPS devices for location information. All exchange of data will be interfaced through Web services and conform to standards such as HL-7 to help the increase of data exchange and interoperability.

Background

Ambulances of the early 1900s were regarded as a means of transportation for the sick and injured from homes, work site, and public places to hospitals, where real treatment could begin. It was not until the advent of cardiopulmonary resuscitation (CPR) and the 1966 publication of a National Academy of Sciences paper entitled, "Accidental Death and Disability: The Neglected Disease of Modern Society," that modern EMS systems came into being (Callahan, 1997). Later, with the introduction of cardiac defibrillation by trained crewmembers and more extensive airway training, the back of ambulances became the sites of true life-saving treatments. While emergency medical services have grown rapidly over the past 30 years, the scope of EMS research has not. Most EMS research focuses on a single intervention or health problem, and it rarely addresses the inherent complexities of EMS systems (Delbridge, Bailey, & Chew, 1998).

It is estimated that EMS systems treat and transport up to 30 million patients per year (NHTSA, 2001) and it is assumed that EMS intervention positively affects patient outcomes, but this is difficult to quantify. Studies have shown that early defibrillation and administration of certain drugs save lives, but other interventions including certain instances of intubations in the field may in fact cause more harm than good (Adnet, Lapostolle, Ricard-Hibon, Carli, & Goldstein, 2001; Vahedi, Ayuyao, & Parsa, 1995). The fact that so few therapies have been examined in outcome studies illustrates a lack of evidence regarding the benefits of many prehospital interventions.

The lack of EMS systems data can be attributed in part to the healthcare industry's delay in utilizing technology; it is one of the last industries to transition to the use of computers for daily operation (Cheung et al., 2001; Foxlee, 1993; Mikkelsen & Aasly, 2001; Tello, Tuck, & Cosentino, 1995). Although some elements of the system are automated (e.g., computer automated dispatch), most EMS personnel record clinical information and other run data using paper and pen. Data collection is therefore limited and highly inefficient. In addition, the patient care report (PCR) that is completed after each EMS transport does not contain data regarding overall patient outcome. The reason for this is that outcome data is held by several different entities, sometimes including other ground and air transport services, hospitals, rehabilitation centers, and physician offices. These various healthcare entities may be affiliated with each other, but seldom are they officially linked and rarely do they exchange prehospital patient information. This lack of information exchange is further hampered by patient privacy laws, incompatible (proprietary) systems, limited data mining methods, and little impetus to form a continuous patient care record. The resultant lack of outcome data severely limits the type and amount of EMS research that can be carried out (Dunford, 2002). Compounding the overall problem is the recognition that serious medical errors can arise in the setting of incomplete data (Foxlee, 1993; Tello et al., 1995). These errors in the handing off of patient care can range from duplicative or delayed therapy to complete lack or inappropriate therapy.

Current Methods

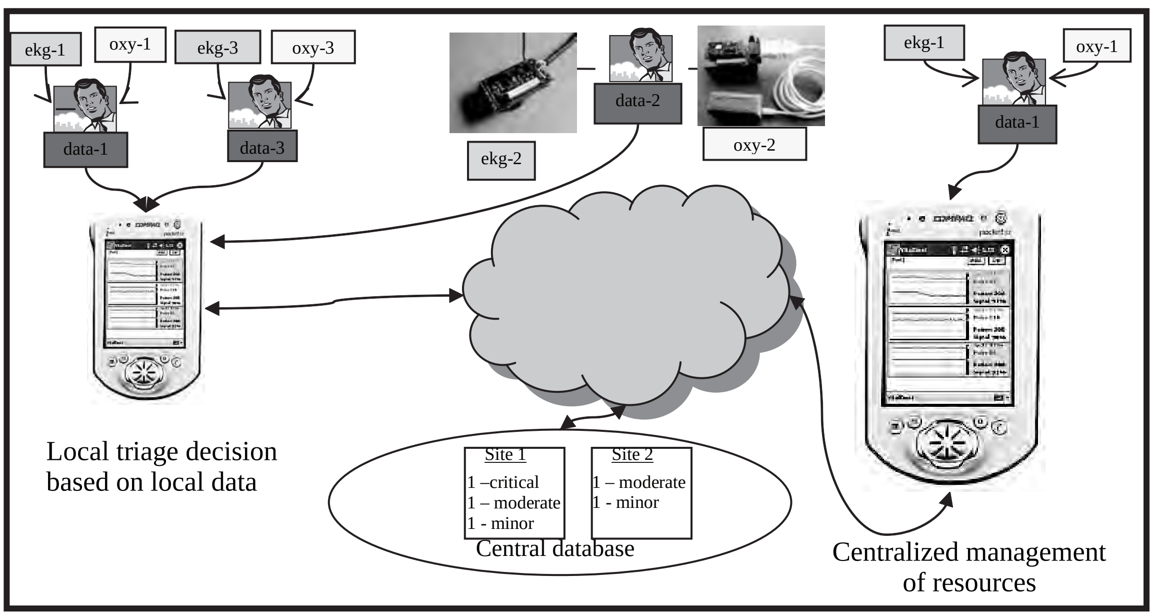
EMS personnel usually work in teams of two and divide the workload at a particular event. While one provider tends to the patient, the other interviews family members or bystanders, sizes up scene conditions, and searches for medications, identification cards, and insurance cards. The NHTSA currently mandates a data set of 40 items to be collected for each patient for each event, including such items as incident location, crewmember identification numbers, patient's social security number, and physical exam findings. Additional insurance and billing information is required by ambulance services so that patients can be billed, while Medicare calls for waivers and prescription forms. It is when the patient's condition is more critical that EMS team members must give their full attention to patient care, forgoing any attempts at data capture or documentation. It is data from these types of events, however, that are of greatest interest to emergency department personnel, researchers, and system administrators. And as EMS systems evolve to offer more advanced care, more time is needed for hands-on patient care and, in turn, more information must be documented (Mears, Ornato, & Dawson, 2002). To overcome these obstacles to better patient care, EMS systems must adopt information systems that streamline the recording, storage, retrieval, and application of quality information.

New methods are being developed to quantify and organize the plethora of data. In 1996, the NHTSA published an article entitled, "EMS Agenda for the Future: Implementation Guide". This article stressed the need for a standardized EMS information system, based on uniform data elements and uniform definitions (NHTSA, 2001). In order to accurately draw conclusions there must be more information regarding care in the field, transportation, emergency department care, hospital care, and final patient disposition. To achieve these goals, EMS information systems must develop new ways to store and retrieve patient data so that patient information is always available. Data must be pooled from a communications center, ambulance personnel, emergency department staff, and finally, other agencies including fire departments, police departments, and medical examiners. Only then can there be a complete database containing all of the information necessary to describe an entire EMS event and facilitate continuous EMS system evaluation and research across multiple systems and to support patient care and EMS-related research (Delbridge, 1998). Our system, iRevive, attempts to address these problems and provides a novel solution in streamlining the data collection. The next section describes the iRevive system built to generate prehospital patient care record.

iREVIVE: THE MOBILE DATABASE SYSTEM

In working toward this vision of compatible EMS database systems that support healthcare data integration, changes must start with regard to how data is originally collected. iRevive is a mobile prehospital database system that allows point-of-care data capture in an electronic format. It consists

Figure 1. Current iRevive system architecture



of a network of wireless, handheld computers running the iRevive application. Wireless vital sign sensors, called VitalDust motes, automatically capture and integrate patient vital sign data directly into each developing patient care record. All of this information can be sent wirelessly from the field to a server that stores and relays selected patient information to receiving hospitals. The stored data is accessible to authorized users for billing and research purposes using Web services (Figure 1).

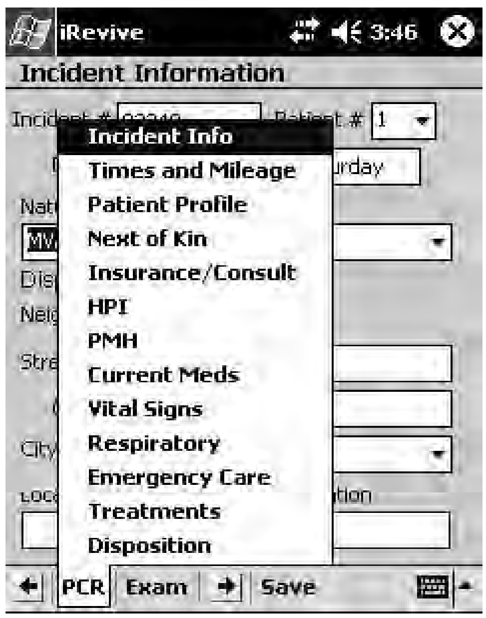
iRevive explores improved documentation of EMS events by streamlining data capture and providing essential prehospital information for subsequent integration with the developing hospital record. Data entry is designed to be logical and intuitive. iRevive can "walk" an EMT through an assessment and remind him/her, for example, to evaluate a patient's neurological response, an essential step that could easily be left out during the high-paced transport of a critically ill patient. Real-time sensors, which collect heart rate and blood oxygen saturation, provide real-time monitoring and enrich the data collection process. During transport or soon after arrival at a hospital, the EMT may choose to complete the PCR using a preset narrative template on the handheld computer. Alternatively, he/she can enter his/her own narrative using either the handheld computer's pen pad or a computer terminal in the receiving hospital's emergency department. Once the PCR is complete and electronically signed, it can be saved (in the central database) and printed (at the hospital) so that it can be included in the patient's paper-based hospital record. The administrative back end aggregates the raw data into summary reports and statistics for operational analysis, performance measurement, and effective managerial decision making.

Prehospital Patient Care Records

The iRevive prehospital patient care record consists of and captures three types of data. The first type of data is the information captured by the EMT which contains the preexisting conditions and the current conditions. Specially, it contains demographic information, a history of the patient's illness or injury (HPI), past medical history (PMH, including medications and allergies), procedural information (e.g., IV access, splinting, and endotracheal intubations), and disposition information (Figure 2). The HPI section includes standardized narratives for more than 30 common complaints, which range from allergic reactions to "dead on scene." The use of standardized narratives allows EMTs to quickly describe each incident using a series of drop-down option boxes, rather than writing an entire incident out in prose—a difficult task on a PDA. The addition of pertinent negatives and additional information in the event of special circumstances is also allowed. These narratives enrich the database and, in the future, will aid in the development of new knowledge-based treatment algorithms.

The second type of data captured is the documentation of critical examinations. It is devoted to an extensive physical exam including head, neck,

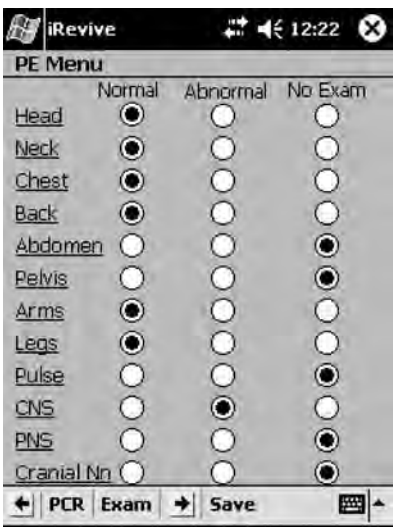
Figure 2. iRevive patient care report sections



chest, abdomen, pelvis, extremities, and nervous system. From the physical exam main menu, users may select and describe any part of the anatomy that has been found to be abnormal (Figure 3a). Selecting an "abnormal" button brings the user to additional pages where abnormal physical findings may be documented in detail using a series of pull-down lists (Figure 3b).

Finally, the third type of data captured and recorded is the sensor data from motes such as the pulse-Oxometer and Vital Dust (Figure 4). The vital signs, such as the pulse rate and SP02 levels, from the patient are automatically recorded and sent continuously from the sensor to the PDA. Critical changes can be immediately noted by the EMT. The EMT no longer has to physically monitor the patient's vital at regular intervals. He/she is automatically provided with more readings which results in a more accurate state of the patient at any particular time.

Figure 3a. iRevive physical exam main menu. If a normal button is selected, a statement is entered in the PCR indicating a normal exam. If an "abnormal" button is selected, then a detailed exam screen opens (CNS = central nervous system; PNS = peripheral nervous system; Cranial Nn = cranial nerves).



Current Application Architecture

The iRevive application is written in C# under the .NET environment and therefore built to run on many types of mobile devices including PDAs, laptops, and wearable computers (Figure 5). iRevive provides a graphical user interface that users can navigate with a stylus. The application is primarily menu driven with customizable drop-down menus that increase efficiency by allowing for quick navigation and data collection. Once the data have been uploaded to a central database, a Web-based interface can be used to edit the PCR on any Internet-enabled PC.

Once field data have been synchronized with the iRevive database, EMS and hospital personnel can instantaneously track current patients and a complete patient care report can be generated and printed. Specific providers have access to the entire record base for billing, supply tracking, and continuous quality improvement (CQI) applications. Other users have the ability to access deidentified data that are nonconfidential for use in overall emergency service research and systems

Figure 3b. A detailed exam screen for the central nervous system (CNS) is shown (GCS = Glascow Coma Score, React. = papillary reactivity).

Figure 4. Sensor Data information transmitted wirelessly to the PDA.

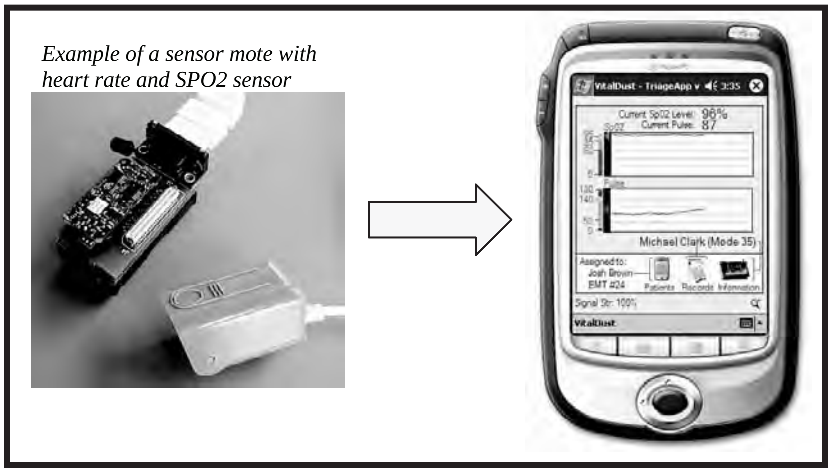
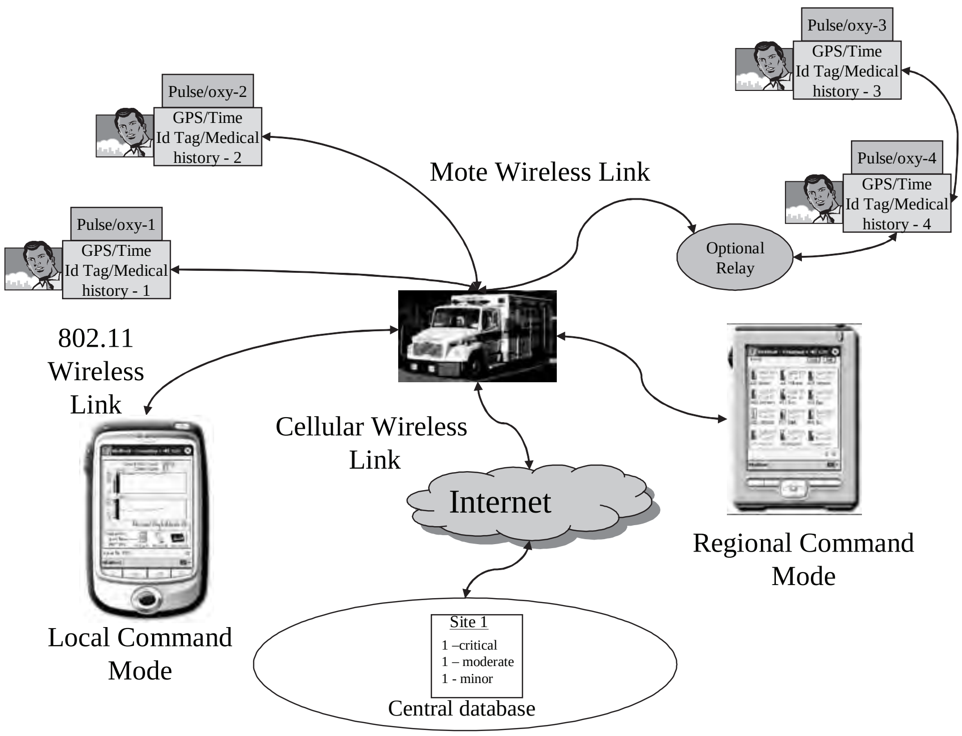


Figure 5. The Future iRevive System Architecture



management. All data transfer is Health Insurance Portability and Accountability Act (HIPAA) compliant and the security is end-to-end (using standard Transport Layer Security). Users must be authenticated to use the system and only authorized users can view specific types of data.

To encourage interoperability, the iRevive system is built around emerging standards such as the National Highway and Traffic Safety Board

Uniform Prehospital EMS Data Set (NHTSA-UPDS) and the Data Elements for Emergency Departments Standard (DEEDS), which is the Centers for Disease Control and Prevention's (CDC's) specification for emergency department data that was created to assist data integration across information resources.

FIELD TRIAL ANALYSIS WITHOUT SENSORS

One ambulance from Professional Emergency Services (Pro) of Cambridge, Massachusetts, was selected to field test iRevive. Pro is a moderate-sized private ambulance company with eight ambulances and 50 emergency field providers who respond to over 16,000 emergency calls per year. iRevive was used in parallel with the service's existing handwritten documentation methods during emergency responses. Data were collected using a Hewlett Packard iPAQ Pocket PC, which was synchronized with a laptop satellite station before being uploaded to the server at the end of each shift. In an attempt to steer further product development, users focused on how the iRevive application could be improved. The effectiveness of iRevive in integrating and streamlining data capture was studied, along with its ability to merge with the current workflow of EMS professionals. Factors examined included ease of use, documentation completion, and content. The use of iRevive in the field was observed by a proficient user and ambulance crew chief. This was further examined via interviews with other ambulance crew members. Printouts of iRevive PCRs were then compared to handwritten reports from the same events.

A Case Study

In one real-life scenario, a call concerning chest pain and difficulty breathing was dispatched to an ambulance. While en route to the location, one EMT began to enter data into iRevive. Information included the incident location, type of response, and other pertinent dispatch information known at the time; the time of response was recorded in the times and mileage page. Upon arrival at the scene, time was recorded again while dispatch was notified of arrival via the radio. The time of arrival at the patient's location in the third-floor apartment was also recorded. The first EMT began an initial assessment, obtaining vital signs, attaching a heart monitor, administering oxygen, and prepping an IV site. At the same time, the second EMT interviewed the patient and bystanders in order to gain a detailed account of the patient's medical history, allergies, medications, and a history of the present illness. All pertinent information was recorded using pull-down menus within the iRevive program.

After recording placement of an IV catheter, oxygen delivery rate, and the type of oxygen device used, the medics determined that the patient should receive one dose each of nitroglycerine and aspirin. This was also recorded in iRevive. Once the patient was placed in the ambulance, a second set of vital signs indicated the need for a second dose of sublingual nitroglycerine, which was recorded in iRevive with a time stamp. En route to the hospital, the first EMT took over use of iRevive to record the time of transport and to complete additional sections as necessary. At the hospital, a verbal report was given to the emergency department staff as the patient's care was transferred to an emergency physician. The iRevive PCR was then completed by an EMT with additional supplemental data pertaining to the transport and final patient disposition. A standardized narrative was completed to finish the event documentation. The PCR was then saved on the handheld and uploaded to the 10Blade server allowing immediate access to the PCR for emergency department, ambulance service, and billing.

Results

iRevive was used in conjunction with the standard method of EMS documentation at Professional Ambulance for 16 emergency transports. Of these transports, 12 were calls for medical help, 3 for trauma, and 1 for an assist. iRevive was used by 12 of the 50 field providers at Pro.

It was noted that both the methods captured the same type of data. However, as the iRevive data were electronic, once captured, they were easily ported to the necessary storage and are now available for any further research or analysis. There were difficulties encountered while capturing the data using iRevive which are presented in the next subsection. These difficulties were taken into consideration for the next version of iRevive.

Difficulties

The main goal behind deploying iRevive in the field was to obtain feedback and find out what problems should be addressed for the next release. Problems experienced can be categorized into two main categories: usage problems and need for more data points (Tollefsen, 2003).

Usage Problems:

- History of present illness page had to be navigated prior to discovering key information. This page would be better placed at the end of the report.
- Placement of times and mileage page at the beginning of the PCR required jumping back from pages in use to record times during procedures.
- Dispatch, scene arrival, scene departure, and hospital arrival times are recorded by the dispatch center; this information could be automatically synchronized with the 10Blade server in place of manually recording each time in iRevive.
- Program did not allow a PCR to be saved unless certain fields were completed.
- Inability to return to saved PCR until synchronized with server caused inability to continue to update report.
- Lengthy pull-down menus for certain items took too long to navigate; these could be improved by reordering or starting in a more appropriate range.

Need for More Data Points:

- Lack of appropriate values in some pull-down menus required manual entry of these values.
- Insufficient space to record multiple allergies.
- Inability to draw a picture of patient or scene. Addition of human figure in order to point to areas injured is recommended. Integration with a digital camera was suggested.

Discussion

Overall, the ability of capturing data in an electronic manner proved to be a huge success. Having electronic data available automatically and as soon as the data are captured is a huge gain from the current state of the art. However, in this domain, the granularity for the type of data and the data content are very important. Even though the current version of iRevive contains all 40 points of the Massachusetts Office of EMS prehospital data set, the need for additional data points was recognized. In order to capture as many data points as possible and make iRevive more attractive to a broader number of EMS providers, several additions have been proposed, including the following:

- Type of response: lights and sirens vs. with traffic flow.

- Location type: to explain difficulties or irregularities (e.g., if the incident occurred at a school, parents would not be present).
- Additional patient information, such as estimated weight.
- Signs and symptoms of chief complaint, such as headache or syncope associated with a cut on a patient's head.
- EMT's impression of the event and patient state (e.g., psychiatric issues).
- Cause of trauma and additional information about the cause, such as severity of automobile damage, and whether the patient was wearing a seat belt.
- Reason for transport (e.g., generalized weakness).
- Additional information about the pain a patient is experiencing during the physical exam.
- Condition of the patient's skin.
- Additional transport information, including position of patient and traffic delays.
- Signature page for capture of patient receipt.

FUTURE VERSION OF iREVIVE

The future version of iRevive aims to increase the usability and efficiency of the complete system. This will be done by the addition of more devices and functionality and increasing the role the ambulance plays (Figure 5).

It was observed the ambulance can play a key role in helping communications and transactions that take place during a response. In the future version of iRevive, the ambulance will play the role of a hub bridging real-time communication from the PDA to the hospital housing the central command. The ambulance will contain a wireless network connection (satellite, cell) as well a global positioning system (GPS) tracking its location and time. Information from the PDA to the central command will be exchanged in real time; the EMTs will be able to receive the patient's information immediately as well as be able to send the condition right away allowing for the hospital to make important decisions such as which hospital can facilitate to the patient's needs. This real-time data exchange will help routing of the ambulances efficiently. Furthermore, the PDAs as well as the sensor motes will also be equipped with GPS and wireless connections. EMTs will be required to match the PDA with the sensor mote of the patient to acquire the vital sign information.

The New Architecture

We plan to acquire and implement the following components (devices):

1. **Sensors:** We plan to integrate additional customizable sensors, medical and nonmedical, that are capable of patient monitoring, data filtration, and ad-hoc networking. An example of such sensors would be chemical sensors to detect levels of toxicity and the GPS sensors to record time and location.
2. **Web services:** In order to ease construction of exchange partnerships and increase flexibility of integration with hospital legacy systems and yet-to-be-released communication protocols, the exchange of data will be transitioned into using Health Level-7 (HL-7) Web services. HL-7 is a widely used data definition and delivery standard that has been recommended by the National Committee on Vital and Health Statistics. The HL-7 data format, a Web service based on open standards that are being accepted by the medical community at large, allows any client to access iRevive's Web service as long as he/she has permission and follows the standards set forth in calling the data. A server node will be the manager of sensor streams. A default is a manager of sensor streams on ambulance which manages data

for connected PDAs and GPS sensors. Other modes for the server node may include:

- Operating with attached databases for further storage.
- Acting as a forwarder to another server that is running with an attached database.
- Act as an aggregator of multiple server nodes running on multiple ambulances.

An example of such a server node would be a daemon running on the ambulance listening for PDAs and hospital databases.

3. **Client device:** A PDA will subscribe to an ambulance to receive sensor streams as well as send PCR data for storage/transmission. Laptops may also be employed at different locations that receive the complete information of which ambulance and sensors are in which locations, and the type of conditions that are being treated. This will help create a FedEx routing of the ambulances.

CONCLUSION

The current methods employed by the EMS for gathering and recording data are out-dated, time consuming, and error prone. New technologies and solutions must be employed to improve the system. iRevive provides a solution to overcome many of the problems. EMTs can save time and efficiently capture the necessary information. Furthermore, the data are in electronic format readily available for exchange and analysis with many different systems. A wireless environment for the exchange of data saves time and provides up-to-date real-time information. Also, iRevive can be employed on a variety of different platforms and devices, including but not limited to PDAs, laptops, tablets, and so forth. It was tested in the field and many improvements were made. More innovative changes are in the process: addition of many more emerging sensor and sensor network technology, embracing of industry standards and accepted data formats, a new level of prehospital care is on the horizon. In time, broad deployment of prehospital applications such as iRevive will allow healthcare providers to monitor and document in real time how various procedures and types of therapy affect patient status and outcome. As more data are accrued, new algorithms will be developed to accurately guide and control all types of medical care. Eventually, automated acute care algorithms will be developed to enable sensor-based, computer-controlled patient care.

ACKNOWLEDGMENT

This material is based upon work support by the National Institute of Health. Any opinions, finding, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Institute of Health.

REFERENCES

Adnet, F., Lapostolle, F., Ricard-Hibon, A., Carli, P., & Goldstein, P. (2001). Intubating trauma patients before reaching hospital—revisited. *Critical Care*, *5*(6), 290–291.

Ananthataman, V., & Han, L.S. (2001). Hospital and emergency ambulance link: Using IT to enhance emergency pre-hospital care. *International Journal of Medical Informatics*, 61(2–3), 147–161.

Callahan, M. (1997). Quantifying the scanty science of pre-hospital emergency care. *Annals of Emergency Medicine*, *30*, 6.

Cheung, N.T., Fung, K.W., Wong, K.C., et al. (2001). Medical informatics—the state of the art in the hospital authority. *International Journal of Medical Informatics*, *63*(2–3), 113–119.

Delbridge, T., Bailey, B., Chew, J., et al. (1998). EMS agenda for the future: Where we are …where we want to be. *Pre-hospital Emergency Care, 2*(1), 1–12.

Dunford, J. (2002). Performance measurements in emergency medical services. *Pre-Hospital Emergency Care*, *6*(1), 92–98.

Foxlee, R.H. (1993). Computer-aided documentation. Quality, productivity, coding, and enhanced reimbursement. *American Journal of Clinical Oncology*, *16*(5), 455–458.

Hill, J., Szewczyk, R., Woo, A., Hollar, S., Culler, D., & Pister, K. (2000, November). System architecture directions for networked sensors. *Proceedings of Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, Cambridge, MA.

Mann, G.E. (2002). *Data capture in the United States healthcare system—the pre-hospital phase of patient care*. Unpublished master's thesis, Boston University School of Medicine–Division of Graduate Medical Sciences.

Mears, G., Ornato, J., & Dawson, D. (2002). Emergency medical services information systems and a future EMS national database. *Pre-hospital Emergency Care*, *6*, 123–130.

Mikkelsen, G., & Aasly, J. (2001). Concordance of information in parallel electronic and paper based patient records. *International Journal of Medical Informatics*, *63*, 123–131.

National Highway Traffic Safety Administration (NHTSA). (2001). National research agenda. Retrieved from *www.nhtsa.dot.gov/people/injury/ems/ems-agenda*

Silver, M.S. (1990). Decision support systems: Directed and non-directed change. *Information System Research*, *1*(1), 47–70.

Silver, M.S. (1998). User perceptions of decision support system restrictiveness: An experiment. *Journal of Management Information Systems, 5*(1), 51–65.

Tello, R., Tuck, D., & Cosentino, A. (1995). A system for automated procedure documentation. *Computational Biology, 25*(5), 463–470.

Tollefsen, W. (2003). *The history, development and field trials of iRevive, a handheld mobile database for use in the pre-hospital setting.* Unpublished masters thesis, Boston University School of Medicine–Division of Graduate Medical Sciences.

Vahedi, M., Ayuyao, A., Parsa, M., et al. (1995). Pneumatic anti-shock garment-associated compartment syndrome in uninjured lower extremities. *The Journal of Trauma*, *38*(4), 616–618.

Welsh, M., Myung, D., Gaynor, M., & Moulton, S. (2003). Resuscitation monitoring with a wireless sensor network. *Circulation*, *108* (Supplement IV), 1037.

This work was previously published in Unwired Business: Cases in Mobile Business, edited by S. J. Barnes & E. Scornavacca, pp. 205-219, copyright 2006 by IRM Press (an imprint of IGI Global).

Chapter 8.10
Outsourcing of Medical Surgery and the Evolution of Medical Telesurgery

Shawna Sando
University of Arizona, USA

ABSTRACT

With rising and often unreasonable costs in the U.S. healthcare system, Americans are becoming more inclined to seek cheaper alternatives. In some cases, Americans do not have to search for such alternatives on their own because their employers are offering them incentives to receive care from a foreign institution. Employees can go abroad to countries, such as India, in order to receive medical services for prices that are at least half of what the procedure would cost in the U.S. This emerging market seems to be beneficial to all involved except U.S. healthcare providers; however, this outsourcing of healthcare services sends a powerful international message. It seems that the U.S. has a healthcare system that cannot adequately serve all economic classes of the American public. In contrast, though India has the proper facilities and professionals, there are concerns regarding malpractice litigation, postoperative care, and possible negative effects on the Indian public. Having given consideration to all affected constituencies, it seems that the outsourcing of medical procedures is in the best interest of lower- and middle-class Americans as well as medical professionals in India. In reality, though medical tourism is receiving much attention, it will most likely not be a pressing concern for the American market in the near future. A widening discrepancy in the Indian public may, however, be cause for nearer concern. This new trend does foreshadow a push for more preventative changes in the business of U.S. healthcare, such as the development of information technology specific to the growing international healthcare market. Whereas, it will initially be beneficial to send patients abroad, with the evolution of technology, the latter ideal will instead be to have medical professionals abroad that care for patients located in the U.S.

INTRODUCTION

The cost of healthcare in the United States seems to be approaching a level that is beyond the economic means of the general public. Even with insurance, expensive surgeries sometimes require thousands of dollars in out-of-pocket costs. Americans have found the cheaper alternative of seeking less expensive healthcare in a foreign country. This practice, which is often coupled with sightseeing and actual tourism, is appropriately referred to as medical tourism. Since there are plenty of advantages for both Americans and their foreign counterparts, medical tourism seems like the natural solution in the short run, and has the potential to become a rapidly growing market. However, when taking all constituencies into consideration, in the long run, this growing market could have some indirect, unfavorable side effects.

ENCOURAGEMENT OF FOREIGN PROCEDURES

Perhaps the largest instigators in this recent trend are the employers. They could be playing a large role in the development of medical tourism simply by introducing the idea to their employees and, in some cases, even offering some economic incentive. In an effort to save on insurance fees, several companies have begun to promote foreign healthcare options.

Blue Ridge Paper Products, based in Canton, North Carolina, is one such venturous company. Their healthcare claims were initially projected to be $36 million by 2006; however, due to these foreign alternatives, their actual claims in 2006 were closer to $24 million (U.S. Senate Hearing, 2006). These savings will be to the advantage of the company and its employees because they will be able to internalize more revenue as well as keep wages reasonable.

Carl Garrett is a technician at a Blue Ridge paper mill who had plans to receive two medical operations in New Delhi, India. The 60-year-old needed both his gall bladder removed and his rotator cuff mended, and was delighted at the opportunity to avoid paying $10,000 in deductibles and out-of-pocket fees (Milne-Tyte, 2006). The cost for Blue Ridge was so low in comparison to the charge that the company would have incurred from U.S. medical fees that they actually offered to return a portion of their savings to Garrett.

In another case, Howard Stabb, a successful business owner, sought physicians in India rather than in the United States. Stabb had chosen not to have health insurance and found that he could save over $150,000 by receiving heart surgery out of the country. He brought a patient advocate with him, and both of them agreed that his decision to seek services abroad was best for both his personal well-being and for the financial stature of his company (U.S. Senate Hearing, 2006).

THE PREFERENCE TOWARDS INDIA

One might wonder, of all the third-world countries where one might be able to receive care, why would India be a good choice? Why would the global leader in importing foreign patients be half-way around the world? Though a possible reason could be coincidence, as in, Indians were the first to create such a market, there are more likely, tangible reasons for this phenomenon. For instance, there is less of a language barrier in India as opposed to some other possible foreign destinations. Most Indians begin learning English in school at a young age, so odds are that the healthcare professionals would be able to communicate with an American in English. Also, helping to make their foreign patients feel secure with the care that they will be receiving, many Indian physicians have been trained in the west (Rai, 2006). Much American hesitation to travel abroad for care comes from uncertainty about the quality of care and knowledge and experience of the foreign physicians. Thus, knowing that an Indian physician has an American medical education would make the possibility more appealing.

Not only might the qualifications of the physicians testify to the care that they can provide, but the experiences of former patients can also be a strong point of persuasion. If individuals that are considering receiving care abroad know someone that can give a recommendation based on experience, they are much more likely to do it themselves. With the volume of medical tourism in India, half a million patients per year and growing, they have a large constituency that can vouch for the quality of care provided in the country (Hutchinson, 2005). Maggie Ann Grace, patient advocate for Howard Stabb, has stayed with patients both in India and the United States. She noted in a statement to the U.S. Senate that she, "would sooner leave her loved ones in the care of doctors and nurses in the Indian Hospital" (U.S. Senate, 2006). She attributed this opinion mostly to the fact that nurses in the U.S. are too busy and because of this, care is given based on the level of the patient's severity and the nurse's schedule.

Cost is also to India's advantage, not only in comparison to U.S. prices, but to other countries as well. In a 2005 comparison of estimated fees for a coronary artery bypass graft surgery, India's more expensive hospital was still $15,000 less expensive than a Mexican hospital and $5,000 cheaper than Thailand (U.S. Senate, 2005). Thus, for reasons that range from financial to human interaction, India is a strong competitor in the international market for medical tourism.

THE AMERICAN PERSPECTIVE

American disdain for the healthcare system has the potentional of becoming a detriment to U.S. society. The situations of both Carl Garrett and Stabb seem to favor medical tourism. It is a favorable option for the patient and their employer, but are there possible adverse affects? On a larger scale, considering all constituencies both directly and indirectly involved, there is a variation in those that benefit from this particular form of outsourcing. For the growing constituency that benefits from and is in favor of medical tourism, there is an opposing set of constituencies that will find this disadvantageous. From 2002 to 2005, there was an increase of 350,000 foreigners that traveled to India for care, this number is projected to increase by 30% each year (Hutchinson, 2005). If this pattern of receiving healthcare abroad does continue to progress, it does not present a sustainable future for the present state of our domestic healthcare system.

Howard Stabb was operated on in India, stayed for a month, and returned with a positive impression of the care provided. His cardiologist in the States reported that he was healthy and that the operation had been successful (U.S. Senate, 2006). That was the best option for Stabb personally, but on a societal level, Stabb's, advertisement for medical tourism may have a negative impact.

Carl Garrett received his operations in the U.S. instead of in India as intended, due to strong opposition. Garrett is part of the United Steelworkers Union, who saw this growing trend toward medical tourism as an excuse to lower the quality of care and a denial of the right to safe healthcare. Stan Johnson, a speaker for the Union, stated that outsourcing healthcare would lead to corporate profiteering and that if we allow traveling abroad for treatment to remain a feasible option, it may one day become the only option (Milne-Tyte, 2006).

Bruce Cunningham, president of the American Society of Plastic Surgeons, presented other concerns, while providing what would most likely be the shared opinion of many American physicians on this subject. He could understand the desire to receive healthcare at a lower cost, but warned of the potential to increase risks (U.S. Senate, 2005). Travel done shortly after having surgery can be risky. Also, many procedures require follow-up care; if an American provider will be administering postoperative care, they should be found and agree to provide care before the patient leaves the country.

There was no evidence found that would justify American physicians having a fear of losing revenue. On the contrary, American physicians and hospitals may begin to feel some reprieve from a diminished patient load. This seems to be the optimistic view of UK physicians. In 2002, it was the National Health Service (NHS) of the UK that began to encourage citizens to seek care abroad in order to alleviate long wait lists (Maini, 2005). Also, patients that would not be eligible for a procedure under NHS would travel abroad to receive it. Thus, it seems that for the present time, such outsourcing is beneficial. For long-term concerns, there is comfort in the fact that for many procedures, proximity is essential. For instance, childbirth and emergency care are not services that could withstand the 20-hour flight.

IndUSHealth, one of many companies that coordinate visits for medical tourists, caters to interested Americans by providing direct information and offering assistance with planning the trip and procedure. Their Web site will most likely convince Americans with limited financial resources that this is their best option for receiving medical care. The company assigns each patient a personal case manager that handles the details of their visit (IndUSHealth site, 2005). Their Web site is subjective; it sheds a positive light on all aspects of receiving care in India. Interested patients should investigate their own reasons for going and whether or not they find any ethical bias against it. Overall, the American perspective is torn. Not necessarily a question of quality vs. cost, there are several factors and variables to consider.

THE INDIAN PERSPECTIVE

The healthcare system in India is complex. The healthcare facilities are divided first into two categories, government and private, and under each of those two categories are three main branches: tertiary, secondary, and primary (the lowest level of care) (Van Hollen, 2003). The government institutions are generally considered inferior to the private due mostly to overcrowding, less funding, and less expensive care. As one goes further from the city, the medical facilities become less technologically advanced, have minimal staff, and lowered sanitation standards. The divide between the private and public sector is growing. In fact, the spending on public healthcare is the sixth lowest in the world. On the other hand, India is among the top 20 nations for spending in the private sector (Nundy & Sengupta, 2005). This is due to spending on healthcare being a lower priority. To compensate, the government encourages growth in the private sector. The government uses subsidies and tax exemptions as some methods of encouraging the private sector. This increasing divide between public and private healthcare factors into not only the care provided, but also the providers of the care. Healthcare professionals prefer working in the private sector due to higher wages and a more pleasant working environment. It is this lack of interest in public sector employment that has propelled entrepreneurial endeavors such as venues for medical tourism, because the Indians involved in the development and implementation of these endeavors do benefit from them.

Unfortunately, there does also seem to be negative repercussions for Indian patients both in the private and public sectors. It seems probable that their care and access to the best physicians might be compromised by the influx of foreigners with more elastic checkbooks. One large downfall of the private sector is that it is not regulated so there are no standards of quality or costs (Nundy & Sengupta, 2005). Thus, their success is measured by profit rather than the actual care being provided. Unfortunately, the private sector also tends to pull the focus away from the public health situation. Since India has plenty of internal health concerns such as widespread diseases and rural communities with minimal healthcare facilities, it seems that a greater focus should be put on the public sector. The problem is that the attraction to the private system is prevalent both to the care providers and patients. In such a profit driven situation, to make this shift from privately to publicly focused is challenging.

COMPARATIVE DATA

All Indian hospitals involved are approved by the WHO supported Joint Commission International. The mission of the Joint Commission International is to improve the safety and quality of care in the international community. Many contributing Indian hospitals are part of a private national chain; for instance, the Apollo Hospitals in Chennai and New Delhi and the Wockhardt Hospitals, one of which is in Mumbai. There is also a chain in Delhi, Fortis Healthcare.

Just as the market for medical tourism has grown in India due to enticing profit margins, reciprocally, American consumers are similarly convinced by the savings. All of these options have a major financial advantage over U.S. providers.

Americans also benefit from some insurance relief since some insurance companies that traditionally only covered emergency, out of country procedures have evolved to a coverage plan for non-emergency situations. BCBSAZ, Cigna, Humana, and Aetna are examples of providers that support the concept of receiving healthcare abroad (personal communication, March 3, 2007). Aetna would only be willing to cover a non-emergency if it is proven to be medically necessary. Perhaps the growing popularity of coverage for procedures conducted abroad has a correlation to the rising insurance premiums. According to the Kaiser Family Foundation, insurance premiums have risen 87% in the last 6 years (Galles, 2007). In 2006, the average annual premium for coverage for a family of four was $11,500 (Fitzpatrick, 2004). That means that the premium was greater than the gross earnings for a full-time minimum wage worker. It would be economically impossible for them have insurance, which explains why 43 million people in the U.S. do not have health insurance (Hutchinson, 2005). This trend of forcibly keeping people from having insurance seems to be a force that is perpetuating the outsourcing of medical care.

THE MALPRACTICE ISSUE

Speaker of the Union, Stan Johnson, expressed an additional concern that most would not consider, having already been strongly convinced by the savings that receiving care abroad affords. However, consumers should consider the possibility of a malpractice lawsuit and who, if any, would be the responsible party. In accordance with their agenda, IndUSHealth has a reasoned response to concerns about malpractice litigation. They claim that India has a similar court system to the U.S. and American patients have the right to file cases there. In reality, however, it seems that though American patients would have the right to file a case in India, the proceedings of a lawsuit would differ from the American process. The major difference between the systems of the two countries is that in the U.S., such a case would be tried in a state court, while in India there are more specific consumer courts. The trend in U.S. courts is to present the patient as victimized and almost guilt the jury into awarding the patient a large sum of the rich doctor’s income. In India, on the other hand, there is a less-biased perspective. Awards are given based on damage; lawyers are not permitted to accept cases that may have facetious intentions. Culturally, doctors in India are so highly respected that most patients would not challenge the decision of their doctor, though there is a possibility that India’s system may be swaying toward the favor of the consumer. In 1992, the Kerala High Court proposed an extension to the Consumer Protection Act to include malpractice and negligence. Due to opposition from physicians, the proposal was reviewed by the supreme court but, to the favor of consumers, was upheld. It was not only upheld in Kerala, but courts were established in all states to handle the complaints of consumers (Studdert, Mello, & Brennan, 2004).

However, there has not been much change toward the American mentality of a more persuasive litigation process. This is dually reflected in the contrasting price of malpractice insurance. According to the Pacific Research Institute, Indian doctors pay $4,000 per year for insurance while American doctors have been known to pay 25 times that amount (Brunell, 2007).

Although this may seem to correlate with the view of foreign malpractice laws as being too relaxed, within both healthcare systems there is a problem of keeping record of errors and any resulting injuries. There is currently an emerging program in the U.S. to hold doctors and hospitals accountable for recording errors (Studdert, et al., 2004). No such program has been initiated in India, but perhaps that will change with an increased influx of foreign patients. A responsible system to track liability would, in turn, keep malpractice suits in both countries honest and less biased.

Overall, even if an American patient were willing to chance a system that had no tendency to be in the consumers' favor, considering the financial aspect of such a process may cause them to reconsider. Financially, there is the added cost of traveling back and forth, as well as lawyer's fees. Also, if compensation is awarded, it would be a much smaller amount than one would expect in the U.S. (Milne-Tyte, 2006).

TECHNOLOGICAL INNOVATIONS

Consider a more cooperative approach to cross-cultural care, where liability is spread amongst providers in two separate countries; there are several technicalities that tend to hamper innovation in healthcare. However, new developments may also serve as solutions to greater struggles. It seems that advances in technology will correlate with the popularity of medical tourism.

Telesurgery, a specific branch of telehealth, provides the opportunity for surgeons that are physically located in different locations to communicate through operative videoconferencing. One surgeon observes the procedure through a camera and then offers visual and auditory feedback to the surgeon at the operating site (Cheah, Lee, Lenzi, & Goh, 2000). Among the more developed teleradiology, telepathology, and teleoncology, telesurgery seems to be the branch of telecommunications that is directly applicable to medical tourism. International telesurgery proved to be successful in two laparoscopic cases between the United States and Singapore. More specifically, an experienced U.S. surgeon observed a less experienced surgeon perform both a radical nephrectomy and a varicocelectomy (Lee, Liew, Fabrizio, Li, Jarrett, & Kavoussi, 2000). This practice would be beneficial to the medical tourism market both by alleviating the distrust of foreign physicians, as well as keeping cost minimal by still undergoing the procedure in a foreign institution. A more recent 2003 study of procedures conducted between the United States and Brazil noted a feasible future for telesurgery. In two cases of telepresence surgery, a U.S. surgeon directed a robot attached to a laparascope in a bilateral variococelectomy, and the other robot was used for needle placement in a percutaneous nephrolithotomy (Rorigues, Mitre, Lima, & Fugita, 2003).

However, there are barriers to implementing this technology, the language divide, for one. One would assume the physicians in this example corresponded in English, but it cannot always be assumed that that will be feasible. Additionally, the cost of higher bandwidth communication lines is still expensive, which makes it difficult for such technology to access foreign countries (Lee, &

Png, 2000). The financial emphasis on healthcare is truly an overpowering barrier, which explains why the success of telehealth relies on the private sector. It seems that an established relationship with a university that is willing to do research and provide resources, as well as a partnership with several private healthcare organizations to incorporate, is key to establishing a successful telehealth program. Not only is a structural foundation necessary, but a substantial need for such an innovation is also required. The success of the UltraClinics telehealth solution to mammogram screening was due to the yearly demand of over 48 million mammograms and the one million additional that require biopsy (Weinstein, et al., 2007). As of now, though there are crowded emergency rooms and stymied appointment schedules in the U.S. and UK, the demand is not overpowering, thus, there has been no push for the combination of medical tourism and telesurgery.

Beyond financial barriers it seems that in the future, when demand as well as technology has increased, this may be a strong asset to the growing medical tourism market. Telesurgey supplements both problems of communication and postoperative care. Comprehensive hospital information systems, such as the system implemented at the U.S. Veterans' Affairs institutions, have initiated electronic health records. With the correct considerations of patient privacy, the ability to transfer records with such technological ease will improve communication and understanding across country borders. Based on the UltraClinics example, increased use of technology can keep the patient more informed, accelerate their care, and allow more time for the physician to care pre- and postoperatively (Weinstein, et al., 2007). There are more benefits than detriments to the globalization of healthcare, but it will require a more cooperative attitude toward sharing advances in research, techniques, and technologies.

CONCLUSION

Medical tourism will most likely continue to grow in popularity. Though it is negative propaganda for the arguably outrageous cost of healthcare in the U.S., it is helping to ease some strain on the American system. Essentially, American physicians are keeping most of the patients that can afford the procedure, losing middle-income patients that opt to save money abroad, but still have the burden of the low-income population that cannot afford the less-expensive option. This seems to prove that the medical environment in the U.S. is now driven by money rather than the desire to provide quality care to each patient. However, the money lost (see Table 8.1) to foreign markets seems to have little effect on the U.S. healthcare market. As long as there are overcrowded emergency rooms and the need to schedule doctor's visits months in advance, like the UK, the U.S. will probably consider this new trend to be beneficial. The fact that the market in India has a 30% consumer increase yearly is significant, but not enough to severely damage the American market in the near future. If we do come to a breaking point where benefit becomes burden, the U.S. healthcare system would be forced

Table 8.1. (Chart: U.S. Senate, 2005)

Procedure	**U.S. Cost (appr)**	**Indian Cost**	**Airfare (appr)**	**Money Saved**
Heart Bypass	$70,000	$6,000	$2,000	$62,000
Angioplasty	$41,000	$6,000	$2,000	$33,000
Hip Replacement	$37,000	$5,000	$2,000	$30,000
Spinal Fusion	$55,000	$8,000	$2,000	$47,000

to make some changes, such as the compromise of telesurgery, that would attract the middle-class constituency and minimize the outsourcing of medical services.

No major preventative changes have been made as of yet, due mostly to one major barrier; policymakers and government officials cannot agree on the best way to control costs. Some possible solutions would be implementing price controls or strict budgets (Fitzpatrick, 2004), although it may be time to consider a system that provides healthcare to every American citizen. With millions of uninsured citizens, people will avoid seeking care until it is necessary. The U.S. still spends more on healthcare than any other industrialized nation. As with many shifts toward outsourcing, the need for change originated in the home country, which led to the need to seek out supplementation, not substitution, in foreign markets.

FUTURE RESEARCH

There is still much to be determined about the future of medical tourism and economic solutions to healthcare, domestically. Though price difference and the savings to be had abroad are well advertised, there are few documented patient experiences. It would be beneficial, both for prospective patients and providers, to have a well-rounded description of the medical tourism industry. Inclusive in such a description should be what drove each patient to seek care abroad, their experience throughout the process, whether they benefited overall from the savings and what recommendations they would give to prospective patients. Such documentation should also include reasoning behind choosing the country in which care was given, it seems that regional specificity may emerge along with this growing trend (i.e., for cardiac surgery one would seek care in Mexico, while facilities in India would specialize in dentistry). Realistically, it would be to the advantage of foreign providers to specify their fields of care, in the effort to heighten the quality of care that they can provide. Though this may reduce some of the competition in the market, it will increase the appeal to American consumers who are wary about the quality of care provided abroad.

Telesurgery, an innovative technology that may prove to be a solution to several problems, has already been successful in a few operations. However, there will have to be a great deal more data collected to ensure the usefulness of this innovation. Additionally, the cost of such a procedure would have to be reduced before it would be willingly adopted into the market. Overall, since this market is still in its fledgling stages, there is little data from which concrete conclusions can be made, making further research essential to future progress.

REFERENCES

Brunell, D. (2007). Medical tourism creates global competition. *The Columbian*. Retrieved April 21, 2007, from http://www.columbian.com

Cheah, W., Lee, B., Lenzi, J., & Goh, P. (2000). Telesurgical laproscopic cholecystectomy between two countries. *Surgical Endoscopy, 14*(11), 1085.

Fitzpatrick, C. (2004). *Facts on cost of health care*. Retrieved March 11, 2007, from http://www.nchc.org/facts/cost.shtml

Galles, J. (2007). Unsustainable health care premium increases. *Greater Charlotte biz*. Retrieved March 11, 2007, from http://www.greatercharlottebiz.com/article.asp/id=62

Hutchinson, B. (2005). Medical tourism growing worldwide. *UDaily*. Retrieved April 25, 2007, from http://www.udel.edu

Indushealth Web site. (2005). Retrieved April 24, 2007 http://www. indushealth. com

Lee, B., Png, D., Liew L., Fabrizio, M., Li, M., Jarrett, J., & Kavoussi, L. (2000). Laparoscopic telesurgery between the United States and Singapore. *Annals Academy of Medicine Singapore*, *29*(5), 665-668.

Maini, A. (2005). India: Untapped potential in the medical tourism market. *Healthcare Management*, Issue dtd. Janurary 15-31.

Milne-Tyte, A. (2006). Medical tourism meets healthy opposition. *American Public Media*. Retrieved September 13, 2006, from http://marketplace.publicradio.org/shows/2006/09/13/PM200609137.htm

Milstein, A., & Smith, M. (2006). America's new refugees seeking affordable surgery offshore. *The New England Journal of Medicine*, *355*, 1637-1640.

Nundy, S., & Sengupta, A. (2005). The private health sector in India. *BMJ Journal, 331*, 1157-1158.

Rai, S. (2006). Union disrupts plan to send ailing workers to India for cheaper medical care. *The New York Times,* October 11. Retrieved February 13, 2007, from http://www.boston.com/yourlife/health/other/articles/2006/10/11/

Rodrigues Netto Jr., N., Mitre, A., Lima, S., & Fugita, O. (2003). Telementoring between Brazil and the United States: Initial experience. *Journal of Endourol.*, *17*(4), 217-20.

Studdert, D., Mello, M., & Brennan,T. (2004). Medical malpractice. *New England Journal of Medicine*, *350*, 283-92.

U.S. Senate Hearing. (2006). *The globalization of health care*. Retrieved March 12, 2007, from http://www.medicaltourisminsight.com/articles/0701-senate.htm

Van Hollen, C. (2003). *Birth on the threshold: Childbirth and modernity in South India*. Los Angeles: University of California Press.

We Care Health Services. (N.D.). *The comparative costs between India and other developed countries like U.S., UK, and Singapore approximate figures in U.S. dollars*. Retrieved February 15, 2007, from http://content.nejm.org/cgi/content/full/355/16/1637

Weinstein, R., López, A., Barker, G., Krupinski, E., Descour, M., Scott, K., et al. (2007). The innovative bundling of teleradiology, telepathology, and teleoncology services. *IBM Systems Journal*, *46*(1), 69-84.

ADDITIONAL READINGS

Bradley, W. G., (2004). Offshore teleradiology. *Journal of the American College of Radiology, 6*, 244-248.

Firth-Cozens, J., & Cording H. (2004). What matters more in patient care? Giving doctors shorter hours of work or a good night's sleep? *Quality and Safety in Healthcare, 13*, 165-166.

Fontanarosa, P. B., Rennie, D., & DeAngelis, G. D. (2004). Postmarketing surveillance: Lack of vigilance, lack of trust. *Journal of the American Medical Association, 292*, 2647-2650.

Franken, E. A., Berbaum, K.S., Brandser, E.A., D'Alessandro, M.P., Schweiger, G.D., & Smith, W.L. (1997). Pediatric radiology at a rural hospital: Value of teleradiology and subspeciality consultation. *American Journal of Roentgenology, 168*, 1349-1352.

Graschew, G., Roelofs, T. A., Rakowsky, S., & Schlag, P. M. (2006). E-health and telemedicine: Digital medicine in the virtual hospital of the future. *International Journal of Computer Assisted Radiology and Surgery, 1*, 119-135.

Halliday, B. E., Bhattacharyya, A. K., Graham, A. R., Davis, K. R., Leavitt, S. A., Nagle, R. B., et al. (1997). Diagnostic accuracy of an international static imaging telepathology consultation service. *Human Pathology, 28*(1), 17-21.

Herzlinger, R. E. (2006). Why innovation in healthcare is so hard. *Harvard Business Review 84,* 5, 55-66.

Kayser, K., Szymas, J., & Weinstein, R. S. (2005). *Telepathology and telemedicine—Communication, electronic education and publication in e-health*. Berlin: VSV Interdisciplinary Medical Publishing.

Olszak, A. G., & Descour, M. R. (2005). Microscopy in multiples. *IEEE oe magazine 5*(5), 16-18.

Weinstein, R. S., Bhattacharyya, A., Halliday, B. E., Yu, Y-P, Davis, J. R., Byers, J. M., et al. (1995). Pathology consultation services via the Arizona-International Telemedicine Network. *Archives d'Anatomie et de Cytologie Pathologiques, 43,* (4), 219-226.

Wysocki, W. M., Komorowski, A. L., & Aapro, M. S. (2005). The new dimension of oncology: Teleoncology Ante Portas. *Critical Reviews in Oncology/Hematology, 53*(2), 95-100.

This work was previously published in Outsourcing and Offshoring of Professional Services: Business Optimization in a Global Economy, edited by A. Gupta, pp. 157-166, copyright 2008 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 8.11
Towards Cognitive Machines:
Multiscale Measures and Analysis

Witold Kinsner
University of Manitoba, Canada

ABSTRACT

Numerous attempts are being made to develop machines that could act not only autonomously, but also in an increasingly intelligent and cognitive manner. Such cognitive machines ought to be aware of their environments, which include not only other machines, but also human beings. Such machines ought to understand the meaning of information in more human-like ways by grounding knowledge in the physical world and in the machines' own goals. The motivation for developing such machines range from self-evidenced practical reasons such as the expense of computer maintenance, to wearable computing in health care, and gaining a better understanding of the cognitive capabilities of the human brain. To achieve such an ambitious goal requires solutions to many problems, ranging from human perception, attention, concept creation, cognition, consciousness, executive processes guided by emotions and value, and symbiotic conversational human-machine interactions. An important component of this cognitive machine research includes multiscale measures and analysis. This article presents definitions of cognitive machines, representations of processes, as well as their measurements, measures, and analysis. It provides examples from current research, including cognitive radio, cognitive radar, and cognitive monitors.

INTRODUCTION

Computer science and computer engineering have contributed to many shifts in technological and computing paradigms. For example, we have seen shifts (1) from large batch computers to personal and embedded real-time computers, (2) from control-driven microprocessors to data- and demand-driven processors, (3) from uniprocessors to multiple-processors (loosely coupled) and multiprocessors (tightly coupled), (4) from data-path processors to structural processors, for example, neural networks (Bishop, 1995), quantum processors (Nielsen & Chuang, 2000) and biomolecular processors (Sanz, Chrisley & Aaron Sloman, 2003), (5) from silicon-based processors to biochips (Ruaro, Bonifazi, & Torre, 2005), (6) from vacuum tubes to transistors to microelectronics to nanotechnology, (7) from large passive sensors to very small smart active sensors (Soloman, 1999), (8) from local computing to distributed computing and networkwide computing, (9) from traditional videoconferencing to telepresence (e.g., WearTel and EyeTap; Mann, 2002), (10) from machines that require attention (like a palmtop or a wristwatch computer) to those that have a constant online connectivity that drops below the conscious level of awareness of users (like autonomic computers; Ganek & Corbi, 2003; IBM, 2006), and eyeglass-based systems (Haykin & Kosko, 2001; Mann, 2002), (11) from crisp-logic-based computers to fuzzy or neurofuzzy computers (Pedrycz & Gomide, 1998), as well as (12) from control-driven (imperative) systems to cognitive systems such as cognitive radio (Haykin, 2005a), cognitive radar (Haykin, 2006), active audition (Haykin & Chen, 2005), and cognitive robots. These remarkable shifts have been necessitated by the system complexity which now exceeds our ability to maintain them (Ganek & Corbi, 2003), while being facilitated by new developments in technology, intelligent signal processing, and machine learning (Haykin, Principe, Sejnowski, & McWhirter, 2006).

Since the 1950s, philosophers, mathematicians, physicists, cognitive scientists, neuroscientists, computer scientists, and computer engineers have debated the question of what could constitute *digital sentience* (i.e., the ability to feel or perceive in the absence of thought and inner speech), as well as machine consciousness or artificial consciousness (e.g., Cotterill, 1988; Dennett, 1991; Klivington, 1989; Kurzweil, 1990; Rumelhart & McClelland, 1986; Minsky, 1986; Neumann, 1958; Posner, 1989; Searle, 1980; Penrose, 1989, 1994; Searle, 1992). Consequently, many approaches have been developed to modeling consciousness, including biological, neurological, and engineering (practical). The approach to cognition taken in this article is mostly engineering in which the behavior of a system can be observed, measured, characterized, modeled, and implemented as an artifact, such as a cognitive robot (either isolated or societal) to improve its interaction with people, or a cognitive radio to improve the utilization of a precious resource, that is, the frequency spectrum. In general, the intent of such cognitive systems is to improve their performance, to reduce waste in resource utilization, and to provide a testbed for learning about cognition and cognitive processes. If an approach is purely reductionist, it may not be capable of describing the complexities of cognition. Since our engineering approach considers not only the individual components of a system, but also their interactions, it may be capable of describing the dynamics of cognitive processes. Although engineering approaches have serious limitations (e.g., Chalmers, 1997; Parsell, 2005), they are intended to produce a range of specific practical outcomes.

The next section provides a few definitions and models of consciousness, and serves as a preamble for several applications. The article also identifies a few problems with the current status of cognitive machines.

What Is Cognition?

According to the Oxford Dictionary, cognition is "knowing, perceiving, or conceiving as an act." The Encyclopedia of Computer Science (Ralson, Reilly, & Hemmendinger, 2003) provides a computational point of view of cognition consisting of the following three attributes: (1) cognition can be described by *mental states and processes* (MSPs) intervening between input stimuli and output responses, (2) the MSPs can be described by algorithms, and (3) the MSPs should lend themselves to scientific investigations.

Another view of cognition is suggested by Pfeifer and Scheier (1999) as an interdisciplinary study of the general principles of intelligence through a synthetic methodology termed *learning by understanding.*

Cognition also includes language and communications, as studied in different forms (e.g., Fischler & Firschein, 1987; Roy, 2005; Roy &

Pentland, 2002). Although the language and communications of humans and machines differ significantly, their roles are similar. More specifically, Haikonen (2003) defines cognition as the association of auxiliary meanings with percepts, the use of percepts as symbols, the manipulation of these symbols, reasoning, response generation, and language.

COGNITIVE MACHINE MODELS

Haikonen Model

Pentti Haikonen of the cognitive technology at Nokia in Helsinki has advanced the concept of cognitive machines in the context of the mind-body problem. He sees emulation of the cognitive process of the brain through a cognitive architecture (consisting of the flow of perception, inner speech, inner imaginary, and emotions), rather than the classical architecture involving rule-based artificial intelligence and neural networks. His architecture does not involve rule-based computing. To explain the scope of the architecture, Haikonen gives a good illustrative example of five classes of behavioral machines (mobile robots) that have external environmental sensors to monitor the environment, and internal self-sensors to monitor the internal states of the machine itself:

1. *Simple-Reflex Machine.* In response to an obstacle, the robot backs off and turns a little so that it could avoid the obstacle. This behavior is driven by *recognition* (i.e., sensing, characterization, and classification) alone, and its operation requires no memory. Although the robot is "aware" of the obstacle (by detecting it) and responds to it, this behavior is too primitive to call it a conscious act.
2. *Simple-Reflex Machine with Smart Memory.* If memory and some processing capabilities are added, the robot not only can sense the obstacles, but also can record their locations and the responses to each obstacle, and then can compute the "best" path by deleting any loops in the path, using known optimization algorithms. Thus, its "awareness" has been expanded from the encounter with the obstacles to the layout of the maze traversed. However, this behavior is still too simple to call it consciousness.
3. *Machine with Meaningful Perception and Associative Memory.* If an associative memory (Hinton & Anderson, 1981; Kohonen, 1987) is added to the previous robot, the robot cannot only sense, but can also perceive its environment, and can learn from the experiences encountered so far. Furthermore, if a *cost function* is established within the robot (i.e., a function capable of discriminating between a "pain" for dangerous obstacles, and a "pleasure" for useful objects) then the "seen" obstacle can evoke "images" of its past encounters, and a response can be produced even before the actual contact with the object. This *perception* goes beyond recognition in that it is an active process of searching and finding threats and opportunities through evoked "imagery" of the object and associated actions. In a way, such a machine exhibits an *attention* process, seeking "satisfaction," as controlled by a good-bad discrimination criterion and its own needs.
4. *Machine Associative Memory and Reporting.* If representational symbols (words) and a corresponding symbol system in the form of a language (e.g., Chomsky, 2006; Dawkins, 1990) are established within the robot, then the robot not only can exhibit the behavior just described, but also can report (declare) on its perception and actions during and after the perceived events, as well as on its inner self-sensed percepts. The robot can

also associate meaning with the declaration of its peers.

5. *Machine with Self Awareness and a "Mind."* The previous two classes of behavior were based on an "open-loop" control process. If feedback is added to the flow of percepts, then the robot could perceive its own declarations and understand them silently, without the need to act upon them. This would allow the robot to distinguish between the percepts caused by the external environment and those caused internally. The "inner images" could lead to "imagining." If another good-bad criterion could be developed and applied to the "inner images," then the robot could develop a "mental content" or a form of a "mind."

In summary, the Haikonen cognitive machine architecture involves (1) sensory preprocessing circuits to derive representations of the sensed external events, (2) introspective feedback loops that discriminate between the external representations, and broadcast the outcomes to other loops, (3) associative cross-coupling of those loops, and (4) attention control through adaptive thresholding.

Such a machine requires: (1) distributed signal representations, (2) associative processing and learning, (3) perception processes, (4) sensory attention and inner attention, (5) the flow of inner "speech" and "imagery," as well as (6) evaluation of the significance (relevance) of reasoning, motivation, and action. Thus, consciousness might emerge spontaneously in such a complex system. A demonstration of a low-complexity system exhibited emotions, but not consciousness (Haikonen, 2004).

Similar emergent consciousness in autonomous agents is described by Freeman (2001) and Cotterill (2003). Both approaches are consistent with Grossberg's view of the mind and brain in which new states are being synthesized continuously to form a better adaptive relationship with the environment (Grossberg, 1982) and with Grossberg's adaptive resonance theory (ART) (Grossberg, 1988), as well as with the dynamic system approach to consciousness and action (Thelen & Smith, 2002).

A major criticism of Haikonen's model is that there is no systematic approach to measuring the basic system reactions or emotions or perception of time (Parsell, 2005). This article addresses a part of this problem.

Franklin's Model

Functional consciousness of an autonomous agent (in the form of an intelligent distribution agent, IDA) was described by Stan Franklin from the Conscious Software Research Group at the University of Memphis (Franklin, 1995, 2003). An IDA can exhibit several functions of consciousness as elicited by Bernard Baars under his *global workspace theory* (Baars, 1988). An implementation of the concept includes a system to communicate with U.S. Navy sailors in a natural language.

Aleksander's WISARD and MAGNUS

Igor Aleksander of the Neural Systems Engineering group at the Department of Electrical and Electronic Engineering at the Imperial College of Science, Technology and Medicine in London has an extensive background in neurocomputing, including parsimonious neural networks (Aleksander, 1989). He has been working on a neural pattern recognition system called WISARD (Aleksander, 1998) and a neural representation modeler called MAGNUS (Aleksander, 1998, 2003) and the artificial neural consciousness system (Aleksander, 2003, 2006). His model adds the ability to predict (anticipate) foreseeable events. Aleksander considers *prediction* as the key feature of consciousness.

Taylor's CODAM Model

John Taylor of the Department of Mathematics at King's College in London has developed many models of perception and consciousness, with focus on attention as the prerequisite for consciousness. His thesis is that the application of engineering control theory to attention could show how consciousness is created. This view is motivated by his extensive knowledge of experimental data of visual illusions, memory, attention, and motor control through a number of imaging techniques. One of the techniques is the *magnetoencephalography* (MEG), capable of detecting the very small (femtoTesla) magnetic fields generated by the brain. MEG produces a fine temporal resolution (milliseconds), but poor spatial resolution (millimetres). The problem of spatial resolution can be resolved by the functional *magnetic resonance imaging* (fMRI) and *positron emission tomography* (PET) whose temporal resolution is poor (tens of seconds), but spatial resolution is very good. His studies identified the parietal lobe as an important brain region in the creation of biological consciousness. This has lead to the development of the *corollary discharge of attention movement* (CODAM) model (Taylor, 2001, 2002, 2003). This model has been used to study mental diseases such as schizophrenia and autism.

Over 15 other models amenable for implementations have been described at a Workshop on Models of Consciousness (Sanz et al., 2003). A brief review of the state of cognitive systems engineering is presented in Hoffman, Klein, and Laughry (2002).

EXAMPLES OF COGNITIVE SYSTEMS

The models described in the previous sections are designed to provide insight into the development of cognition and consciousness. Practical applications may not require the complexities involved. To this end, Simon Haykin of the Adaptive Systems Laboratory at McMaster University in Hamilton has defined (Haykin, 2005b) a cognitive machine as an intelligent system that (1) is aware of its surrounding environment (i.e., outside world), (2) uses understanding-by-building to learn through interactions with the environment and other means, and (3) adapts its internal states to statistical variations in input stimuli by making corresponding changes in adjustable system parameters in real-time with specific objectives (e.g., reliability, efficiency, active sensing) as determined by the application of interest. The essential attributes of such a machine includes awareness, intelligence, learning, adaptivity with temporal and structural reconfigurability, action, and real-time operation. This characterization of cognitive machines is mostly behavioral, and leads to systems realizable with today's technologies.

Cognitive Radio

The electromagnetic radio spectrum can be considered a precious natural resource that should be utilized efficiently. Some frequency bands are used heavily, while others are underutilized. The utilization of the bands also changes with location and time. By being aware of the actual utilization of the bands, and by dynamic spectrum management (i.e., through frequency allocation and adjustment of the transmitted power), this resource could be used more efficiently.

Cognitive radio based on the software-defined radio (Dillinger, Madani & Alonistioti, 2003) can achieve this goal. A comprehensive description of such a cognitive radio system is provided by Haykin (2005a).

Cognitive Radar

Similar to cognitive radio, cognitive radar is aware of its environment, utilizes intelligent signal processing, provides feedback from the receiver to the transmitter for adaptive illumination based on the range and velocity, and preserves the information contents of radar returns. A network of such systems can also be set up to collaborate with one another as multifunction radars and noncoherent radars. A description of such a system is provided by Haykin (2006).

Active Audition

Active audition involves sound localization, segregation of the target source from interference, tracking of the source, and learning from the experience to adapt to the changing environment (Haykin & Chen, 2005). This problem has many applications, including improved devices for hearing-impaired persons.

Meaning Machines and Affective Computing

Deb Roy of the Cognitive Machines Group at the MIT Media Laboratory has been focusing on human-machine language development for conversational robots (Roy, 2005). Others from that group have been developing machine recognition of affective (emotional) states and their synthesis in order to accelerate and improve learning of individuals (Kort & Reilly, 2002).

Autonomic Computing Systems

Autonomic computing (AC) is a scaled-down form of cognitive machines (Ganek & Corbi, 2003; IBM, 2006; Kinsner, 2005a; Wang, 2003, 2004, 2006). The systems are evolving earnestly because the cost-performance of hardware improvements (speed and capacity) has lead to escalating complexity of software (features and interfaces). However, this increased complexity requires elaborate managing systems that are now six to ten times the cost of the equipment itself. Autonomic computing is intended to simplify this problem by making the systems self-configuring, self-optimizing, self-organizing, self-healing, self-protecting, and self-telecommunicating, thus leading to increased reliability, robustness, and dynamic flexibility. This involves not only the traditional fault tolerant computing (i.e., tolerating hardware and software faults), but also tolerating various faults made by human operators and users, thus shifting attention from the *mean-time-between failures* (MTBF) to the *mean-time-to-recover* (MTTR) in order to make the systems more available. AC applies to desktop computing, portable computing, pervasive computing, and embedded systems.

Other Projects

Sandia Laboratories have been developing several projects related to cognitive machines, including human emulation, augmented cognition, knowledge capture, episodic memory, and naturalistic decision making. UCLA Cognitive Systems Laboratory has been involved in several projects, including evidential reasoning, default reasoning, learning, constraint processing, and graphoids (UCLA, 2006). James Anderson of the Department of Cognitive and Linguistic Sciences at Brown University in Providence has been developing the Ersatz brain project (Anderson, 2002, 2005), among many other neurocomputing projects.

PROCESSING AND METRICS

What do the theoretical models and practical implementations of cognitive machines have in common? The key distinguishing features include (1) intelligent signal processing that is robust and embedded in the machine itself, (2) real-time learning prompted by the awareness of the external and internal environments, and interaction with both, (3) closed-loop feedback control systems engineering so that past experiences could be used in the future, and (4) state and quality metrics to discern between the desirable and undesirable actions.

Intelligent Signal Processing

From the perspective of applied and industrial mathematics, a signal is an electrical representation of various ubiquitous, quantifiable physical variables such as temperature, pressure, flow, and light intensity. Such signals can be analog (continuous with very high resolution), discrete (sampled, but with very high resolution), digital (sampled and quantized to a finite resolution), boxcar (step-wise analog) functions over time (e.g., speech), space (e.g., images), or both (e.g., volumetric Doppler radar). The digital signals are of particular importance to computer-based signal processing which deals with the modeling, analysis/synthesis, feature extraction, and classification of such signals in order to gain insight into the underlying physical process, or to perform specific control tasks with the process.

Signal processing is used in nearly all fields of human endeavor from signal detection in the presence of noise, to fault diagnosis, advanced control, audio and image processing (restoration, enhancement, segmentation, reconstruction, coding, compression), communications engineering, intelligent sensor systems with reconfigurable architectures, business, and *humanistic intelligence* (HI) (Haykin & Kosko, 2001) which utilizes the natural capabilities of the human body and mind, as well as *cognitive informatics* (CI) (Wang, 2002) .

Intelligent signal processing (ISP) is treated in literature very thoroughly (e.g., Haykin & Kosko, 2001). Since many real-world physical systems are time varying, complex (high-dimensional), nonlinear, statistically nonstationary, non-Gaussian, nonlocal, sometimes chaotic, and subjected to unwanted signals (noise), the classical *statistical signal processing* (SSP) (e.g., Oppenheim, Schafer & Buck, 1999; Proakis & Manolakis, 1995) must be augmented by ISP and CI because of the required autonomy and interaction with humans. ISP has been found to be a more useful approach as it employs adaptation and learning to extract the essential information from the acquired signals and noise, without any assumed statistical models of the signals or theirs sources. These signals no longer exhibit additive invariance (short-range dependence), but multiplicative invariance (self-affinity with long-range dependence). The ISP tools include supervised and unsupervised learning through adaptive neural networks (Bishop, 1995), wavelets and their variations (Mallat, 1998), fuzzy rule-based computation (Pedrycz & Gomide, 1998), rough sets (Pawlak, 1991), granular computing (Bargiela & Pedrycz, 2002), genetic algorithms and evolutionary computation (Goldberg, 2002), and blind signal estimation (Haykin & Kosko, 2001). Algorithms for ISP must be implemented to satisfy the B-robustness property (Gadhok & Kinsner, 2006) because of higher-order statistics involved.

CI is concerned with (1) the extraction of characteristic features from signals obtained from measurements and observations, and (2) the measurement and characterization of patterns (i.e., order and correlation) in processes related to perception and cognition (i.e., interaction with humans).

Signals obtained from physical dynamical processes appear to be very complex. Much attention has been given to deterministic and stochastic linear-time-invariant (LTI) signals with a limited-bandwidth power spectrum density and short-tail distributions, leading to processing with finite moments. However, many physical signals are fundamentally different from the LTI signals in that they are invariant to scale rather than to translation (Wornell, 1996). Such signals have different degrees of singularity as measured by their noninteger (fractal) dimensions (Kinsner, 2005b). Correlation in such signals persists from short to very long ranges, with distributions having long tails (infinite moments). In contrast to the well-established LTI system theory, the nonlinear scale invariant (NSI) system theory and applications are still developing. There is also another class of signals, the chaotic signals, originating from nonlinear dynamical systems, such as the AC systems (Kantz & Schreiber, 2004; Principe, Euliano & Lefebvre, 2000; Sprott, 2003). Research is being conducted to measure and characterize such systems.

Real-Time Learning

Real-time learning may not be achievable due to the curse of dimensionality. A partial solution to this problem seems to be forthcoming through approximate dynamic programming (de Farias, 2002; de Farias & Van Roy, 2003, 2004; Haykin, 2005b). This approach is based on Bellman's dynamic programming, and it can be used as an optimal policy to guide the interaction of the learning system with its environment. Other approaches include neurofuzzy (Pedrycz & Gomide, 1998), granular (Bargiela & Pedrycz, 2002), and evolutionary computing (Goldberg, 2002).

Control With Feedback

This component can be implemented in many different ways, from classical control system to neurofuzzy (Pedrycz & Gomide, 1998).

Metrics

Cognitive systems cannot operate without some specific good-bad discriminating criteria, which, in turn, require metrics. Such metrics are being developed (e.g., Kinsner & Dansereau, 2006). The classical energy-based metrics (such as the scalar peak signal-to-noise ratio, PSNR) must be augmented by vectorial information-based or entropy-based metrics, such as the Rényi entropy spectrum, the Rényi fractal dimension spectrum (Kinsner, 2005b), or the *relative* Rényi fractal dimension spectrum (Kinsner & Dansereau, 2006). The latter can also be used for subjective quality metrics to evaluate images and video.

CONCLUDING REMARKS

Intelligent signal processing, real-time machine learning, and cognitive informatics are essential in solving some design and implementation issues of cognitive machines. Although the examples of cognitive radio, cognitive radar, active audition, cognitive robotics, and autonomic computing indicate that such cognitive systems are now feasible, much work must be conducted in all the areas to make the ambitious goals of such systems more practicable. Particular attention must be given to metrics and cognitive informatics.

ACKNOWLEDGMENT

Partial financial support of this work from the Natural Sciences and Engineering Research Council (NSERC) of Canada is gratefully acknowledged.

REFERENCES

Aleksander, I. (1989). *Neural computing architectures: The design of brain-like machines.* Cambridge, MA: MIT Press.

Aleksander, I. (1998, August). *From WISARD to MAGNUS: A family of weightless neural machines* (pp. 18-30).

Aleksander, I. (2003) *How to build a mind: Towards machines with imagination (maps of the mind)* (2nd ed.). New York: Columbia University Press.

Aleksander, I. (2006, May) *Artificial consciousness: An update*. Retrieved August 7, 2006, from http://www.ee.ic.ac.uk/research/neural/publications/iwann.html

Anderson, J. A. (2002, August 19-20). Hybrid computation with an attractor neural network. In

Proceedings of the 1st IEEE International Conference of Cognitive Informatics (pp. 3-12), Calgary, Alberta, Canada. ISBN: 0-7695-1724-2.

Anderson, J. A. (2005, August 8-10). Cognitive computation: The Ersatz brain project. In *Proceedings of the 4th IEEE International Conference of Cognitive Informatics* (pp. 2-3), Irvine, California. ISBN: 0-7803-9136-5.

Austin, J. (1998). *RAM-based neural networks*. Singapore: World Scientific.

Baars, B. (1988). *A cognitive theory of consciousness*. Cambridge, UK: Cambridge University Press. Retrieved August 7, 2006, from http://nsi.edu/users/baars/BaarsConsciousnessBook1988/index.html

Bargiela, A., & Pedrycz, W. (2002). *Granular computing: An introduction*. New York: Springer.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford, UK: Oxford University Press.

Chalmers, D. (1997). *The conscious mind: In search of a fundamental theory*. Oxford, UK: Oxford University Press.

Chomsky, N. (2006). *Language and mind* (3rd ed.). Cambridge, UK: Cambridge University Press.

Cotterill, R. (1988). *Computer simulations in brain science*. Cambridge, UK: Cambridge University Press

Cotterill, R. (2003). CyberChild: A simulation test-bed for consciousness studies. *Journal of Consciousness Studies, 10*(4-5), 31-45.

Dawkins, R. (1990). *The selfish gene* (2nd ed.). Oxford, UK: Oxford University Press.

Dennett, D. C. (1991). *Consciousness explained*. London: Allan Lane/Penguin.

Dillinger, M., Madani, K., & Alonistioti, N. (Eds.). (2003). *Software defined radio: Architectures, systems and functions*. New York: Wiley.

Farias, D. P. (2002). *The linear programming approach to approximate dynamic programming: Theory and application*. Doctoral dissertation, Stanford University, Stanford California. Retrieved August 7, 2006, from http://web.mit.edu/~pucci/www/daniela_thesis.pdf

de Farias, D. P., & Van Roy, B. (2003). The linear programming approach to approximate dynamic programming. *Oper. Res., 51*(6), 850–865.

de Farias, D. P., & Van Roy, B. (2004, August). On constraint sampling in the linear programming approach to approximate dynamic programming. *Math. Oper. Res., 29*(3), 462–478.

Fischler, M. A., & Firschein, O. (1987). *Intelligence: The eye, the brain and the computer*. Reading, MA: Addison-Wesley.

Franklin, S. (1995). *Artificial minds*. Cambridge, MA: MIT Press.

Franklin, S. (2003). IDA: A conscious artefact? In O. Holland (Ed.), *Machine consciousness*. Exeter, UK: Imprint Academic.

Freeman, W. (2001). *How brains make up their minds* (2nd ed.). New York: Columbia University Press.

Gadhok, N., & Kinsner, W. (2006). An implementation of beta-divergence for blind source separation. In *Proceedings of the IEEE Can. Conf. Electrical & Computer Eng.*, CCECE06 (Ottawa, ON; 10-12 May 2006) pp. 643-646, 2006.

Ganek, A. G., & Corbi, T. A. (2003). The dawning of the autonomic computing era. *IBM Systems J., 42*(1), 2003. Retrieved May 2006 from http://www.research.ibm.com/journal/sj/421/ganek.pdf)

Goldberg, D. E. (2002). *The Design of Innovation: Genetic Algorithms and Evolutionary Computation*. New York: Springer.

Grossberg, S. (1982). *Studies of Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition and Motor Control*. Boston: D. Reidel Publishing.

Grossberg, S. (Ed.). (1988). *Neural networks and natural intelligence*. Cambridge, MA: MIT Press.

Pentti, O.A., & Haikonen. (2003). *The Cognitive Approach to Conscious Machines.* New York: Academic. (See also http://personal.inet.fi/cool/pentti.haikonen/)

Pentti, O.A., & Haikonen. (2004). Conscious machines and machine emotions. Workshop on Models for Machine Consciousness. Antwerp, BE.

Haykin, S. (2005). Cognitive radio: Brain-empowered wireless communications. *IEEE J. Selected Areas in Communications, 23*(2), 201-220.

Haykin, S. (2005, September 28-30). Cognitive machines. *IEEE Intern. Workshop on Machine Intelligence & Sign. Proc.*, IWMISP05, Mystic, CT. Retrieved May 2006 from http://soma.crl.mcmaster.ca/ASLWeb/Resources/data/Cognitive_Machines.pdf

Haykin, S. (2006). Cognitive radar. *IEEE Signal Processing Mag.*, 30-40.

Haykin, S., & Chen, Z. (2005). The cocktail party problem. *Neural Computation*, 17, 1875-1902.

Haykin, S., & Kosko, B. (2001). *Intelligent Signal Processing.* New York: Wiley-IEEE.

Haykin, S., Principe, J. C., Sejnowski, T. J., & McWhirter, J. (2006). *New Directions in Statistical Signal Processing.* Cambridge, MA: MIT Press.

Hinton, G. E., & Anderson, J. A. (1981). *Parallel Models of Associative Memory.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Hoffman, R. R., Klein, G., & Laughry, K. R. (2002). The state of cognitive systems engineering. *IEEE Intelligent Systems Mag.*, 73-75.

Holland, O. (Ed.). (2003). *Machine consciousness.* Exeter, UK: Imprint Academic.

Hyvarinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis.* New York: Wiley.

IBM Autonomic Computing Manifesto. Retrieved May 2006, http://www.research.ibm.com/autonomic/

Kantz, H., & Schreiber, T. (2004). *Nonlinear Time series analysis* (2nd ed.). Cambridge, UK: Cambridge Univ. Press.

Kinsner, W. (2005, June 16-18). Signal processing for autonomic computing. *Proc. 2005 Meet. Can. Applied & Industrial Math Soc.*, CAIMS 2005, Winnipeg, MB. Retrieved May 2006 from http://www.umanitoba.ca/institutes/iims/caims2005_theme_signal.shtml

Kinsner, W. (2005, August 8-10). A unified approach to fractal dimensions. *Proc. IEEE 2005 Intern. Conf. Cognitive Informatics*, ICCI05, Irvine, CA (pp. 58-72).

Kinsner, W., & Dansereau, R. (2006, July). A relative fractal dimension spectrum as a complexity measure. *Proc. IEEE 2006 Intern. Conf. Cognitive Informatics*, ICCI06, Beijing, China.

Klivington, K. (1989). *The science of mind*. Cambridge, MA: MIT Press.

Kohonen, T. (1987). *Self-organization and associative memory* (2nd ed.). New York: Springer Verlag

Kort, B., Reilly, R. (2002). Theories for deep change in affect-sensitive cognitive machines: A constructivist model. *Educational Techn. & Society, 5*(4), 2002.

Kurzweil, R. (1990). *The age of intelligent machines.* Cambridge, MA: MIT Press.

Mallat, S. (1998). *A Wavelet Tour of Signal Processing.* San Diego, CA: Academic Press.

Mann, S. (2002). *Intelligent image processing.* New York: Wiley/IEEE.

Minsky, M. (1986). *The society of mind*. New York: Touchstone.

von Neumann, J. (1958). *The Computer and the Brain*. New Haven, CT: Yale Univ. Press.

Nielsen, M. A., & Chuang, I. L. (2000). *Quantum Computation and Quantum Information*. Cambridge, UK: Cambridge Univ. Press.

Oppenheim, A. V., Schafer, R. W., Buck, J. R. (1999). *Discrete-time signal processing* (2nd ed.). Prentice Hall.

Parsell, M. (2005). Review of P.O. Haikonen: The Cognitive Approach to Conscious Machines. *Psyche, 11*(2), 1-6. Retrieved May 2006 from http://psyche.cs.monash.edu.au/book_reviews/haikonen/haikone.pdf

Pawlak, Z. (1991). *Rough Sets: Theoretical Aspects of Reasoning about Data*. New York: Springer.

Pedrycz, W., & Gomide, F. (1998). *An Introduction to Fuzzy Sets: Analysis and Design*. Cambridge, MA: MIT Press.

Penrose, R. (1989). *The Emperor's New Mind*. Oxford, UK: Oxford Univ. Press.

Penrose, R. (1994). *The Shadows of the Mind: A Search for the Missing Science of Consciousness*. Oxford, UK: Oxford Univ. Press.

Pheifer, R., & Scheier, C. (1999). *Understanding intelligence*. Cambridge, MA: MIT Press.

Posner, M. (Ed.). (1989). *Foundations of cognitive science*. Cambridge, MA: MIT Press.

Principe, J. C., Euliano, N. R., & Lefebvre, W. C. (2000). *Neural and Adaptive Systems: Foundations through Simulations*. New York: Wiley.

Proakis, J. G., & Manolakis, D. G. (1995). *Digital Signal Processing: Principles, Algorithms and Applications* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Ralson, A., Reilly, E. D., & Hemmendinger, D. (Eds.). *Encyclopedia of Computer Science* (4th ed.). New York: Wiley.

Roy D. K. (2005). Grounding words in perception and action: Insight for computational models. *Trends in Cognitive Sciences, 9*(8), 389-396.

Roy, D. K., & Pentland, A. P. (2002). Learning words from sights and sounds: A computational model. *Cognitive Science*, *26*, 113-146.

Ruaro, M. E., Bonifazi, P., & Torre, V. (2005). Toward the neurocomputer: Image processing and pattern recognition with neuronal cultures. *IEEE Trans. Biomedical Eng.*, *52*(3), 371-383.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing* (Vols. 1 and 2). Cambridge, MA: MIT Press.

Sandia National Laboratories, "Projects" (2006). Retrieved May 2006 from http://www.sandia.gov/cog.systems/Projects.htm

Sanz, R., Chrisley, R,. & Sloman, A. (2003). *Models of Consciousness: Scientific Report*. European Science Foundation. Retrieved May 2006 from http://www.esf.org/generic/1650/EW0296Report.pdf http://aslab.disam.etsii.upm.es/public/events/moc/report/ScientificReport.pdf

Searle, J. R. (1980). Minds, brains and programs. *Behavioral & Brain Sciences*, 3, 417-424.

Searle, J. R. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.

Sienko, T., Adamatzky, A., Rambidi, N. G., & Conrad, M. (2003). *Molecular computing*. Cambridge, MA: MIT Press.

Soloman, S. (1999). *Sensor handbook*. New York: McGraw-Hill.

Sprott, J. C. (2003). *Chaos and time-series analysis*. Oxford, UK: Oxford Univ. Press.

Taylor, J. G. (2001). *The race to consciousness.* Cambridge, MA: MIT Press.

Taylor, J. G. (2002). Paying attention to consciousness. *Trends Cogn. Sciences,* 6, 206-210.

Taylor, J. G. (2003, June 20-24). The CODAM model of attention and consciousness. *Proc. Intern. Joint Conf. Neural Networks,* IJCNN03, Portland, OR (Vol. 1, pp. 292-297).

Thelen, E., & Smith, L. B. (2002). *A Dynamic System Approach to the Development of Cognition and Action.* Cambridge, MA: MIT Press.

UCLA Cognitive Systems Laboratory. (2006). Retrieved May 2006 from http://singapore.cs.ucla.edu/cogsys.html

Velmans, M. (2000). *Understanding consciousness.* New York: Routledge.

Wang, Y. (2002, August 19-20). On cognitive informatics. *Proc. 1st IEEE Intern. Conf. Cognitive Informatics,* Calgary, AB (pp. 34-42).

Wang, Y. (2003, August). Cognitive informatics models of software agent systems and autonomic computing: Keynote Speech. Proc. Intern. Conf. Agent-Based Technologies and Systems (ATS'03), U of C Press, Calgary, Canada.

Wang, Y. (2004, August). On autonomic computing and cognitive processes: Keynote speech. In *Proceedings of 3rd IEEE Intern. Conf. on Cognitive Informatics*, (pp. 3-4) ICCI04, Victoria, BC.

Wang, Y. (2006, July 17-19). Cognitive Informatics - Towards the future generation computers that think and feel: Keynote speech. In *Proceedings of 5th IEEE Intern. Conf. Cognitive Informatics,* (pp. 3-7) ICCI06, (Beijing, China;, July 17-19, 2006) ISBN 1-4244-0475-4, 2006.

Wornell, G. (1996). *Signal processing with fractals: A wavelet-based approach.* Upper Saddle River, NJ: Prentice-Hall.

This work was previously published in the International Journal of Cognitive Informatics and Natural Intelligence, Vol. 1, Issue 1, edited by Y. Wang, pp. 28-38, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Chapter 8.12
Biotechnology Portals in Medicine

Yoosuf Cader
Zayed University, UAE

INTRODUCTION

The 2005 global revenues of publicly traded biotechnology companies have grown by 18.1% to $63.1 billion (Donn, 2006). Many countries are now investing in research and development in the biotechnology industry as it is believed this 30 year-old industry is moving toward profitability. The stock value in this industry has outperformed the average stock value in many countries. In the pre-genomic era, a typical life sciences company would have marketed diagnostic kits, assays, chemicals, measuring equipment, and research products to name a few. In the genomic era, a new range of products is marketed focusing on molecular medicine. Among these new products are bioinformatics software solutions, storage systems, biotechnology systems, and solutions researching into genes and proteins, tools for analysis of genetic sequence data, integrated systems and solutions for disease research, and new drug discovery (Cader, 2004). The need for biotechnology portals is now more than justified and will be a useful information and knowledge source.

A biotechnology portal is a gateway of comprehensive source of information and knowledge to those interested in knowing about biotechnology and the benefits this industry is offering. It should be considered as the first point of reference for those seeking reliable, quality, and current information and knowledge about issues in biotechnology. In addition, it should be interactive and have the appropriate tools to enable a community of users to share information and knowledge among them. There should also be a commercial component to the biotechnology portal, which should be to generate revenue through advertisements and offers to its target visitors. This revenue is essential to ensure the maintenance and survival of the portal and offer value to all its stakeholders. A biotechnology portal will not be complete unless it provides information on biotechnology stocks to potential investors seeking insights into this industry. A biotechnology portal is like any online business with various objectives such as profits, growth, market share, and innovation.

BIOTECHNOLOGY IN MEDICINE

This article will concentrate on biotechnology in medicine although the benefits of biotechnology have influenced other disciplinary areas such as agriculture and environmental sciences. In medicine, hundreds of biotechnology drug products and vaccines are currently in human clinical trials in advanced countries with many more in the new product development stage.

Linking to Web Sites

A good starting point is to link up with the site http://www.ornl.gov/hgmis/ sponsored by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research, Human Genome Program. This site gives information on the Human Genome Project (HGP), news and announcements, planned user facilities, educational resources, research progress, impacts of research, GTL documents (potential microbial documents), science, technologies behind GTL, Gene Gateway (tools for exploring the sequence), and related department of energy sites. There are also links to related sites.

The biotechnology portal should also link up to an excellent Web site, which makes available several DNA and Gene images (http://www.ornl.gov/hgmis/graphics/slides/images1.html) in an image gallery.

The Wellcome Foundation in the UK (http://genome.wellcome.ac.uk/) gives the latest news, features, and background, and a lot of information about the human genome—exploring genes and its impact on health, disease, and society. The Wellcome foundation is also the primary funding source for the Sanger Institute at Cambridge University (http://www.sanger.ac.uk/). The Sanger Institute is a genome research institute whose aim is to further the knowledge of genomes, particularly through large scale sequencing and analysis.

Another interesting site in the UK that should be included is http://www.geneservice.co.uk/. Geneservice Limited is a contract research organisation and biological resource centre, which supplies genomic products and technical services to both academic and commercial research organisations. These services include whole genome amplification, DNA sequencing, micro-satellite, and SNP genotyping including 10K and 100K mapping, and expression array analysis.

An interesting European Web site is http://www.litbio.org/. This is a Laboratory for Interdisciplinary Technologies in Bioinformatics (LITBIO) applied to genomics, transcriptomics, proteomics, and metabolomics providing international research and development programs with the new analysis strategies of biomedical and biotechnological data. The laboratory consists of five collaborating partners whose links are listed as follows:

- CILEA (http://www.litbio.org/cilea.htm)
- CNR/IEIIT (http://www.litbio.org/ieiit.htm)
- Universita' degli Studi di CAMERINO (http://www.litbio.org/camerino.htm)
- Universita' degli Studi di GENOVA (http://www.litbio.org/genova.htm
- Eurotech S.p.A. (http://www.litbio.org/eurotech.htm

Another interesting link is to this Japanese Web site http://www.genome.ad.jp/kegg/. The Kyoto Encyclopaedia of Genes and Genomes (KEGG) is a bioinformatics resource developed by the Kanehisa Laboratories in the Bioinformatics Center of Koyoto University and the Human Genome Center of the University of Tokyo. They are working toward representing a complete computer representation of the cell, the organism, and the biosphere. This will enable computational prediction of higher-level complexity of cellular processes and organism behaviours from genomic and molecular information.

The biotechnology portal should also be linked to other informative Web sites in other countries including the Australian Web sites.

This Web site (http://www.csiro.au/pub-genesite/index.htm) explains what Australian scientists are doing in the area of biotechnology research. For information about gene technology, policy, and regulations, there should be a link to biotechnology Australia's gateway site at www.biotechnology.gov.au. Another useful link should be to the Victorian government biotechnology Web site. This Web site offers an overview of biotechnology in Victoria and links to industry sectors and biotechnology centres. It also includes the biotechnology strategic development plan for Victoria.

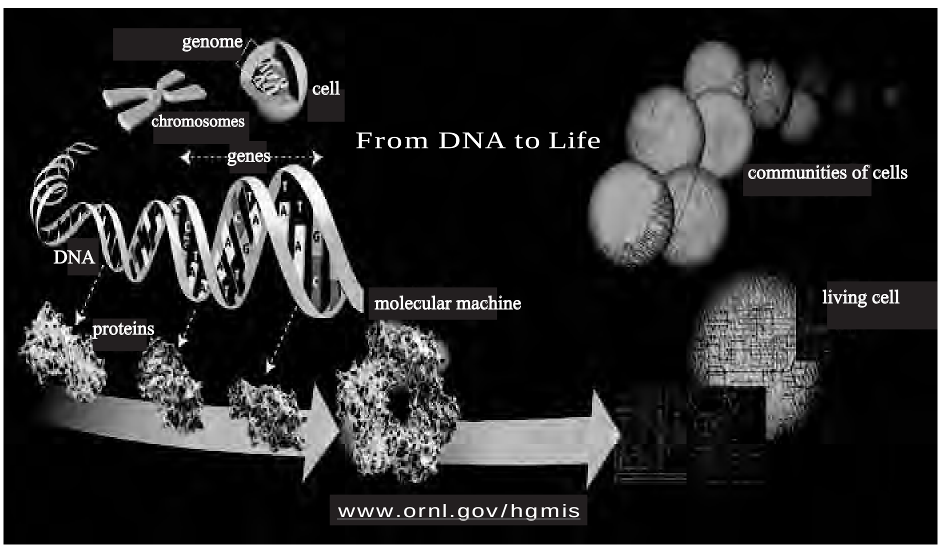
This Web site (http://www.business.vic.gov.au/BUSVIC.458954/INDUSTRY/1226260600/PC_60362.html) offers an overview of biotechnology in Victoria and links to industry sectors and biotechnology centres. It also includes the biotechnology strategic development plan for Victoria. Another useful link would be to http://www.ausbiotech.org. Ausbiotech is the industry body representing the Australian biotechnology sector.

The molecular biology database collection is a public online resource listing key databases. The online version of this journal (Nucleic Acids Research) article has been published under an open access model (Galprin, 2006). The biotechnology portal must definitely be linked to this up-to-date database, which is intended to serve as the initial point from which to find specialised databases that may be of use in biological research. This database includes 858 updates in the 2006 version. It includes major public sequence repositories, gene expression, gene identification and structure, genetic and physical maps, genomic databases, intermolecular interactions, metabolic pathways and cellular regulations, mutation databases, pathology, protein databases, protein sequence motifs, proteome resources, retrieval systems and database structure, and varied biomedical content.

Icelandic's "Decode Genetics"

This is an Icelandic biopharmaceutical company. Its Web site (http://www.decode.com/) shows the type of work it does. The nature of work became

Figure 1. Courtesy of the U.S. Department of Energy Human Genome Program



of much interest worldwide because the company started developing a genealogical database. When the information from this database is combined with health care data (diagnosis, treatment, prognosis, measurement of values in blood, etc.) it will help identify genes linked to certain diseases. The concept is to look at the community as a system of data that can be mined for knowledge.

UK Biomedical Population Collection

A similar study in the UK called "*UK Biomedical Population Collection*" was established between the Wellcome Trust, the Medical Research Council, and the Department of Health, which looks at the interaction between lifestyle, environment, and genetic makeup that causes diseases. This project will collect key medical and lifestyle information from participants between the ages of 45 and 64 over a period of at least 10 years. The Web site associated with this study should also be linked to the biotechnology portal. Further details can be seen in http://www.phgu.org.uk /ecard?reference_ID=2688.

Stem Cells, Embryonic Stem Cells, Embryonic Stem Cell Line, Somatic Stem Cells

The National Health Institute Web site is a useful resource for all one wants to know about stem cells and the ensuing debate about stem cells research (http://stemcells.nih.gov/info/glossary.asp).

Apart from the previous National Health Institute (U.S.) Web site, another useful Web site to which the biotechnology portal should be linked is the National Human Genome Research Institute Web site (http://www.genome.gov/10004765). According to this Web site, the term "cloning" is used by scientists to describe many different processes that involve making duplicates of biological material. In most cases, isolated genes or cells are duplicated for scientific study and no new animal results. The experiment that led to the cloning of Dolly the sheep in 1997 was different. It used a cloning technique called *somatic cell nuclear transfer* and resulted in an animal that was a genetic twin—although delayed in time—of an adult sheep. This technique can also be used to produce an embryo from which cells called *embryonic stem* (ES) *cells* could be extracted to use in research into potential therapies for a wide variety of diseases. This Web site also discusses ethical concerns, policy and regulation, and cloning for the isolation of Human ES Cells.

Content of the Biotechnology Portal

A good biotechnology portal should be innovative and interactive rather than just be a source of information. The design should be creative and welcoming to the browser. The content should offer quality material, which is regularly updated. The technology should enable the portal to be user friendly and fast.

CONCLUSION

Mentioned above are only some of the possible links to the portals and Web sites of the numerous genomic and medical Web sites originating in almost every country. Eventually a good biotechnology portal will be a gateway to most of these Web sites. The portal should be linked to all the biotechnology *hotspots* in the world. The *hotspots* refer to biotechnology research centers and institutes involved in biotechnology research and development activities.

There are numerous other Web sites of private companies that should be liked to the biotechnology portal but they are too numerous to be mentioned here.

Agricultural and environmental aspects of biotechnology are also important components of a biotechnology portal. Other aspects are an executive search section for biotechnologists and links to online publications such as "nature," "science," "today's life science" etc., to name a few of the many academic journals in biotechnology. For a biotechnology portal to be viable, it needs to offer advertising space for a fee for companies wanting to market related products and knowledge sources.

REFERENCES

Cader, Y. (2004). Knowledge management in the life sciences industry. In Y. Cader (Ed.), *Knowledge management: Theory & application in the twenty-first century context* (pp. 1-17). Victoria, Australia: Heidelberg Press.

Donn, S. (2006). *Beyond borders. The global biotechnology report.* Ernst & Young.

Galprin, M. Y. (2006). The molecular biology database collection: 2006 update. *Nucleic Acids Research, 34*(D3-D5). Oxford University Press.

Krane, D. E., & Raymer, M. L. (2003). *Fundamental concepts of bioinformatics.* Pearson Education, inc. (Publishing as Benjamin Cummings, CA, USA). Pages 277-284 (Glossary).

Web Site References

Genome Programs of the US Department of Energy Office of Science. Retrieved June 26, 2006, from http://www.ornl.gov/hgmis/

Genome Science Images of the US Department of Energy Office of Science. Retrieved June 26, 2006, from http://www.ornl.gov/hgmis/graphics/slides/images1.html

Wellcome Trust site of the Human Genome (UK). Retrieved June 26, 2006, from http://genome.wellcome.ac.uk/

The Sanger Genome Research Institute (Cambridge, UK). Retrieved June 26, 2006, from http://www.sanger.ac.uk/

Geneservice Contracts research company (UK). Retrieved June 26, 2006, from http://www.geneservice.co.uk/

Kegg: Kyoto Encyclopedia of Genes and Genomes. Retrieved June 26, 2006, from http://www.genome.ad.jp/kegg/

Commonwealth Scientific Industrial Research Organisation of Australia. Retrieved June 26, 2006, from http://www.csiro.au/pubgenesite/index.htm

Biotechnology Australia (A Government Initiative). Retrieved June 26, 2006, from www.biotechnology.gov.au

Business Victoria (Australian Victorian Government Initiative). Retrieved June 26, 2006, from http://www.business.vic.gov.au/BUSVIC.458954/INDUSTRY/1226260600/PC_60362.html

AusBiotech is the industry body representing the Australian biotechnology sector. Retrieved June 26, 2006, from http://www.ausbiotech.org

Icelandic biopharmaceutical company. Retrieved June 26, 2006, from http://www.decode.com/

Public Health Genetics Unit (UK). Retrieved June 26, 2006, from http://www.phgu.org.uk/ecard?reference_ID=2688

National Institute of Health Resource for Stem Cell Research (U.S.). Retrieved June 26, 2006, from http://stemcells.nih.gov/info/glossary.asp

National Human genome Research Institute (U.S.). Retrieved June 26, 2006, from http://www.genome.gov/10004765

KEY TERMS

Bioinformatics: A new and exciting field, which developed due to the power of super computers enabling faster processing and storage of terabyte levels of data, using computer software tools in storing, organising, and computationally analysing biological information.

Biotechnology: It is the application of knowledge based on cellular and molecular processes of living organisms and its ability to act as molecular machines to make new proteins with the potential to become new drugs; to develop new industrial processes and new genetically modified high yielding crops. There is also a convergence of rapid advances in information technology with the rapidly growing knowledge of the cellular and molecular processes resulting in the new opportunities.

Biotechnology Portal: It is a gateway of a comprehensive source of information and knowledge to those interested in knowing about biotechnology and the benefits this industry is offering. It should be considered as the first point of reference for those seeking reliable, quality, and current information and knowledge about issues in biotechnology.

Chromosomes: The DNA molecule containing a cell's genome. These strands of DNA form the 23 pairs of chromosomes inside the cell nucleus with approximately 3 billion DNA sub units or chemical base pairs called nucleic acids, denoted by the letters A, T, G, and C. Each pair of nucleic acid is called a base pair. These are like rungs on a spiralling staircase.

Cloning: The term cloning is used by scientists to describe many different processes that involve making duplicates of biological material. In most cases, isolated genes or cells are duplicated for scientific study, and no new animal results (http://www.genome.gov/10004765)

DNA: Deoxyribonucleic acid (see Nucleic acids) is known as the molecule of life.

Drug Discovery: The process involved in discovering new drugs.

Embryonic Stem Cells: Primitive (undifferentiated) cells derived from a 5-day preimplantation embryo that have the potential to become a wide variety of specialized cell types (http://stemcells.nih.gov/info/glossary.asp)

Embryonic Stem Cell Line: Embryonic stem cells, which have been cultured under *in vitro* conditions that allow proliferation without differentiation for months to years (http://stemcells.nih.gov/info/glossary.asp)

Gene: A specific sequence of nucleotides in DNA (deoxyribonucleic acid) or RNA (Ribonucleic acid). There are 26,261 genes distributed along the 23 pairs of chromosomes in humans. Each gene contains a particular set of information coding for a particular protein in a cell.

Genome: The sum total of an organism's genetic material.

Nucleic Acids: There are approximately 3 billion DNA sub units or chemical base pairs, called nucleic acids in the human genome, denoted by the letters A, T, G, and C. Each pair of nucleic acid is called a base pair. These are like rungs on a spiralling staircase.

Pharmacogenomics: Field that uses information about an individual's genetic makeup to maximize the efficacy of treatments, while at the same time minimizing the unwanted side effects (Krane & Raymer, 2003).

RNA: Ribonucleic acid (see Nucleic acids).

Somatic Stem Cells: Non-embryonic stem cells that are not derived from gametes (egg or sperm cells) (http://stemcells.nih.gov/info/glossary.asp)

Stem Cells: Cells with the ability to divide for indefinite periods in culture and to give rise to specialized cells (http://stemcells.nih.gov/info/glossary.asp).

This work was previously published in the Encyclopedia of Portal Technologies and Applications, edited by A. Tatnall, pp. 89-93, copyright 2007 by Information Science Reference, formerly known as Idea Group Reference (an imprint of IGI Global).

Chapter 8.13
Soft Statistical Decision Fusion for Distributed Medical Data on Grids

Yu Tang
Georgia State University, USA

Yan-Qing Zhang
Georgia State University, USA

ABSTRACT

This chapter introduces the decision fusion as a means of exploring information from distributed medical data. It proposes a new method of applying soft data fusion algorithm on the grid to analyze massive data and discover meaningful and valuable information. It could potentially help to better understand and process medical data and provide high-quality services in patient diagnosis and treatment. It allows incorporation of multiple physicians into one single case to recover and resolve problems, and integration of distributed data sources overcome some limitations of geographical locations to share knowledge and experience based on the soft data and decision fusion approach.

INTRODUCTION

Healthcare service is a complex industry nowadays. It is one of the most critical components of the modern human-oriented service. Informatics is an essential technology to health care (Dick & Steen, 1991) and has been applied to this field as long as computers have existed. Information technology can be one of the major drivers of e-health activity, both directly and indirectly. E-health offers new opportunities to further improve the quality and safety of services because technology makes possible the high level of information management. However, it raises an important issue of how to utilize and integrate an impressive amount of medical data efficiently to provide high-quality and safe services.

Grid computing has emerged to address this issue. It was first developed in the scientific community and can be used as effective infrastructures for distributed high-performance computing and data processing (Foster, Kesselman, & Tuecke, 2001). The features of grid computing make the designation of an advance decision support system possible.

How to apply data fusion in a distributed medical decision system on the grid is still an open problem. In our previous research on this subject, the following observations were made that should guide our further work:

1. Massive data are collected in different organizations. With an explosion in size of database, discovering meaningful and valuable information from different datasets on grids is still a critical issue that affects decision-making in this area. There is an urgent need for a new computation technique to help service providers to process, analyze, and extract meaningful information from the rapid growing data.
2. The need for efficient, effective, and secure communication between multiple service providers for sharing clinical knowledge and experience is increasing. Traditional techniques are infeasible for analyzing large datasets that may maintain over geographically distributed sites.
3. The need for finding an efficient way to integrate data, knowledge, and decision from different parties is increasing.

These first two observations suggest an answer: build a grid-based system that enables the sharing of application and data in an open, heterogeneous environment. The last observation suggests an answer to build a soft fusion mechanism to do summarization, and it may result in higher accuracy of diagnosis and better treatment.

RELATED WORK

There are several research groups whose work can contribute to grid-based data fusion on e-health. We first discuss decision support on the grid in the grid community, then we will introduce some related works about the medical decision support from the health community, and finally we will present some related works about soft data fusion and our proposal for solving this problem.

Decision Support on the Grid

A decision support system is defined as any computer program that assists decision-makers to utilize data and models to solve problems (Gorry & Morton, 1971; Keen & Morton, 1978; Sprague & Calson, 1980). Usually, it requires access to vast computation resources and processes a very large amount of data to make a decision. Grid computing is one approach to solving this problem. It has emerged as a paradigm with the ability to provide secure, reliable, and scaleable high-speed access to a distributed data resource. Compared to traditional distributed techniques, it has many advantages like resource sharing, high-performance services. The grid offers significant capability for designation and operation of complex decision support system by linking together a number of geographically distributed computers (Ong et al., 2004).

A grid-based decision support system can be used in a broad range of problems, from business to utilities, industry, earth science, health care, education and so on. Most researchers focus on simulation and visualization for specific processes such as air pollution (Mourino, Martin, Gonzalez, & Doallo, 2004), flooding crisis (Hluchy et al., 2004; Benkner, et al., 2003), and surgical procedures (Narayan, Corcoran-Perry, Drew, Hoyman, & Lewis, 2003; CrossGrid project) and then support decision-makers to make decisions on the basis of simulation results.

Medical Decision Support

The term medical decision support system describes a set of computer applications that are designed to assist health service providers in clinical decision-making. It can provide assessment or specifics that are selected from the knowledge base on the basis of individual patient characteristics (Hunt, Haynes, Hanna, & Smith, 1998; Delaney, Fitzmaurice, Riaz, & Hobbs, 1999). It is typically designed to integrate a medical knowledge base, patient data, and an application to produce case-specific advice.

The decision support system has been used in health care since the 1960s. There is evidence that using a medical decision support system may increase compliance with clinical pathways and guidelines and reduce rates of inappropriate diagnostic tests (Australia's Health Sector). It can support increased use of evidence by clinicians in direct patient care, resulting in better patient outcomes. However, the use of computerized medical decision systems is not commonplace. The results achieved have been rather low and the progress is slow (Reisman, 1996). Two identified barriers are lack of sources of knowledge and system development (Shortliffe, 1986), and lack of communication among profusion of different systems (Hobbs, Delaney, Carson, & Kenkre, 1996). As many researchers say, there is a rapidly growing need to improve medical decision-making in order to reduce practice variation, preventable medical errors (Poses, Cebul, & Wigton, 1995; Bornstein & Emier, 2001; Sintchenko & Coiera, 2003) and become feasible in the real world.

Soft Data Fusion

Data fusion is the amalgamation of information from multiple sources. It can be classified as either hard fusion and soft fusion.

All data fusion efforts are initiated to be used in particular research areas. It is still a "wide open field based on the difference in technology, the expectations by the users, and the kinds of problems that biologists and life scientists try to solve" (Freytag, Etzold, Goble, Schward, & Apweiler, 2003). Fusion system application can be found in the domain of hydrological forecasting (Abrahart & See, 2002), health care and medical diagnose (Laskey & Mahoney, 2003), and engineering (Chow, Zhu, Fischl, & Kam, 1993).

There are different approaches in the literature to fuse data. Some approaches use statistical analysis, while others use AI techniques like fuzzy logic (Chen & Luo, 1993), learning algorithms based on neural networks (Myers, Laskey, & DeJong, 1999), and Bayesian networks or uncertainty sets (Singhal & Brown, 1997) to handle the uncertainty.

RESEARCH PROBLEM

In the healthcare industry, many organizations that could be located in different places collect data. In the traditional way, data is fragmented, and it is inconvenient for service providers to share experience and knowledge. Physicians would change medical decisions if they had enough "knowledge." Assume that they may collaborate under some agreements on some concerned problems. The following are some typical scenarios of example processes in e-health:

- Health service providers collaborate on the analysis of newly discovered disease or pathogenic bacteria.
- Health service providers collaborate on the estimate of the patient's state and providing appropriate treatments.
- Health service providers share the experience and knowledge with others.
- Health service providers get the support from others to make decisions for uncertain cases.

In these scenarios, collaboration across geographical location is needed to enable the sharing of data and knowledge and then make a decision. Greater benefits can be achieved if data integration is used rather than simply data collection

(Sensmeier, 2003). Our research problems can be described as the following:

- **Quality of the medical decision.** Liability and reliability are two main issues of the medical decision system. Decision support tools must be carefully designed so that they are reliable and accurate (Sloane, Liberatore, & Nydick, 2002). How does one utilize grid computing to perform the data process? How does one use grid-based distributed data mining to discover knowledge? How does one make use of massive distributed data to improve the quality of service?
- **Collaboration.** How do the different parties collaborate with each other? How can we get data about patients and transmission through the Internet? Can data be integrated in different levels? How does one integrate the data, information, knowledge, and decisions from different organizations?

The following are our goals:

- **For service providers.** Share knowledge and experience to make high QoS decisions and choose the optimal treatment for the individual patient.
- **For patients.** Get better medical care that includes more accurate diagnosis results and better treatments.

RESEARCH PLAN

In order to achieve all the requirements to make a system available, we need to integrate the network with multiple health service providers and patients. Grid computing allows flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources; it can be used as effective infrastructures for distributed high-performance computing and data processing (Foster, Kesselman, & Tuecke, 2001). The grid-based system of logical architecture is presented in Figure 1.

Figure 1 shows the logical architecture diagram of the system. The computers with service provider applications are the nodes of the grid. Health service providers communicate, exchange data, and share experience and knowledge through the grid service. Computers on the grid can be desktop, laptop, pocket PC, cell phone, and so on. Users of this system can be doctors, specialists, and assistants. They — users and computers — can communicate with each other through standard or wireless Internet. The typical data workflow is described in the following scenarios:

- Data of patient's current situation is sent to doctor.
- Decision support system starts to analyze data.
- Decision support system invites other doctor / system on the grid to participate in the diagnosis.
- Decision support system collects results from the grid and fuses results, then it generates decision.
- Doctor sends back results to patient.

In order to perform such tasks, the decision support system has three integrated modules: grid service module, fusion service module, and user service module. The software architecture is shown in Figure 2.

Several modules work in the system to carry out tasks. The user friendly interface can accept tasks from and send back responses to users; the grid service agent module is used to provide basic services to manage the grid, coordinate actions, and resolve resources among nodes. The fusion agent module is used to implement different levels of data fusion. The data analysis agent module is used to analyze data and make diagnosis and decisions. Both fusion and data analysis modules work with the medical database. These three com-

Figure 1. Logical architecture

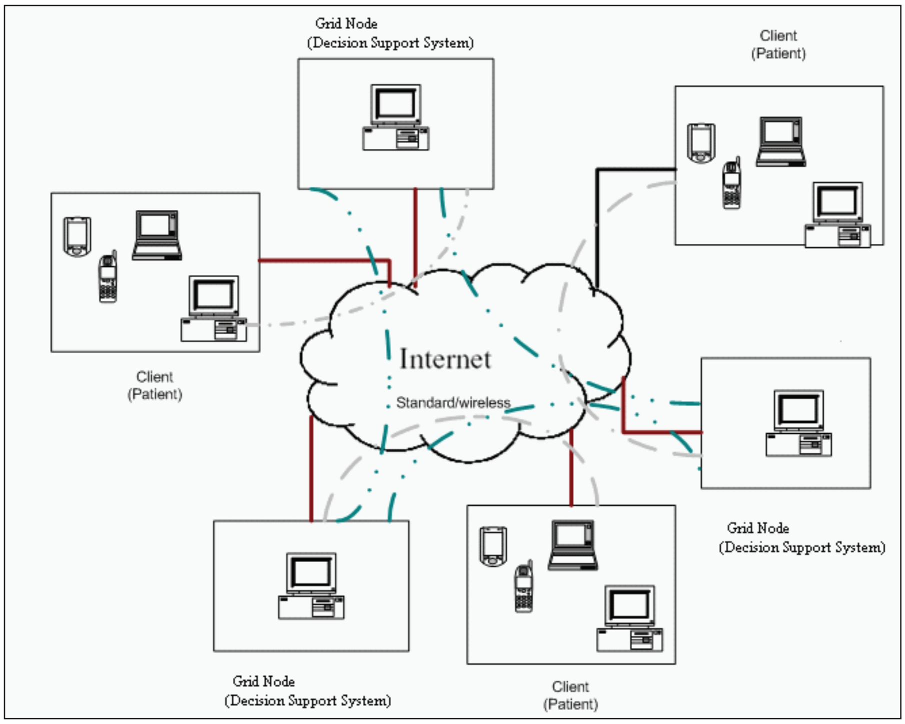
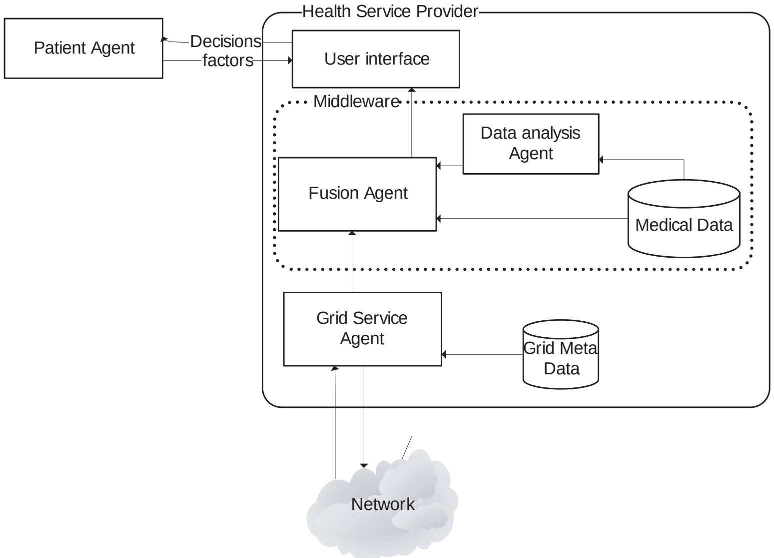


Figure 2. Software architecture



ponents are composed of middleware to provide services to the user.

To illustrate how this will work, we will first describe some key issues that are associated with grid-based distributed knowledge management technologies, as well as methodology for decision-making. We will then discuss plans to develop a feasible fusion mechanism for data, knowledge, and decision assembly.

Knowledge Management Technologies

Some factors that influence the quality of a medical decision system include: the quality of the underlying knowledge base used in the system; the incomplete dataset; and the conversion of knowledge into electronic form. Grid-based distributed data mining is used in this system to solve these problems and thus improve the quality of decisions. We will discuss these following issues:

- **Data privacy.** Medical data are sensitive and proprietary (Shamos, 2004); many hospitals and organizations treat health and medical data as their own property and are not willing to share with others. Proposed solutions to this issue include de-identification (Tracy, Dantas, & Upshur, 2004; Li & Chen, 2004; Kline, Johnson, Webb, & Runyon, 2004) and data-centered (Du & Zhan, 2002; Kantarcioglu & Clifton, 2002). We propose to follow both ways to protect data privacy.
- **Data preparation.** One barrier that influences the quality of a decision is the quality of medical data. They are often incomplete or out of time. The crucial information is missing when the decision is made. Data preparation is important to generate a high-quality decision. Conceptual reconstruction is one of options to be used to solve this problem (Aggarwal & Parthasarathy, 2001).
- **Data communication.** There is a security concern about this system. The use of grid security infrastructure (GSI) (Welch et al., 2003) allows secure authentication and data exchange over an opened network.
- **Data mining methods.** Using data mining in medical decisions can generate a decision of high accuracy (Kusiak, Kern, Kernstine, & Tseng, 2000). Several data mining algorithms are available for a decision-making system including: classification trees, case-based reasoning, neural network, genetic algorithm, fuzzy set approach, SVM, and so on. Our ongoing work is based on SVM and case-based reasoning. We will test and evaluate other algorithms in our future research.

Grid Toolkit

Grid computing is one of the innovative distributed computational models. It can offer high-performance computing and data processing abilities for the distributed mining and extraction of knowledge from data repositories. Grid applications are used in many fields including scientific computing, environmental monitoring, geohazard modeling, and business. It can be used as an effective infrastructure for distributed computing and data mining (Foster, 2001).

Grid technology is growing up very quickly and is going to be more and more complete and complex both in the number of tools and in the variety of supported applications (Cannataro, 2001). Compared to traditional distributed techniques, it has many advantages like resource sharing and high-performance services. Many existing tools are designed to provide functionalities of integration, resource management, access, process large datasets, and support the knowledge discovery process.

Globus Toolkit (Globus project group) is one candidate to implement grid management. It is a well-known grid middleware for providing grid resource management, security management, and other grid facilities. Globus Toolkit 3.0 is the first grid platform to fully support the OGSA/OGSI standard. It provides functionalities to discover, share, and monitor the resources. It also provides abilities for the mutual authentication of services and protection of data (Butler, 2000). The fusion module and application module are built on the base of Globus Toolkit.

Distributed Knowledge Discovery

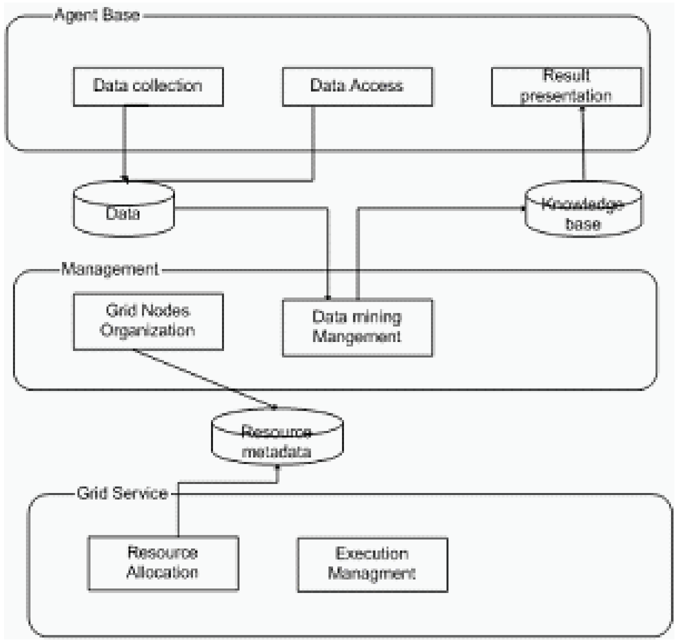
The distributed knowledge discovery is a process that applies artificial intelligent theories and grid techniques to extract knowledge from distributed databases on the grid. Such processes can be implemented in the following steps:

1. **Data preparation.** The first step of this process is data preparation. Data retrieved from different databases on the grid needs to be pre-processed for two main reasons:
    - *Medical data is sensitive.* The data in this domain cannot be obtained without privileges. In order to protect data security, we follow two different ways: de-identification and data-centered. For de-identification, medical data in all organizations can be categorized into two parts: One part is pure medical data without any identified personal information of patients, it is accessible for all partners on the grid; the other part is full dataset with private patient information, it is only accessible for the owner. For data center, data processing is performed on a local data source. After processing, analysis results are based on data not data itself exchanged among the grid. Data and different levels of analysis results can be input of different level of fusion agent.
    - *Medical data may not be complete.* In order to minimize noise caused by incomplete data, pre-process data is necessary. The idea of conceptual reconstruction can be used to fill the missing data. One refill process is performed by calculating the mean and deviation of individual data values of each attribute; the other one is performed by finding the approximate patterns and calculating the mean of patterns. Different algorithms are used to calculate and fill missing values.
2. **Data exchange.** The second step of this process is data exchange. Grid tools provide secure, high-throughput data transfer. This model is set in terms of layer, just like many other grid-based knowledge discovery systems. Services provided by this model are set in a three-layer infrastructure as depicted in Figure 3.
3. **Data analysis.** The third step is data analysis. Several methods are used to analyze data. To take advantage of the grid, different data analysis applications can be used on different machines. Statistical algorithms and data mining algorithms are used to train medical data and generate results for a given case. In addition, they are used to discover useful knowledge on the distributed database. The knowledge explored includes membership functions and fuzzy rule sets. It aims to generate a number of fuzzy rules and membership functions by applying data mining algorithm to a collection of dataset on grids.
4. **Fusion analysis.** The system has the self-developing ability to analyze fusion results. Fusion logs are kept in a database, which enable system learning from outside resources.

Fusion Technologies

In real-world applications, a very large amount of data may be kept in the distributed database and

Figure 3. Three-layer grid-based architecture



can be accessed at an acceptable rate. Collaboration among many organizations is an important issue in making decisions with high accuracy. To combine data, information, and decisions from different parties, fusion is used in some research (Azuaje, Dubitzky, Black, & Adamson, 1999; Phegley, Perkins, Gupta, & Dorsey, 2002). The fusion technologies can be applied to different application domains.

We propose four possible levels of fusion with interactive discussion including different approaches to manage the fusion process. The integration of data and decision can occur in these four distinct levels — data, information, knowledge, and decision. Fusion service is based on basic grid mechanisms. It is built on top of grid services. Decisions made on the basis of local data are collected and fused. Such process aims to generate reliable decisions for health service providers by applying data mining algorithm and AI technologies to a collection of dataset on grids. It overcomes the disadvantages of human fusion and machine fusion: The former is limited by knowledge and subjective experience, while the latter is inflexible and relies on data excessively.

On account of the features of medical data, we propose soft fusion in our work, including fuzzy logic and simple weighting / voting. Neural network fusion and Bayesian fusion would be our future test.

Four-Level Fusion

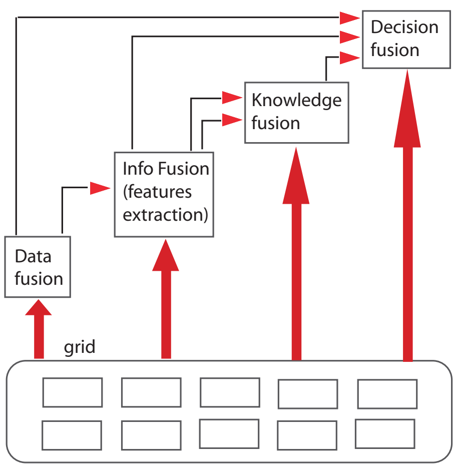
A four-layer fusion framework is proposed for integration. It includes data fusion, information fusion, knowledge fusion, and decision fusion.

- **Data fusion.** This is the lowest level of fusion. The system collects data from multiple sources on the grid and provides categorized data that requires further processing and analysis to users. Users can make decisions or the system can generate decisions based on the process results of collected raw data.

- **Information fusion.** This is the second level of fusion. The system collects data that have common features from multiple sources on the grid and produces a more informative dataset for the users.
- **Knowledge fusion.** This is the third level of fusion. The system finds relevant features of data for each data source on the grid and summarizes knowledge from multiple nodes into a new knowledge base. Users can make decisions using knowledge base.
- **Decision fusion.** This is the highest level of fusion. The system gathers decisions and combines decisions coming from multiple nodes on the grid. The result is given as a system decision. In this chapter, we propose a dynamic decision fusion mechanism. The decision-making process is a negotiable process. It is not only a gather-and-combination procedure, but also allows decision-makers who have different opinions to discuss the issue of concern and make a final decision.

The four-layer frame work is presented in Figure 4.

Figure 4. Four-layer fusion



Hybrid Interactive Fusion

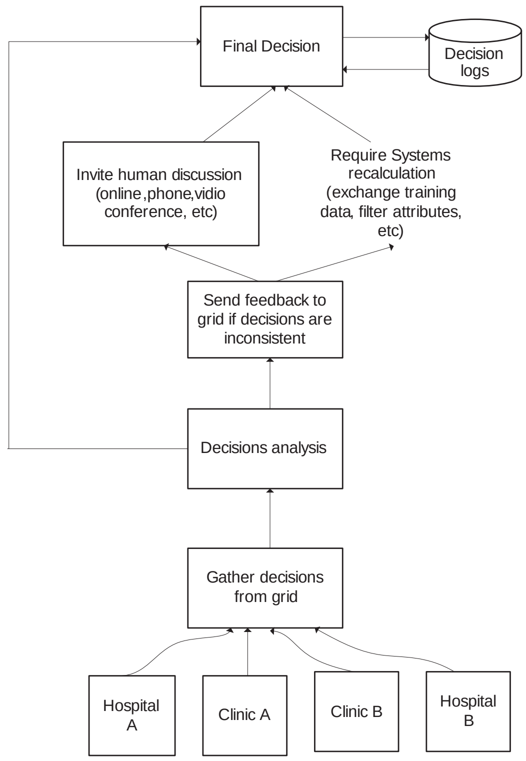
To the best of the authors' knowledge, most of the current fusion systems have one feature in common: They fuse data or decisions on the basis of gathered items. They follow the simple flow — gather data, summarize, and give results. The limitation is clear; it is inflexible and relies excessively on computers.

We propose a dynamic fusion system with interactive discussion between different nodes on the grid. When the gathered decisions are not consistent, some further actions are performed. If service providers are available, they are invited to take a Web conference or telephone conference to discuss among each others. If they are not available, the system may be required to re-draw conclusions by exchanging training datasets or processing training datasets, including filter some unnecessary attributes using different algorithms to handle incomplete data. Compared to traditional fusion systems, it offers higher reliability and diagnosis accuracy by allowing users to confer with others in order to get consistency. This fusion process is described in Figure 5.

The dynamic fusion process simulates the human decision-making process. In the real world, such processes may involve decision-making, discussion, and re-decision-making. The dynamic fusion process is carried out as follows:

1. Systems analyze local datasets separately. Different data analysis applications, such as SVM, gene algorithm, and neuron network, can be used in different sites. The size of datasets varies from node to node.
2. Consistent threshold is set by the user. It includes information about the number of users with consistent and confidence. It provides an indication of how certain we are about the decision fusion and what the acceptable fusion is. Consistent threshold is measured as <X, Y>. X and Y are values between zero to one. A value close to 1 for

Figure 5. Fusion process



X means high consistency and for Y means high certainty. Similarly, a value close to 0 for X means low consistency and for Y means low certainty. <0.7, 0.7> is one example of consistent threshold, which means the acceptable result is that at least 70% of the decisions with average confidence 0.7 are identical. The higher the values of X, Y are, the more difficult to get satisfied answers but the more reliable results are.

3. Collect decisions from the grid and calculate consistent parameter <x, y>. Assume set S and set T are declared as decisions with the same results and total decisions, respectively.

   x = (size of S) / (size of T)
   y = mean of confidence values in S

   For example, there are eight systems in this medical group, and for some given case, six systems have the same diagnosis with confidence (0.8, 0.6, 0.5, 0.9,1, 0.75).
   So, x = 0.75
   and y = 0.76

4. If consistent confidence is less than consistent threshold, that is,

   <x, y> < <X, Y>, x< X and y < Y,

   the further fusion process is activated. Otherwise, results with confidence from different resources of the grid are input of fusion algorithm like voting/weight and neural network to generate the final system result.

5. Results come from the grid and the final system results are written into the database

as fusion history. Every result counts in this system because it will be used as input by fusion algorithm and make contribution for future fusion. Systems accumulate knowledge in this way.

In the best case, all decision-makers on the grid have the same results in the first round of fusion. In this situation, no further fusion is needed. But in the worst case, decisions coming from different makers can be very different. Eight people may have eight different results for the same case. Exchange of opinions and discussion are necessary. Some decision-makers may change their decisions after discussion, and the fusion process is not implemented until it gets satisfied results.

Potential users of this system are doctors, physicians, specialists, and their assistants. The system can provide partial functionalities without human interactions in certain conditions. There are three types of the further fusion processes according to different types of users in the context of the grid:

- **Human-to-human.** The fusion process is implemented in a human-to-human environment. It occurs when users of systems are available and the system works as a decision assistant. The purpose of the system is to provide suggestions for users to make decisions by getting consistent results from group users. Once the system collects results from the grid, it determines the degree of consistency and compares it with the threshold. If it is lower than consistent, the system activates the further fusion including online discussion, e-mail, phone conference, and video conference. Doctors will discuss as to the given case just like the real-world situation. The system fusion decision starts again after some doctors change decisions.
- **Machine-to-machine.** The fusion process is implemented among machines. It occurs when systems work automatically without human interactions. Once the system determines to carry out the further fusion process, one of the two proposed methods can be implemented: re-analyze data using a different training dataset; or re-analyze data using the same training dataset, but only some important attributes are take into account. The first method involves data exchange and low-level data fusion, the latter one involves data preparation and middle-level information fusion. This process may repeat several times until it gets satisfied consistency.
- **Human-to-machine.** Not all doctors are available on the grid. The fusion process is implemented in a hybrid way. If the system needs to carry out the further fusion, it can have two parts: invite available doctors to discuss directly like in the human-to-human situation; or suggest system re-analyze data by providing part of the local set to remote systems or determining the attributes that are used for further analysis. Then, fuse the results again.

The proposed fusion mechanism follows the way humans make decisions. The AI technologies make it smarter and more reliable.

CONCLUSION

A novel method for data fusion with interactions among decision-makers is described. This method takes advantages of observation of other decision-makers' opinions and then modifies the result. This method simulates the process of human decision-making. It involves decision-making, decision fusion, discussion, and re-making, re-fuse. It can improve reliability and flexibility of the fusion system.

The goal of this project is to develop a generalized decision-making and fusion system on grid approach to improve accuracy of diagnose.

A system using SVM for learning from medical data and fuzzy logic for making decisions is designed. Simulations based on Wisconsin Breast Cancer database and Heart Disease (UCI ML Repository) show that the new system is effective in terms of decision accuracy. Even more promising is that higher accuracies are possible if other AI techniques are used.

REFERENCES

Abrahart, R. J., & See, L. (2002). Multi-model data fusion for river flow forecasting: An evaluation of six alternative methods based on two contrasting catchments. *Hydrology and Earth system science, 6*, 655-670.

Aggarwal, C. C., & Parthasarathy, S. (2001). Mining massively incomplete data sets by conceptual reconstruction. In *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 227-232).

Australia's Health Sector. (2002). *Electronic decision support for Australia's health sector.* Report to Health Ministers by the national electronic decision support taskforce.

Azuaje, F., Dubitzky, W., Black, N., & Adamson, K. (1999). Improving clinical decision support through case-based data fusion. *IEEE Transaction on Biomedical Engineering, 46*(10).

Benkner, S., Engelbrecht, G., Backfrieder, W., et al. (2003). Numerical simulation for eHealth: Grid-enabled medical simulation services. In *Proceedings of the PARCO2003, Parallel Computing*, Dresden, Germany.

Bornstein B. H., & Emier, A. C. (2001). Rationality in medical decision-making: A review of the literature on doctors' decision-making biases. *Journal Eval. Clin. Pract., 7,* 97-107.

Butler, R., et al. (2000). Design and deployment of a national-scale authentication infrastructure. *IEEE Computer, 33*(12), 60-66.

Cannataro, M., Talia, D., & Trunfio, P. (2003). Design of distributed data mining on the knowledge grid. *CACM*, *46*(1), 89-93.

Chen, T. M., & Luo, R. C. (1999). Multilevel multiagent based team decision fusion for autonomous tracking system. *Machine Intelligence & Robotic Control, 1*(2), 63-69.

Chow, J. C., Zhu, Q., Fischl, R., & Kam, M. (1993). Design of decision fusion rule for power system security assessment. *IEEE Transaction on Power System*, *8*, 858-864.

CrossGrid Project. (n.d.) Retrieved from http://www.crossgrid.org

Delaney, B. C., Fitzmaurice, D. A., Riaz, A., & Hobbs, R. (1999). Can computerized decision support systems deliver improved quality in primary care. *BMJ, 319*, 1281.

Dick, R. S., & Steen, E. B. (Eds.). (1991). *The computer-based patient record: An essential technology for health care*. Washington, DC: National Academy Press.

Du, W., & Zhan, Z. (2002). Building decision tree classifier on private data. *IEEE International Conference on Data Mining workshop on Privacy, Security, and Data Mining, 14*.

Foster, I., Kesselman C., & Tuecke, S. (2001). The anatomy of the grid: Enabling scalable virtual organizations. *International Journal of Supercomputer Applications, 15*(3).

Freytag, J., Etzold, T., Goble, C., Schward, P., & Apweiler, R. (2003). Information and process integration: A life science perspective. *Schloss Dagstuhl International Conference and Research Center for Computer Science*. Retrieved from http://www.dagstuhl.de/03051/

Globus Toolkit. (n.d.). Retrieved from http://www.globus.org

Gorry, A., & Morton, M. S. (1971). A framework for information systems. *Sloan Management Review, 13*, 56-79.

Hluchy, L., Tran, V. D., Simo, B., Habala, O., Astalos, J., & Gatial, E. (2004). Flood forecasting in CrossGrid project. In *Proceedings of the 2nd European Across Grids Conference*, Nicosia, Cyprus.

Hobbs, F. D., Delaney, B. C., Carson, A., & Kenkre, J. K. (1996). A prospective controlled trial of computerized decision support for lipid management in primary care. *Family Practice, 13*, 133-137.

Hunt, D. L., Haynes, B. R., Hanna, S. E., & Smith, K. (1998). Effects of computer-based clinical decision support systems on physician performance and patient outcomes. *JAMA, 280*, 1339-1346.

Kantarcioglu, M., & Clifton, C. (2002). Privacy-preserving distributed mining of association rules on horizontally partitioned data. In *Proceedings of the ACM SIGMOD Workshop on Research Issue on Data Mining and Knowledge Discovery* (pp. 24-31).

Keen, P. G. W., & Morton, M. S. (1978). *Decision support systems: An organizational perspective*. Reading, MA: Addison-Wesley.

Kline, J. A., Johnson, C. L., Webb W. B., & Runyon, M. S. (2004). Prospective study of clinician-entered research data in the emergency department using an Internet-based system after the HIPAA privacy rule. *BMC Medical Informatics and Decision Making, 4*, 17.

Kusiak, A., Kern, J. A., Kernstine, K. H., & Tseng, B. L. (2000). Autonomous decision-making: A data mining approach. *IEEE Transaction on Information Technology in Biomedicine, 4*(4).

Laskey, K. B., & Mahoney, S. M. (2003). *Knowledge and data fusion in probabilistic networks.* Retrieved from http://ite.gmu.edu/~klaskey/papers/KDFML_Laskey_Mahoney.pdf

Li, Y. C., & Chen, S. (2004). Building Taiwan health grid for health decision making. *ISGC Workshop.*

Mourino, J. C., Martin, M. J., Gonzalez, P., & Doallo, R. (2004). Air pollution modeling in the CrossGrid Project. In *Proceedings of the Computational Science-ICCS 2004: 4th International Conference, Part 1*, Karkow, Poland (pp. 132-139).

Myers, J. W., Laskey, K. B., & DeJong, K. A. (1999). Learning Bayersian networks from incomplete data using evolutionary algorithms. In *Proceedings of the Genetic and Evolutionary Computation Conference.*

Narayan, S. M., Corcoran-Perry, S., Drew, D., Hoyman, K., & Lewis, M. (2003). Decision analysis as a tool to support an analytical pattern-of-reasoning. *Nursing and Health Sciences, 5*(3), 229-243.

Ong, M., Ren, X., Allan, G., Kadirkamanathan, V., Thompson, H. A., & Fleming, P. J. (2004). Decision support system on the grid. *International Conference on Knowledge-based Intelligent Information and Engineering System (KES2004),* New Zealand.

Phegley, J., Perkins, K., Gupta, L., & Dorsey, K. J. (2002). Risk-factor fusion for predicting multifactorial diseases. *IEEE Transaction on Biomedical Engineering, 49*(1).

Poses, R. M., Cebul, R. D., & Wigton, R. S. (1995). You can lead a horse to water — Improving physicians' knowledge of probabilities may not affect their decisions. *Med. Dec. Making,* 65-75.

Reisman, Y. (1996). Computer-based clinical decision aids. A review of methods and assessment of systems. *International Journal of Medical Informatics, 21*(3), 179-197.

Sensmeier, J. (2003, Fall). Advancing the state of data integration in healthcare. *Journal Healthc Inf. Manag, 17*(4), 58-61.

Shamos, M.I. (2004). *Medical and workplace privacy*. Retrieved from http://lorrie.cranor.org/courses/fa04/work2.ppt

Shortliffe, E. H. (1986). Medical expert systems-knowledge tools for physicians. *West J Med., 145*(6), 830-9.

Singhal, A., & Brown, C. (1997). Dynamic Bayes net approach to multimodal sensor fusion. In *Proceedings of the SPIE — the International Society for Optical Engineering* (Vol. 3209, pp. 2-10).

Sintchenko V., & Coiera, E. W. (2003). Which clinical decisions benefit from automation? A task complexity approach. *International Journal of Medical Informatics, 70*, 309-316.

Sloane, E. B., Liberatore, M. J., & Nydick, R. L. (2002). Medical decision support using the analytic hierarchy process. *Journal of Healthcare Information Management, 16*(4), 38-43.

Sprague, R. H., & Calson, E. D. (1980). Building effective decision support systems. *Management Information Systems Quarterly, 4*(4), 1-26

Tracy, C. S., Dantas G. C., & Upshur, R. E. (2004). Feasibility of a patient decision aid regarding disclosure of personal health information. *BMC Medical Informatics and Decision Making, 4*, 13.

UCI ML Repository. (n.d.). Retrieved from http://www.ics.uci.edu/~mlearn/MLSummary.html

Welch, V., Siebenlist, F., Foster, I., Bresnahan, J., et al. (2003). Security for grid services. *12th International Symposium on High Performance Distributed Computing (HPDC-12).* IEEE Press.

This work was previously published in Artificial Intelligence and Integrated Intelligent Information Systems: Emerging Technologies and Applications, edited by X. Zha, pp. 59-75, copyright 2007 by IGI Publishing, formerly known as Idea Group Publishing (an imprint of IGI Global).

Index

A

B

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Z